THE UNIVERSITY OF CHICAGO

STOCHASTIC CAPTURE OF CHROMATIN TOPOLOGICAL DOMAINS BY NUCLEAR MATRIX BOUND RNA POLYMERASE II DETERMINES MONOGENIC Vκ CHOICE

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

COMMITTEE ON IMMUNOLOGY

BY

SOPHIYA KARKI LAMSAL

CHICAGO, ILLINOIS

DECEMBER 2016

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SUPPLEMENTARY FIGURES

LIST OF SUPPLEMENTARY TABLES

ACKNOWLEDEGEMENT

I would like to express my sincere gratitude to my PhD supervisor Dr. Marcus Clark, whose immense knowledge, guidance and encouragement throughout my graduate education brought me here. His warm and personable nature was nurturing in everyway. I greatly appreciate the effort he put into understanding my strengths and weaknesses, and guiding me through my academic journey. I am thankful to him for trusting me with this project and providing necessary criticisms along the way. He generously provided all the resources and initiated collaboration that was necessary for completion of this work. I am also grateful for his time and energy that he put into encouraging my scientific queries.

I would also like to thank my thesis committee members, Dr. Barbara Kee, Dr. Fotini Gounari, Dr. Alexander Ruthenburg and Dr. Wei Du for their advice and support. Their expertise in the field of immunology and gene regulation provided helpful insight for experimental designs. I am also grateful for their suggestions and critical evaluation of my manuscript and thesis. I would like to thank our collaborators Heping Xu from Dr. Harinder Singh's laboratory at Cincinnati Children's Hospital, Shiladitya Banerjee from Dr. Aaron Dinner's laboratory at the University of Chicago and Mark Maienschein-Cline from University of Illinois at Chicago for their expertise. I would specifically like to thank Elizabeth Davis from Michelle Le Beau's laboratory at the University of Chicago, who mentored me personally during initial years of my PhD. I am grateful to all members of Clark laboratory, in particular Margaret Veselits, who provided support both at personal and professional level, and helped me retain a positive attitude. I would also like to thank Keith Hamal, Malay Mandal and Kaitlin Mclean, whose kind and collaborative nature contributed to my productivity in the laboratory.

My work would not have been complete without help from the Imaging Core at the University of Chicago. I would like to thank Christine Labno and Vytas Bindokas at the core for their technical support. Their expertise helped me get through some major hurdles. I would like to thank Christine, in particular, for being considerate and available.

I would like to thank Committee on Immunology (COI) Cluster for taking me into the program and providing me a platform for intellectual growth, and opening doors for my future pursuits. I would also like to thank COI administrative staff for keeping me informed with graduate school requirements.

Graduate school journey can be emotionally draining. I would like to thank my friends who helped me manage my personal turmoil and enjoy the process. I would like to thank Maria Ayala , Jasmin Quandt, Janine Woehlk, Soumi Mondal, Priya Mathur, Mehrnoosh Haghi and Akinola Emmanuel from Gounari lab, who witnessed my ups and downs, and supported me at every step. I would like to thank Karlynn Neu and Jen La for their constant encouragement, and Niraj Ohja, his family and Tara Bhatt for their love and care.

I would not reach this point without support from my parents, siblings and extended family. I would like to thank them for having faith in me and cheering for me all these years. Their support provided all the strength necessary to get through this process. I would also like to thank my in-laws for being supportive of my goals and chosen career paths.

Finally, I would like to thank my dear husband Kamal Lamsal, who stood by me throughout this journey. His persistent nature motivated me to keep going even in the face of failure, and his faith in me motivated me to aim higher. I am grateful to him for making frequent trips to Chicago and fulfilling my longing for family. I would like to thank him for his patience, love and care and hope that his support continues to stay with me.

Abbreviations

AgR	Antigen receptors
BAFF	B cell activating factor of the TNF family
CDK4/6	Cyclin-dependent kinase 4/6
Cer	Contracting element for recombination
ChIP	Chromatin immunoprecipitation
CLP	Common lymphoid progenitor
CTCF	CCCTC-binding factor
Eκi	Intronic kappa enhancer
FISH	Fluorescent *in situ* Hybridization
FOXO1	Forkhead box protein O1
GLT	Germline transcription
GSD	Ground state depletion microscopy
HR	Homologous recombination
HSC	Hematopoietic stem cells
IP	Immunoprecipitation
LMPP	Lymphoid primed progenitors
MAR	Matrix attachment region
MPP	Multi-potent progenitors
NHEJ	Non-homologous end joining
NM	Nuclear matrix
NMP	Nuclear matrix-associated proteins
PAIR	PAX5-activated intergenic repeat

Pre-BCR	Pre-B cell receptor
RAG	Recombination activation gene
RB	Retinoblastoma protein
RME	Random monoallelic expression
RSS	Recombination signal sequences
SATB1	Special AT-rich sequence-binding protein 1
SLC	Surrogate light chains
S/MAR	scaffold attachment regions

Abstract

Mono-specificity of antigen receptors (AgR) is dependent upon monoallelic expression of V(D)J recombined AgR allele. By using combined DNA-RNA Fluorescent *in situ* Hybridization (FISH) and single-cell RNA-sequencing, we report that in the context of *Igκ*, allelic choice for recombination is determined by germline Vκ expression. Allelic choice is associated with asymmetric localization of chosen allele within a rich niche of elongating RNA-Polymerase II (e-Pol II) where Vκ germline expression is initiated upon downregulation of nuclear-matrix (nm) bound cyclin D3. This e-Pol II niche is cylindrical in shape, and allows efficient contraction between Vκ and Jκ compared to the alternate niche. Vκ transcription from the chosen allele occurs within CTCF-defined loops, whereby transcription is predicted to happen by random loop capture by fixed e-Pol II arrayed along the cylinder. Multiple Vκ segments can be active within a cell from the chosen allele, however, only one Vκ gene productively interacts with Jκ for recombination. Based on polymer chain simulation of *Igκ* within a cylindrical niche, the physical constrain imposed by e-Pol II niche both favors CTCF defined loops formation and positioning of loops that reflects observed preference for distal and proximal Vκ usage. Absence of cyclin D3 is also associated with expression of other monoallelic genes such as olfactory receptor (OR) genes, protocadherin genes and randomly monoallelic genes. We propose that in the context of monoallelic and monogenic genes, allelic choice for transcription is determined by asymmetric localization of single allele in rich niche of e-Pol II, whereas in the case of monoallelic but clustered genes, this asymmetric localization is followed by random loop capture transcription across the cluster, allowing monogenic choice for expression. Therefore, *Igκ* transcriptional regulation reported here, shedes light on mechanism of both monoallelic and monogenic choice.

1. Introduction

B cell development

B-lymphocytes are one of the two major components of the adaptive immune system that have evolved to have specific role during immunity. B lymphopoiesis occurs in the fetal liver during embryogenesis and in the bone marrow for the rest of life. B cells are continually generated from long-term repopulating hematopoietic stem cells (HSCs), which have tremendous self-renewal and differentiation potential. HSC differentiation can take lymphoid or myloid route depending upon expression of lineage specification factors as well as the cytokine milieu they reside in. B cells arise from lymphoid primed progenitors (LMPP) that also retain the potential to differentiate to T, NK, granulocytes, DC and macrophage lineage(Rieger and Schroeder 2012). Progression into common lymphoid progenitor (CLP) stage and commitment to B lineage is achieved by concerted action of lineage specification and commitment factors such as EBF1, PU.1, Pax5, E2A and Ikaros (Fig 1.1). These differentiation events are shaped by interactions between early progenitors and specialized mesenchymal stromal cells that secrete important survivals factors such as IL-7 and B cell activating factor of the TNF family (BAFF). Once commitment to B lineage is achieved, the major goal of development events thereafter is to orchestrate highly-regulated proliferative and differentiation events that will eventually generate mature B cells capable of providing protective immunity.

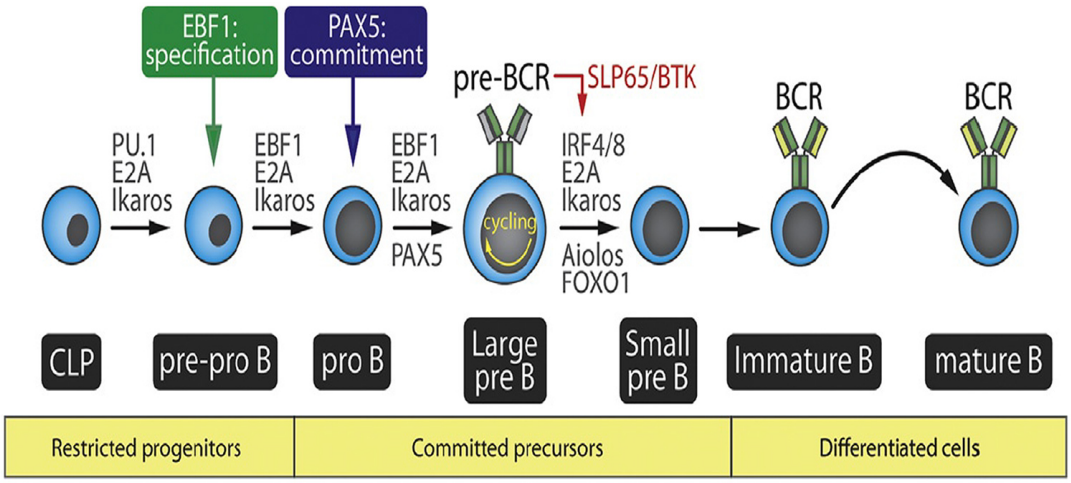


Figure 1.1 B cell commitment.

Hematopoiesis is initiated in the bone marrow by hematopoietic stem cells that differentiate into multi-potent progenitors (MPP). MPPs have potential to differentiate into common lymphoid progenitors (CLP) or common myloid progenitors (CMP), which collectively can give rise to all different cell types in the blood. B cells originate from CLPs, which express lineage specification and commitment factors such as PU.1, Pax5, E2A, EBF1 and Ikaros. After commitment, expression of specific transcription factors drive proliferative and recombination events that leads to expression of BCR and maturation as B cells (adapted from de Almeida et al.2015).

The protective role of B cells relies primarily on the diversity of their surface B cell receptors (BCR) that have capacity to bind unlimited foreign antigens, and initiate effector functions that can eliminate the invaders. BCR and its secreted form called antibody is made by pairing of two identical heavy (*IgH*) and light (*IgL*) chains, created by rearrangement of variable (V), diversity (D) and joining (J) gene segments at *IgH* and *IgL* loci (also see section 1.10). The lymphocyte-specific endonucleases recombination activation gene-1 (RAG1) and RAG2 cooperate to induce double-strand DNA breaks at specific recombination signal sequences (RSS) that flank the V, D, and J gene segments. Subsequent repair of these breaks by non-homologous end-joining (NHEJ) repair machinery, then makes *IgH* and *IgL* polypeptide chains that assemble and express as BCR on the cell surface. BCR so expressed is not only important for antigen encounter but also for providing critical pro-survival signal called tonic signal that helps maintain B-cell survival (Srinivasan et al. 2009).

Stages of B cell development

Bone marrow progenitors that have committed to B lineage undergo series of developmental stages before they can mature as B cells. These stages can be broadly categorized into proliferative and differentiation stages, with each stage defined mutually exclusive of each other (Fig 1.2). Entry into B lineage is marked by differentiation of pre-pro-B cells into pro-B cells ($CD19^{+}B220^{+}IgM^{-}CD43^{+}$), where *IgH* chain undergoes RAG-mediated V(D)J rearrangement. In-frame VDJ rearranged *IgH* chain pairs with surrogate light chains (SLC), λ5 and VpreB and assembles with the signaling subunits Igα and Igβ making a pre-B cell receptor (pre-BCR) that is expressed on the cell surface. These cells, now called large pre-B, express both IL-7R and pre-BCR and undergo proliferation that is primarily driven by IL7R signaling.

Importance of IL-7R signaling is evident by studies showing severely impaired B-lymphopoiesis in mice bearing mutant IL7Rα (a component of IL7R) (Peschon et al. 1994). This defect manifests as early as the pre-pro-B cell stage. After 2-7 rounds of proliferation (Melchers 2005), large pre-B cells exit cell cycle and further differentiate into small pre-B cells ($CD19^+B220^+IgM^-CD43^-$), where they initiate *IgL* rearrangement. In-frame VJ rearranged *IgL* chain that successfully pairs with *IgH* chain makes a BCR that is expressed on the surface of immature B cells ($CD19^+B220^+IgM^+CD43^-$). Cells that undergo successful primary rearrangement express BCR from single *IgH* and *IgL* allele, and show exclusion of the second allele by allelic exclusion mechanisms that can happen either at the initiation level, at the level of feedback inhibition or both (section 1.12) (Vettermann and Schlissel 2010).

Expression of BCR in the immature stage is a major checkpoint of B lymphopoiesis, where newly made BCRs are tested for auto-reactivity. ~85% of newly made BCRs are autoreactive and therefore mechanisms exist to trigger anergy (lack of reaction), secondary rearrangement or deletion of B cells that express them. In particular, anergy and secondary rearrangement in next *Igκ* allele or the *Igλ* alleles (another light chain) by "receptor editing" mechanism, are triggered by low affinity BCR interactions with self-reactive soluble antigens (Melchers 2015, Rolink et al. 1993), whereas deletion is mediated by high affinity interactions specifically with membrane bound antigens (Hartley et al. 1991, Nemazee and Buerki 1989). Cells that undergo successful primary or secondary rearrangements survive and express monoallelic BCR with *IgH* and *Igκ* or *Igλ* chain originating from single allele. However, there is also evidence for dual expression of *Igκ* or *Igλ* in ~5-10% of B cells in the

periphery as a result of receptor editing (Gerdes and Wabl 2004). These cells then exit the bone marrow and migrate to the spleen, where they go through transitional stages before they encounter foreign antigens.

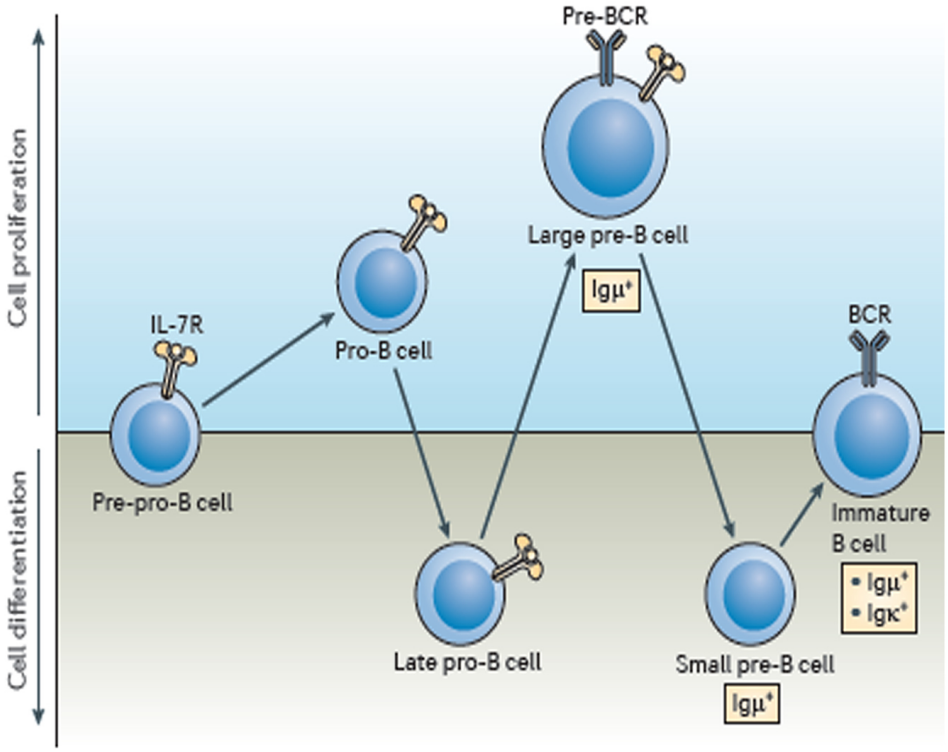


Figure 1.2 Stages of B cell development.

B cell development can be divided into two major stages: proliferation and differentiation. Entry into B lineage is marked by expression of IL-7R along with B cell commitment factors. Response to IL7 leads to first proliferative burst in pro-B stage. After proliferation, the cells differentiate into late pro-B stage, where IgH chain is rearranged. Cells that successfully rearrange their heavy chain pair with surrogate light chain and assemble with signaling units Igα and Igβ to express a pre-BCR in large pre-B stage. These cells undergo 2-7 rounds of proliferative burst before they enter small pre-B where IgL chain rearrangement occurs. Pairing of successfully rearranged IgH and IgL chain then makes a BCR that is expressed on the surface of immature B cells (adapted from Clark et al. 2014).

Orchestration of proliferation and differentiation

During B cell development, proliferative and differentiation (rearrangement) events are segregated to preserve the genomic integrity of cells. In large pre-B cells, simultaneous expression of the IL-7R, which promotes proliferation, and the pre-BCR that promotes rearrangement, creates a major hurdle to achieve this mutual exclusion. However, studies have shown that mutual exclusion of these events is achieved *in vivo* (Clark et al. 2014), and is mediated by modulation of external cues as well as activation of antagonistic signals downstream of the IL7R and the pre-BCR within the cells (Fig 1.3). For example, in the bone marrow, early B cell progenitors that are proliferative (pro-B and large pre-B) reside in a niche of mesenchymal stromal cells that express high levels of IL-7, which is important for proliferation (Tokoyoda et al. 2004). Furthermore, the stromal cells also actively participate in restricting progenitors in this IL-7 niche by secreting a chemokine CXCL12, which acts as a chemoattractant for the progenitors that express its cognate ligand, CXCR4. Small pre-B cells, on the other hand, reside away from this niche, and show down-regulation of CXCR4, while augmenting pre-BCR signaling.

IL-7R signaling favors proliferation while antagonizes recombination. For example, large pre-B cells residing in IL-7 niche, signal via their IL7R receptor and activate STAT5, which both targets expression of cell cycle genes to promote proliferation and repress *Igκ* associated enhancer (Eκi) to antagonize *Igκ* accessibility (Mandal et al. 2011). Similar to IL7R mutant, mice with deletion of *Stat5a* and *Stat5b* show arrest at the pre-pro-B cell stage (Yao et al. 2006). Furthermore, IL7R signaling activates PI3 kinase that antagonizes expression of forkhead box

protein O1 (FOXO1), which is an inducer of RAG1 and RAG2 (Herzog et al. 2008, Clark et al. 2014). Upon migration away from IL-7 niche, and exit from cell cycle, signaling via pre-BCR prevails and antagonizes further IL-7R signaling.

Dominance of pre-BCR signaling over IL-7R signaling has two major implications. One is that further *IgH* rearrangement is prevented by mechanisms of feedback inhibition and second, *IgL* rearrangement is induced by attenuation of cell cycle and expression of components of recombination machinery. Pre-BCR has a direct role in *Igκ* locus accessibility and recombination, as transgenic *Igμ (IgH)* , expression alone leads to increase in accessibility in *Rag2*−/− pro-B cells (Young et al. 1994, Stanhope-Baker et al. 1996). Similarly, mutation of pre-BCR components enhances proliferation (Flemming et al. 2003, Xu et al. 2007) while, constitutive expression of the receptor skews the balance to light chain recombination (van Loo et al. 2007).

Pre-BCR signaling through downstream adapter molecule, SLP65 (also known as BLNK), is important for differentiation, as evident by developmental arrest at large pre-B stage in SLP65 deficient mice (Jumaa et al. 2001). SLP65 induces expression of IRF4, Ikaros and Alios transcription factors which down-regulate SLC components (Thompson et al. 2007, Ma et al. 2008), important for differentiation. It also inhibits proliferation by repressing c-myc expression(Ma et al. 2010) and helps stabilize FOXO transcription factors that directly induce RAG expression (Amin and Schlissel 2008). Furthermore, SLP65 induced expression of IRF4/8 transcription factors has a direct role in *Igκ* transcription. Ras–Erk pathway under

pre-BCR also has important role *Igκ* recombination(Mandal et al. 2009, Shaw et al. 1999). This pathway targets expression of transcription factor, E2A, which directly activates *Igκ* associated enhancers (Eκi and 3'κ), necessary for recombination, while also repressing expression of its inhibitor id3 (inhibitor of DNA binding 3)(Mandal et al. 2009, Kee, Quong, and Murre 2000). These data suggests that various opposing signaling pathways are in place to orchestrate proliferation and rearrangement during B cell development, that helps preserve the genomic integrity of recombining cells .

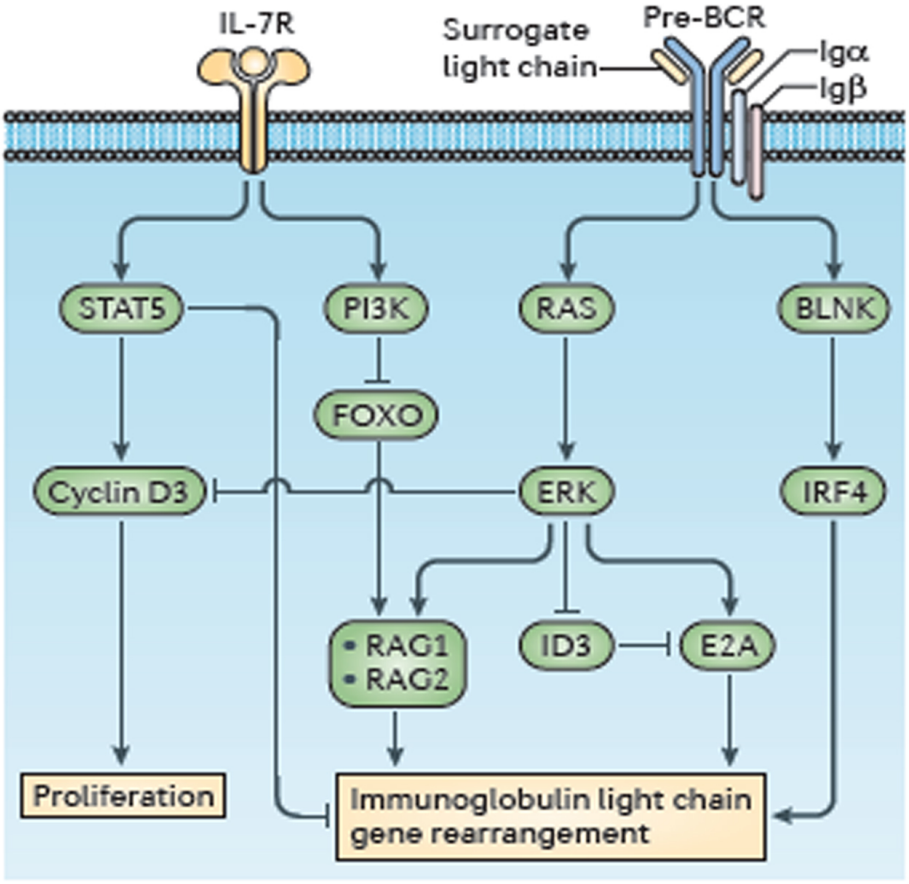


Figure 1.3 Orchestration of proliferation and differentiation.

Signals downstream of IL-7R and pre-B cell receptor coordinate to allow proliferation versus rearrangement. IL-7R signal activates STAT5, and leads to activation of cyclin D3 that mediates proliferation. Binding of STAT5 to the *Igκ* locus inhibits rearrangement simultaneously. PI3K activated downstream of IL-7R phosphorylates FOXO transcription factors and targets them for degradation, preventing expression of RAG1 and RAG2. Activation of Ras-Erk pathway downstream of pre-BCR, on the other hand, opposes proliferation by targeting degradation of cyclin D3 and relieving E2A from Id3 inhibition, necessary for GLT and recombination of *Igκ*. Furthermore, pre-BCR signal via BLNK targets expression of IRF4, which is necessary for *Igκ* transcription (adapted from Clark et al.2014).

Igκ locus and regulatory elements

In mice, *Igκ* stretches across 3.2 Mb in chromosome 6, with >100 variable genes (V) arrayed on the 5' end of the locus and 5 joining gene (J) segments on the 3' end, with single constant (C) segment (Fig 1.4)(Martinez-Jean, Folch, and Lefranc 2001). Vκ is grouped into 18 families based on sequence homology. Members of Vκ family are semi-clustered, but may also intersperse within other families. Each Vκ gene is named starting with family name followed by the gene's position, numbered in ascending order from 3' to 5' end of the locus. The variable gene segments are organized into distal, intermediate and proximal loops. During recombination, distal V-segments are found in close proximity to J-Cκ cluster, as evident by juxtaposition of Vκ and Jκ seen by two-color FISH assays and direct interaction of Vκ with Jκ and downstream enhancers seen by chromosome conformation capture assays (Stadhouders et al. 2014). This juxtaposition of Vκ and Jκ is suggestive of locus contraction, implying that the linear topology of *Igκ* changes drastically during recombination.

Accessibility of *Igκ* is dependent upon several *cis*-acting regulatory elements such as promoters, enhancers and silencer elements (Fig 1.4). Eκi located between Jκ and Cκ is both important for germline transcription (GLT), *Igκ* rearrangement and maintenance of κ:λ ratio (Inlay et al. 2002). 3' Eκ acts independently of Eκi, however, shares overlapping roles with Eκi in mediating *Igκ* GLT, locus demethylation and recombination and additionally has specific role in mediating high-level transcription of rearranged *Igκ* (Inlay et al. 2006). Both Eκi and 3'Eκ enhancer have been implicated in directing monoallelic association of *Igκ* to the peri-

centromeric heterochromatin as a way of facilitating allelic exclusion (Hewitt et al. 2008, Inlay et al. 2002). Increased interactions of distal region to these enhancer elements suggest that they are also involved in locus contraction during recombination(Stadhouders et al. 2014).

In addition to enhancer elements, *Igκ* also has silencer elements that constitute barriers that keep the locus insulated from nearby regions. Presence of multiple enhancer elements particularly necessitates *Igκ* to harbor insulator that keep their activity in check. DNAse hypersensitivity assay identified two hypersensitivity sites (HS) along *Igκ* locus located 5.5kb and 12kb upstream of Jκ1. The proximal HS element termed "silencer in the intervening sequence" (sis) and distal HS called "contracting element for recombination" (cer) are both located between Vκ and Jκ and are occupied by known insulator CCCTC-binding factor (CTCF) and are involved in both insulator function and *Igκ* contraction (Xiang, Park, and Garrard 2013). This is particularly evident by phenotype of *sis/cer* double knockout mice, which show dramatically skewed proximal Vκ usage, suggesting that the two elements cooperate to insulate proximal Vs from interaction with enhancers (Xiang, Park, and Garrard 2014) and facilitate locus contraction and distal usage. These data suggests that the enhancer and silencer elements have important roles in mediating locus accessibility, contraction and boundary enforcement that are important for proper organization and functionality of *Igκ.*

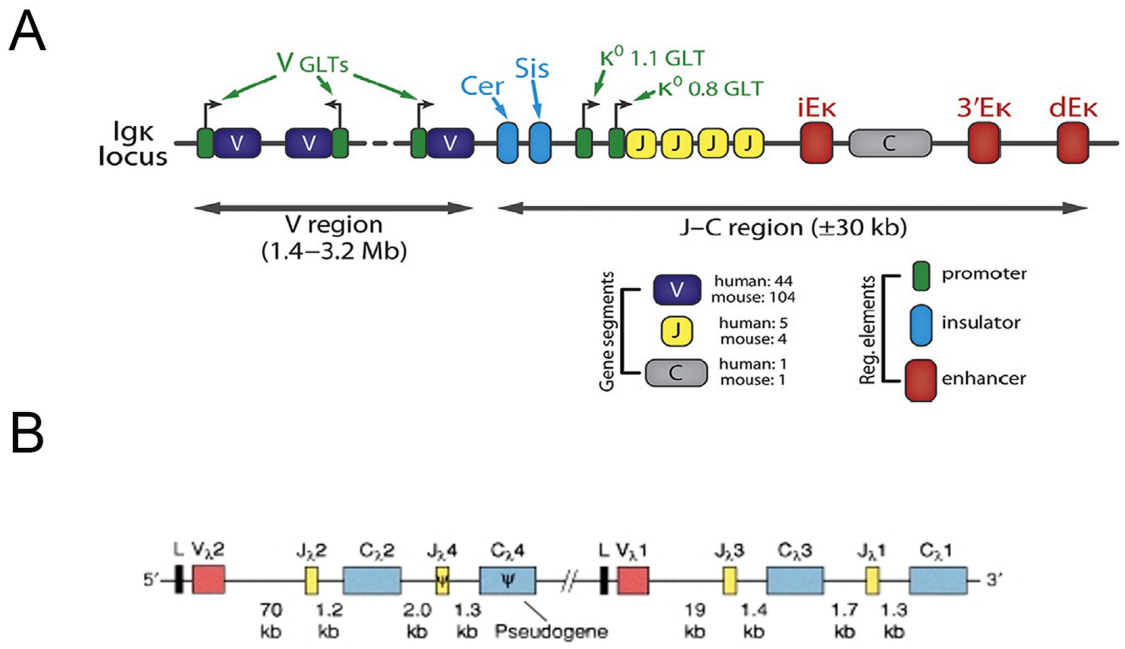


Figure 1.4 *Igκ* and *Igλ* locus

(A) In mice, *Igκ* locus spans across 3.2 Mb of chromosome 6, and comprises, from 5' to 3', 174 V gene segments (~96 functional) that belong to 18 families, 5 J segments (4 functional) and a single Cκ. There are three defined enhancers at the *Igκ* locus: Eκi, 3'Eκ and dEκ. The enhancers are highly conserved in termed of location, and have an important role in Vκ-Jκ recombination and tissue-specific *Igκ* expression (adapted from de Almeida et al. 2015).

(B) *Igλ* locus is located on chromosome 16 that spans 240 kb. The locus comprises 2 *Igλ* V genes, 4 Jλ genes and 4 Cλ genes which belong to 2 subgroups. Each of the Cλ gene is preceded by one (or two) Jλ gene(s), organized into two clusters: J2-C2-J4-C4 and J3-J3P-C3-J1-C1. Jλ-Cλ clusters are then preceded by respective V genes (adapted from http://cig.salk.edu/bicd_140_W99/lecture5.htm).

3D topology of *Igκ*

The 3D topology of *Igκ* locus during recombination constitutes pre-organized V loops that have progressively undergone contraction towards Jκ for recombination, a function attributed to a number of proteins, including PAX5, YY1 and CTCF (Shih and Krangel 2013, Atchison 2014, Fuxa et al. 2004). Among these proteins, CTCF has emerged as one with several functional implications during B cell development. For example, similar to *IgH*, CTCF binding to multiple sites within the *Igκ* locus is predicted to organize the locus into higher-order DNA loops, also called topologically associated domains (TADs) (Shih and Krangel 2013, Stadhouders et al. 2014), whereas its binding to *cer* element has been shown to be important for mediating long range interactions between Vκ and Jκ and enhancer elements at 3' end (Fig 1.4) (Xiang, Park, and Garrard 2013).

CTCF's role as a chromatin organizer is made possible by cohesion, which can clamp homodimers of CTCF molecules bound to DNA at large distances (Degner et al. 2011), thereby bringing them together to form the base of the loop. Based on chromosome confirmation capture assays, CTCF-loops form more frequently between convergent forward and reverse CTCF binding sites than between convergent sites in the same orientation or divergent sites (facing away from each other) (Rao et al. 2014). The loops so formed regulate enhancer-promoter interactions throughout the genome. Insulation of these interactions is associated with CTCF binding to the loop base also called boundaries, whereas support for these interactions is associated with CTCF binding at or near gene promoters or directly with RNA-polymerase II(Ong and Corces 2014, Chernukhin et al. 2007). In the context of *Igκ*, CTCF binds to *sis*

element (Fig 1.4) and insulates proximal Vκ-promoters from *Igκ* enhancers (also see section 1.4) (Ribeiro de Almeida et al. 2011), whereas supports transcriptional activation, particularly of the distal Vκ segments, by binding at or close to their promoters (unpublished). What is unclear is how CTCF-mediated 3D topology of *Igκ* affects V accessibility, and whether single or multiple loops participate in V gene accessibility and recombination in a single cell. Furthermore, whether loop accessibility follows locus contraction or vice-versa is not known. Additionally, a 3D model that predicts Vκ usage based on CTCF-formed TADs is lacking.

Spatial remodeling of *Igκ* during development

As B cell development progresses, *Igκ* alleles also undergo spatial remodeling through various stages, as depicted by their nuclear relocation from nuclear periphery to the center of the nucleus and locus contraction and decontraction events that follow. These events are temporally and stage-specially regulated to repress accessibility and recombination in earlier stages and to promote them in small pre-B stage. Nuclear localization of genes has been co-related with expression status of gene segments, with laminar localization and association with heterochromatin linked to gene repression(Bickmore and van Steensel 2013, Akhtar and Gasser 2007), and relocation of genes off the lamina and association with euchromatin linked to gene activation. Similar mode of regulation has been noted for *Igκ*, where nuclear localization changes all throughout the development (Goldmit et al. 2005, Kosak et al. 2002).

In embryonic stem cells and *WT* pre-pro-B stage, both *Igκ* alleles are located at the nuclear periphery. However, upon entry into pro-B stage, both *Igκ* loci relocate away from the

nuclear periphery (Kosak et al. 2002) and towards the interior of the nucleus, although this relocation does not guarantee locus accessibility in this stage. In small pre-B cells, one of the relocated alleles binds Ikaros and heterochromatin protein HP1γ and associates with pericentromeric heterochromatin, whereas the other allele associates with euchromatin (Goldmit et al. 2005), and shows preferential demethylation and deposition of RAG and activating epigenetic marks(Goldmit et al. 2005). These data suggests that *Igκ* undergoes nuclear relocation symmetrically at both alleles, however, is recruited to heterochromatin or euchromatin in an asymmetric manner.

After relocation from nuclear periphery to the interior of the nucleus, *Igκ* undergoes locus contraction. Contraction brings distal V genes in close proximity to the J cluster, where RAG is deposited (Ji et al. 2010). Minimal RAG deposition across V region, underscores the importance of this contraction event (see section 1.7). However, whether locus contraction is symmetric or asymmetric for the two alleles, and how that relates to heterochromatin and euchromatin localization of individual allele has not been investigated.

Epigenetic regulation of *Igκ* locus

Increase in *Igκ* accessibility in small pre-B cells happens concurrent with or prior to locus contraction. Interestingly, levels of accessibility are drastically different between Vκ and Jκ. For example, the J-Cκ cluster shows several fold higher accessibility compared to the V region in a v-abl transformed pre-B cell line, as seen by DNAse hypersensitivity assay (Fitzsimmons et al. 2007). This increase in hypersensitivity at J-Cκ is consistent with prominent deposition of epigenetic marks such as H3 and H4 acetylation (Fig 1.5)(Xu and Feeney 2009), H3K9Ac, H3S10pK14Ac (Mandal et al. 2015) as well as RAG1 and RAG2 binding (Ji et al. 2010) noted at this region. These epigenetic landscapes assembled on J-Cκ depend upon pre-BCR signaling as exemplified by Erk-mediated deposition of H3K9Ac, H3S10pK14Ac marks over this region (Mandal et al. 2015).

Differential epigenetic regulation between Vκ and Jκ is also mirrored by repressive histone modifications (H3K27me3) in pro-B cells, where they are more prominent on J-Cκ compared to Vκ segments (Xu and Feeney 2009, Mandal et al. 2015, Mandal et al. 2011). These repressive marks at J-Cκ are dependent upon IL-7R signaling, whereby activated STAT5 downstream of IL-7R recruits methyltrasferse Ezh2 , which decorates the region with H3K27me3 marks(Mandal et al. 2011). These data suggest that the epigenetic landscape of J-Cκ is directly linked to the signals received by the IL-7R and the pre-BCR, which are inherently antagonistic for *Igκ* accessibility in pro-B stage and supportive in small pre-B stage.

Differences in V and J accessibility remarkably also apply to other antigen receptors loci, where V segments are relatively devoid of canonical activating or repressive histone marks and RAG binding compared to the J cluster (Xu and Feeney 2009). V segments do appear to have some H3K4me2 marks, which are more marker of poised chromatin than active or repressed, but are not as prominent as other canonical marks at J-Cκ. It is likely that other unidentified histone modifications mark the V regions, however, these have not been extensively studied or reported. These data imply that while J accessibility is marked by epigenetic mechanisms, additional means are in place to mark V accessibility at AgR loci.

Since activating epigenetic marks reflect an active locus, studies have examined how these marks are associate with monoallelic recombination of *Igκ*. In CD43$^-$ small pre-B cells, activating epigenetic mark at J-Cκ is assembled in a monoallelic manner, with *Igκ* allele located in euchromatin showing preferential deposition (Goldmit et al. 2005) at J-Cκ. This suggests that epigenetic modification together with euchromatin localization are directly involved in determining allelic choice for recombination. However, the presumed monoallelic accessibility of J-Cκ based on this finding goes against biallelic germline transcription of Jκ (see section 1.8), which is another marker of accessibility. Therefore, although epigenetic modifications differentially mark Vκ and Jκ, whether they also determine monoallelic *Igκ* recombination is still unclear.

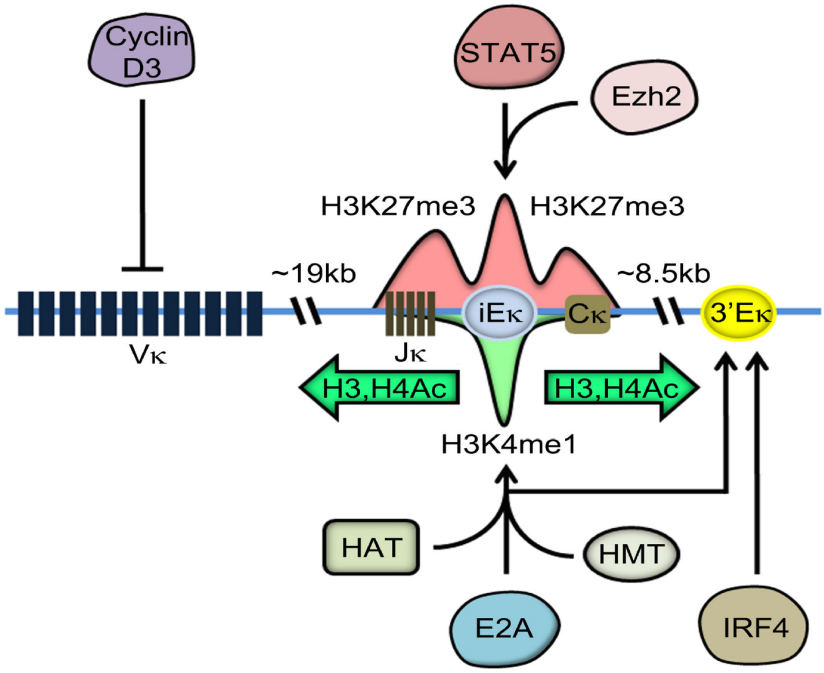


Figure 1.5 Epigenetic regulation of *Igκ*.

J-Cκ cluster is epigenetically regulated during stages of B cell development. In pro-B cells and large pre-B cells, Eκi located within the J-Cκ cluster is decorated by H3K27me3 marks. This repressive mark is deposited by histone methyltransferase Ezh2, recruited by STAT5 tetramer binding Eκi. In small pre-B cells, E2A replaces STAT5 tetramer and binds Eκi, recruiting histone acetyl and methyltransferases that deposit H3/H4Ac and H3K4me1marks respectively. Vκ gene segments do not contain appreciable repressive or activating histone marks. There is however role for cyclin D3 in their regulation (image courtesy of Malay Mandal).

Accessibility of *Igκ* locus: germline transcription

Apart from activating epigenetic marks, another hallmark of *Igκ* accessibility is germline transcription (GLT) of Vκ and Jκ gene segments that occurs prior to or concurrent with recombination. Unlike Vκ-Jκ rearranged transcripts, which are coding, germline Vκ and Jκ transcripts are non-coding. Germline transcription has been confirmed for all antigen receptor loci, and is regulated in a lineage and stage-specific manner(Abarrategui and Krangel 2009). However, its biological relevance is still unclear.

GLT forms the basis of "Accessibility model of V(D)J recombination" proposed by Yancopoulous and Alt, who suggested that these developmentally timed germline transcripts "open" chromatin surrounding V,D and J gene segments and make them accessible for rearrangement (Yancopoulos and Alt 1985). Although the model is widely favored, whether germline transcription is a cause or effect of accessible chromatin could not be tested for a long time. Krangel and colleagues inserted a transcription terminator within TCR Jα gene cluster, and found that terminating germline Jα transcription specifically abrogated rearrangement with Jα segments 3' of the transcriptional terminator, and not 5' end, suggesting that germline transcription mediates accessibility (Abarrategui and Krangel 2006) and recombination, rather than just being an outcome of ongoing recombination.

Jκ GLT proceeds from two promoters located 3.5kb (distal) and 100 bp (proximal) 5' of Jκ1, giving rise to 1.1kb ($\kappa^0$1.1) and 0.8kb (κ^0 0.8) transcripts as spliced products between each promoter and Cκ, respectively (Fig 1.4)(Martin and van Ness 1990). Germline Jκ-Cκ

transcription from the two promoters is important for Vκ-Jκ recombination, as deletion of both promoters results in impaired Vκ-Jκ rearrangement(Cocea et al. 1999). Studies have shown that vast majority of germline transcripts are produced from distal promoter(Amin et al. 2009). The proximal promoter has been shown to be primarily important for secondary rearrangement (Vettermann et al. 2015).

Unlike Jκ GLT, Vκ GLT proceeds from single promoter located 5' of every V segment, making transcripts with 100-400bp long 3'UTR(de Almeida, Hendriks, and Stadhouders 2015, Duber et al. 2003). Interestingly, more than 50% V segments are in reverse transcriptional orientation and occupy intermediate region of the locus. There are several functional elements at ~200bp promoter region upstream of each V, of which only two universal elements have been identified. The first is a tissue-non-specific TATA box that is necessary for recruitment of transcription factors, whereas the second is a conserved ATGCAAAT octamer motif that is tissue specific and is important for high level of Ig transcription (Maston, Evans, and Green 2006). Vκ promoters not only drive germline transcription but also drive transcription of newly rearranged Vκ-Jκ transcript, which undergoes splicing to Cκ to produce full-length light chain. Similar to epigenetic modifications, germline transcription has also been studied on an allele-by allele basis. At least for Jκ, germline transcription is biallelic (Amin et al. 2009). However, whether Vκ germline transcription is mono or biallelic is not known

Transcription factors involved in *Igκ* development and recombination.

Lineage and stage specific changes leading up to *Igκ* recombination can only be achieved by expression of specific transcription factors (TF). Those with important role in B cell development include Ikaros, Alios, IRF4/8, PU.1, PAX5, Spi-B, Spi-C, E2A, EBF1, FOXO and STAT5. Chromatin immunoprecipitation (ChIP) studies in different stages of B cell development have allowed us investigate binding of several TFs across *Igκ* locus, however, mechanistic analysis of their role in *Igκ* development and recombination is still at its early stage.

Ikaros family members Ikaros and Alios have been shown to interact with Nucleosome Remodeling Histone Deacetylases (NURD) and SWI/SNF complexes, allowing them to act as both repressors and activators (Kim et al. 1999). Ikaros is one of the first transcription factors necessary for differentiation of multi-potent progenitors to CLPs(Sellars, Kastner, and Chan 2011). *In vitro* experiments have shown a role for Ikaros together with Alios in repression of pre-BCR SLC components as well as cell cycle genes to help transit from proliferating to resting pre-B stage (Ma et al. 2010, Ma et al. 2008, Thompson et al. 2007). Ikaros likely has specific role in *Igκ* rearrangement in small pre-B cells, as it is recruited to almost all Vκ genes and regulatory elements at *Igκ*, which is in contrast to pro-B stage, where no recruitment can be found (Fig 1.6) (Ferreiros-Vidal et al. 2013, Stadhouders et al. 2014). As mentioned before, Ikaros in association with HP1γ can also recruit *Igκ* to pericentromeric heterochromatin, and contribute to establishment of allelic exclusion(Goldmit et al. 2005).

IRF4 and IRF8 transcription factors also have important role during B cell development, since mice double deficient of the two factors are arrested in pre-B stage(Lu et al. 2003). IRF4/8 deficient cells are highly proliferative and show impairment of *Igκ* rearrangement. The two transcription factors have redundant roles in inducing GLT and rearrangement, as expression of either is enough for this function (Johnson et al. 2008, Ma et al. 2006). IRF4/8 are in the same transcription network as Ikaros/Alios, since both IRF4 and IRF8 can induce Ikaros/Alios expression and similar to Ikaros/Alios induce *Igκ* GLT and recombination (Ma et al. 2008). Both Ikaros and IRF4 also regulate expression of chemokine receptor CXCR4 that promotes migration of pre-B cells in response to the chemokine CXCL12 (Johnson et al. 2008, Schwickert et al. 2014). However, Ikaros and IRF4 have opposing role in nuclear positioning of *Igκ*. While Ikaros binding to 3'Ek recruits one *Igκ* to peri-centromeric heterochromatin by binding to Sis element, IRF4 recruits allele away from peri-centromeric heterochromatin(Johnson et al. 2008).

PU.1's role in specification and survival of early lymphoid progenitors is well established. Recently, PU.1 has also been implicated in *Igκ* recombination via 3'Eκ binding (Hodawadekar et al. 2012). PU.1 ChIP-seq data of pro-B cells showed PU.1 binding to all *Igκ* regulatory regions and many Vκ gene promoters (Schwickert et al. 2014, Whyte et al. 2013), implying its role beyond specification. PU.1 ChIP-seq of small pre-B cells is lacking. It shares common function with SpiB and SpiC, both of which have been shown to be able to recruit other transcription factors to 3'Ek. PU.1 binding to 3'Ek increases recruitment of E2A and PAX5, both of which have important roles in *Igκ* germline activation and locus contraction (Bemark et al. 1999).

PAX5 serves as an important TF necessary for commitment to B cell lineage, and shows a broad range of expression from early progenitors to mature B cell stage. Mice deficient in PAX5 show arrest in early pro-B stage (Nutt et al. 1997). PAX5 deficient B cells do express other B cell specific genes, however, they are not restricted to B lineage, suggesting that PAX5 not only activates B lineage-specific program, but also simultaneously represses non-lineage specific genes (Schaniel et al. 2002). PAX5 has also been implicated in controlling *IgH* rearrangement, since PAX5 deficient pro-B cells show ~100-fold lower recombination efficiency of distal VH family VHJ588, despite normal proximal recombination efficiency (Nutt et al. 1997). 3D- FISH showed that PAX5 deficient pro-B cells are in an extended configuration, suggesting that PAX5 has a role in *IgH* locus contraction (Fuxa et al. 2004).

Recently, several PAX5-activated intergenic repeat (PAIR) elements were identified along the distal region of *IgH* locus. These elements are bound by PAX5, E2A, YY1 and CTCF/cohesion complexes, and are suggested to participate in distal locus contraction, which is also supported by chromosome conformation capture assays (Ebert et al. 2011, Medvedovic et al. 2013, Verma-Gaur et al. 2012). In *Igκ* locus, there are two tandem PAX5 binding elements upstream of κ^0 promoter, termed KI and KII (Ferradini et al. 1996, Tian et al. 1997). Mutation of these binding elements decreased *Igκ* rearrangement to 1/10 of the normal, however, did not affect GLT. However, deletion of PAX5 itself, instead of the elements had more dramatic effect on GLT, as seen by failure to initiate *Igκ* GLT by PAX5 deficient IgH^+ pre-B cells (Sato et al. 2004). PAX5 may also be involved in nuclear relocation of *Igκ*, since PAX5 deficient mice show preferential

Igκ localization at the nuclear periphery (Sato et al. 2004), which is thought to be repressive. These results suggest that PAX5 has broad function during B cell development.

EBF1 similar to PAX5 is important for early B lineage specification, as mice lacking in EBF1 shows complete block in B lymphopoiesis at the pre-pro-B cell stage (Lin and Grosschedl 1995). Developmental block of E2A deficient mice can be rescued by ectopic expression of EBF1, suggesting that E2A initiates EBF1 transcription (Seet, Brumbaugh, and Kee 2004). EBF1 is part of a positive feedback loop with E2A, FOXO1 and IL7R that cooperate with pre-BCR. EBF1 like PAX5 also restricts alternative lineages (Pongubala et al. 2008). In the context of *Igκ*, EBF1 has been shown to bind distal Vκ region, although no binding is observed at *Igκ* regulatory elements (Schwickert et al. 2014). This binding pattern suggests a potential role of this factor in locus contraction. EBF1 has also been shown to control expression of important modulators of *Igκ* rearrangement such as RAG1, RAG2 and FOXO1 (Zandi et al. 2008).

E2A in concert with PAX5 and EBF1 is part of a global network of transcription factors necessary for initiation of B lineage program(Lin et al. 2010). E2A comprises of E12 and E47 transcription factors that are encoded by *Tcfe2a* gene and arise by its alternative splicing. *Tcfe2a*$^{-/-}$ mice show complete block at the pre-pro-B cell stage, even before DH-JH rearrangement (Bain et al. 1994). E2A has specific role in *Igκ* rearrangement, as shown by specific binding of this factor on Eκi(Murre et al. 1989). E2A binds all three enhancers and multiple Vκ genes. Mutation of E2A binding element on Eκi shows severe defect in *Igκ* rearrangement, which phenocopies deletion of the entire enhancer, suggesting that E2A is the

driver of Eκi activity(Inlay et al. 2004). E2A binds to E-box sites distributed non-randomly throughout Vκ. Most E2A occupancy is found close to Vκ gene segments (within 200bp) either in 5' or 3' end. Nearly 37% of all functional Vκ genes show E2A occupancy (Fig 1.6) (Stadhouders et al. 2014), mostly favoring forward oriented Vκs. Presence of E2A on Vκ gene segments is also associated with their usage in the periphery and interaction points for locus contraction. The importance of E2A in *Igκ* recombination is further highlighted by the observation that by ectopic expression of E2A proteins together with RAG1 and RAG2 promotes *Igκ* GLT and Vκ-Jκ rearrangement in non-lymphoid cells (Romanow et al. 2000). E2A also appears to be necessary for secondary rearrangement and receptor editing since these processes are severely affected in immature B cells of E2A-heterozygous mice (Quong et al. 2004). Furthermore, increased activation of *Igκ* locus is also partly dependent upon synergistic interaction between IRF4 and E2A, as shown by IRF-4 mediated recruitment of E2A to the 3'Eκ in pro-B cells (Lazorchak, Schlissel, and Zhuang 2006). Therefore, similar to PAX5, E2A also has a broad role during B cell development.

FOXO factors are expressed downstream of PI3K and AKT signaling pathways (Zhang et al. 2011). AKT kinases phorsphorylates FOXO factors FOXO1 and 3a, which are then targeted for degradation (Amin and Schlissel 2008). PI3K and AKT are activated by IL-7R signals, and stimulate cell proliferation. Therefore, when IL-7R signaling attenuates, during differentiation, PI3K-AKT pathway activity is also diminished, leading to stabilization of FOXO transcription factors that are important for differentiation (Clark et al. 2014). More importantly, FOXO factors have been shown to directly up-regulate RAG expression (Amin and Schlissel 2008, Dengler et

al. 2008).

Another transcription factor important for B cell development is STAT5. Complete inactivation of STAT5 arrests cell in pre-pro-B cell stage (Yao et al. 2006). IL-7R signaling in pro-B cells activates STAT5 that controls pro-B survival and proliferation (Malin et al. 2010). In addition, activated STAT5 inhibits *Igκ* transcription by hampering E2A binding to the Eκi(Mandal et al. 2011). STAT5 binds to Eκi as a tetramer and recruits PRC2 protein EZH2, which decorates Eκi element and Jκ cluster with repressive H3K27me3 marks. Attenuation of IL7R signaling leads to increased binding of E2A to Eκi, thus, activating the locus for recombination.

Overlap in binding pattern of multiple transcription factors at various sites in the *Igκ* locus suggests their cooperative mode of action in *Igκ* accessibility and recombination. However, mechanistic analysis of how these TFs are involved in these events is lacking. Furthermore, their binding in pro-B cells and small pre-B cells in the absence of apparent histone modifications (see section 1.7) across Vκ suggests for their role as 'pioneer' factors. How these factors coordinate with RNA-Pol II and RAGs that mediate germline *Igκ* transcription and recombination, respectively, has also not been investigated yet.

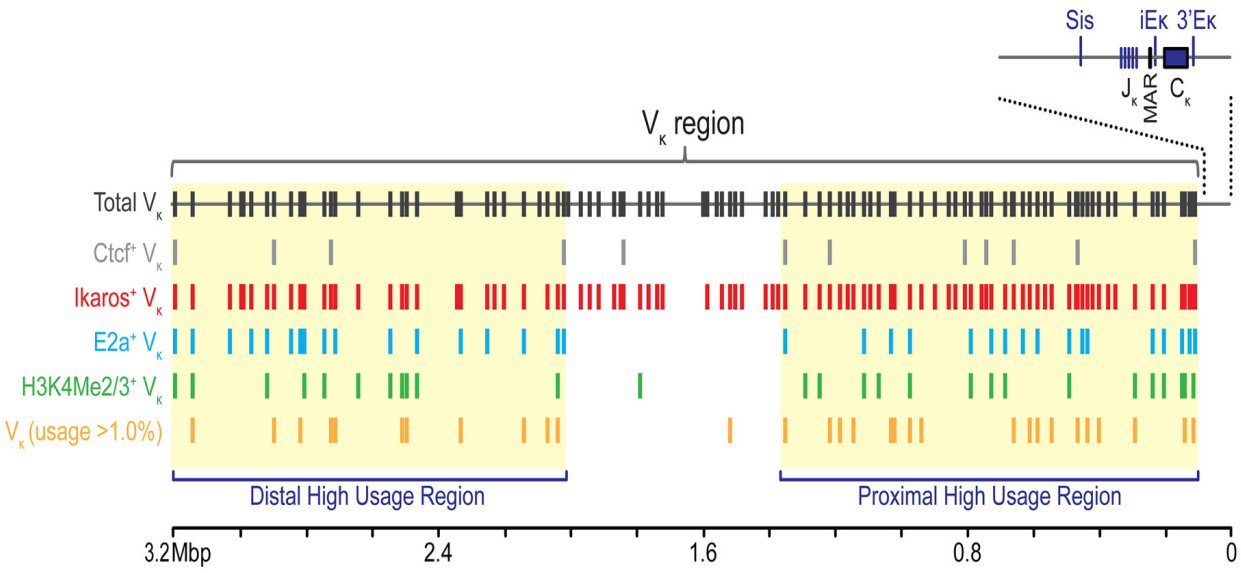


Figure 1.6 Transcription factor binding across *Igκ*.

ChIP seq data for CTCF, Ikaros, E2A, H3K4me2/3 acquired from mutiple studies is shown along with Vκs that constitute >1% of the repertoire. Ikaros binds almost all V genes, whereas E2A and H3K4me2/3 marks are only present on a subset of Vs. This representation does not show peak height, and therefore does not discriminate regions of high and low binding. As can be seen, the intermediate Vs shows minimal TF binding, except for Ikaros (adapted from Stadhouders R et al. 2014) .

Vκ-Jκ rearrangement

All antigen receptors undergo V(D)J rearrangement for expression of their receptors that provide them their unique identity. While rearrangement between V,D and J gene segments occurs at the DNA level, association with the constant region (s), occurs at the level of mRNA processing (Fig 1.7A). Rearrangement is mediated by RAG-1 and RAG2. The RAG-1/2 complexes recognize specific RSSs flanking Vκ and Jκ genes, and make double-stranded breaks that allow recombination between the cleaved segments. RSSs consist of conserved heptamer and nonamer elements that are separated by a less conserved spacer region of 12 or 23 base pairs (Fig 1.7B). Recombination only occurs between segments that are separated by 12-bp and 23-bp spacers. DNA double stranded breaks then get repaired through classical non-homologous end-joining pathway (NHEJ) resulting in VJ products(Matthews and Oettinger 2009). NHEJ is mediated by a suite of proteins that includes Ku70, Ku80, artemis, DNA-PKcs, XRCC4, DNA ligase IV, and XLF(Lieber, Yu, and Raghavan 2006).Due to absence of D segments *Igκ* locus only undergoes Vκ-Jκ rearrangement.

V(D)J rearrangement is regulated at various levels. Rearrangement only occurs in lymphoid cells and at specific stage of development, with rearrangement occurring only in pro-B (*IgH*) and pre-B (*Igκ*) stages for B cells and in pro-T (*Tcr-β*) and pre-T (*Tcr-α*) stages for T cells. In addition, there is temporal regulation, with DJ rearrangement always preceding V-DJ rearrangement for heavy chain (*IgH*) and *Tcrβ* loci. Furthermore, rearrangement shows cell cycle associated regulation, with rearrangement restricted to G_0-G_1 stage.

Recombination is organized in specialized sub-nuclear compartments called “recombination centers” (primarily the J-C cluster), that are marked by high density of H3K4me3, RAG binding and transcription factors and can provide a scaffold for chromosomal organization and locus contraction that are necessary for recombination (Ji et al. 2010, Matthews and Oettinger 2009). Presence of a conserved matrix attachment region (MAR) in J-Cκ cluster predicts that “recombination center” is directly anchored to the nuclear matrix (Yi et al. 1999). How accessible V loops and J-Cκ cluster are organized in the recombination center and whether multiple V loops compete for recombination with Jκ needs to be tested.

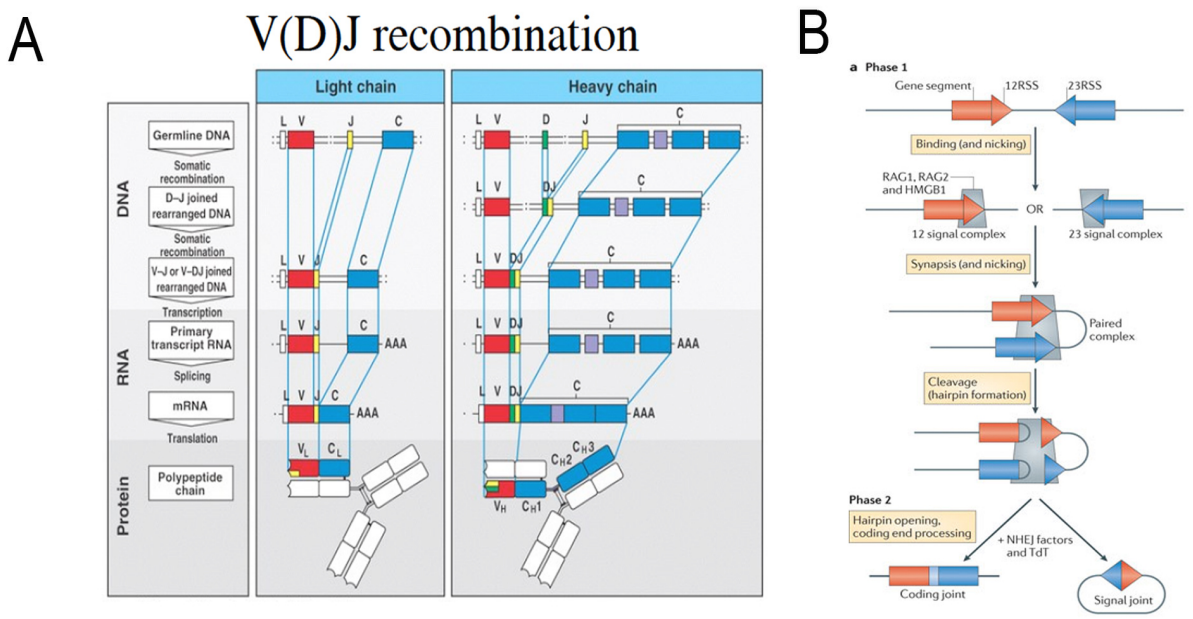


Figure 1.7. Schematics of V(D)J recombination of IgH and IgL.

(A) BCR is made of IgH and IgL chains that undergo V(D)J and VJ rearrangement, respectively. Rearrangement is initiated between D and J segments before V-DJ segments in IgH locus, and later between V and J segments in IgL locus. Transcription of rearranged locus is initiated from promoter of chosen V segment and spans chosen D-J segment and the constant region(s). RNA intervening V(D)J and C region is then spliced to make IgH and IgL polypeptide chains, which then pair up to make the BCR (adapted from Immnobiology 2005 edition).

(B) Antigen receptor gene segments are flanked by recombination signal sequences called heptamer and nonamer, which are separated by either 12bp or 23bp spacers. The RSS sequences are therefore also called 12RSS and 23RSS. In the first phase, RAG1 and RAG2 together with chromatin remodelers including HBGB1 bind to the 12RSS or the 23RSS. Capture of second RSS results in the formation of the paired complex called synapse. At this synapse, RAGs make double stranded breaks between the gene segments and the RSSs. In the second phase, non-homologous end joining (NHEJ) DNA repair factors rejoin the DNA ends. Repair typically includes non-templated nucleotide addition by terminal deoxynucleotidyl transferase (TdT) and nucleotide loss before being joined to form the coding joint. RSS ends then join to form the signal joint (adapted from Schatz and Ji. 2011).

Vκ usage

High diversity of antigen receptors is expected based on stochastic combinatorial process of V,D and J joining. However, for *Igκ*, Vκ and Jκ show bias both in their pre-selection repertoire in the bone marrow and post selection repertoire in the periphery (Aoki-Ota et al. 2012). Vκ usage in C57BL/6 mice was determined by Nemazee's group using 5' coding end amplification (5' RACE), whereby two invariant primers, one that bound to Cκ in the 3' end and another that bound to universal linker adjoined to the 5' end of Vκ transcripts were used to amplify all possible rearranged products. Of 101 functional Vκs from BM, lymph node and spleen samples combined, strikingly, only seven Vκ gene segments (1-135, 9-120,10-96,19-93,6-23,6-17 and 6-15) were found to make up ~ 40% of the bone marrow repertoire, each used at 5%-7% frequency. Most Vκ gene segments used were either clustered along the distal (1-135, 9-120, 19-93) end or the proximal (6-17 and 6-15) end.

Interestingly, major Vκs used in bone marrow (BM) also dominated in spleen and lymph node B cells, and overall usage pattern remained similar. There were slight changes, with frequencies going down for one Vκ 1-135, and going up for others such as 10-96 and 6-15. The change in frequencies might be a result of negative selection for 1-135, and positive selection for 10-96 and 6-15. These highly represented Vκs together with other low, but consistent expressers (twenty total) made up 72% of the sample overall.

For Jκ, primary rearrangements are favored at Jκ1 and Jκ2, whereas secondary rearrangements are favored at Jκ4 and Jκ5 (Max, Seidman, and Leder 1979).These data suggest

that the primary Vκ-Jκ rearrangement is highly skewed than originally thought, and therefore Vκ usage may not be completely stochastic. It is likely that positioning of V loops with respect to recombination center determines which V segments is used for recombination. Based on skewed *Igκ* repertoire, the seven highly used Vs must be aptly placed to favor formation of recombination synapse. It is possible that multiple highly used Vs are accessible per cell and that monogenic choice is dependent upon which V ends up making the synapse.

Allelic exclusion of *Igκ*

AgRs can undergo allelic exclusion at initiation or feedback inhibition level to ensure their monoallelic expression. *IgH* chain has been reported to undergo allelic exclusion primarily via feedback mechanism, where expression of a successful pre-BCR leads to decontraction of the uncombined *IgH* allele and silencing of sense and anti-sense GLT from Eμ enhancer(Melchers 2005). In the context of *Igκ*, silencing of Vκ GLT transcription and locus contraction of second allele by feedback inhibition does occur, however, is usually overridden by need for secondary rearrangements on the other *Igκ* allele or *Igλ* alleles due to high instances of auto-reactive BCR expression. In contrast, in cases of successful primary rearrangements, the excluded allele (primarily J-Cκ cluster) shows reduced deposition of RAGs compared to the chosen allele, remains methylated and associates with heterochromatin in an Ikaros, ATM and 3'E –dependent manner (Mostoslavsky et al. 1998, Goldmit et al. 2005, Hewitt et al. 2009). The chosen allele on the other hand is packaged with activating histone marks, such as H3K4me3 and H3Ac, associates with euchromatin and shows preferential binding of RAGs (Goldmit et al. 2005, Farago et al. 2012). However, these studies have their own caveats as they failed to discriminate

between asymmetric initiation from feedback inhibition, and therefore which of these mechanisms dominate is unclear.

Recently, asynchronous replication has been proposed as a possible determinant of *Igκ* allelic choice and signifies allelic exclusion at the initiation level. Asynchronous replication has also been demonstrated for *Igλ*, *IgH* and *Tcrβ* chain (Skok et al. 2001, Mostoslavsky et al. 2001) as well as imprinted genes, OR genes and X-chromosome genes(Kitsberg et al. 1993, Chess et al. 1994, Takagi 1974), suggesting that asynchronous replication is primarily true of genes that have potential for monoallelic expression. In the context of *Igκ*, asynchronous replication has been studied in relation to recombination, showing that early replicating allele is almost always fated for recombination (Farago et al. 2012), which suggests that something about the early replicating allele makes it preferable for recombination. However, how asynchronous replication relates to allelic choice imposed by epigenetic modifications and germline transcription of *Igκ,* and which of the three mechanisms is the dominant mechanism mediating allelic choice and exclusion in not known

Cell cycle

Early replicating allele associated with monoallelically active and recombining allele in autosomal and immune cells, respectively, suggests an intricate relationship between cell cycle, transcription and recombination. In both of these cell types, surface receptors activated by mitogenic signals initiate proliferation by upregulating D-type cyclins that include D1, D2 and D3 cyclins. In quiescent state G_0, when cyclins are not expressed, the retinoblastoma (Rb) family of transcriptional repressors, including the pRb, p107, and p130 proteins, repress transcription factor E2F that is responsible for expression of positive cell cycle regulators. Rb family members mediate E2F inhibition in their hypophosphorylated state. However, upon phosphorylation by the cyclin D–cyclin dependent kinase 4/6 (Cdk 4/6) complex in G_1, Rb family members are inactivated, leading to release of E2F and expression of cell cycle genes. G_1 exit and entry into S phase thereafter is mediated by activation of cdk2 by cyclin E (Fig 1.8)(Coleman, Marshall, and Olson 2004). Replication of DNA followed by mitosis then makes one complete cycle.

The role of D type cyclins in mediating entry into active G_1 phase from quiescent G_0 phase, makes them important initiators of cell proliferation. During hematopoiesis, all D-cyclins are expressed, albeit at different levels(Passegue et al. 2005). Mice deficient for a single D-cyclin do not have drastic hematopoietic defects, suggesting that there is functional redundancy among different cyclins (Ciemerych et al. 2002). However, mice deficient in all three D-cyclins are embryonically lethal, that manifests with heart defects and hematopoietic failure showing significant reduction in peripheral red blood cell numbers (Kozar et al. 2004).

Absence of cyclin D3 has the most deleterious effects on B and T lymphopoiesis. *Ccnd3*$^{-/-}$ mice show specific developmental arrest at pro-B during B lymphopoiesis and at DN3 during T lymphopoiesis (Cooper et al. 2006, Sicinska et al. 2003). Although both cyclin D2 and cyclin D3 are expressed in B cell progenitors, only cyclin D3 is required for early B cell development and for the proliferation of pro-B and pre-B cells (Cooper et al. 2006). Cyclin D3 shows preferential localization within the pro-B nuclei and binds to CDK4/6, showing minimal overlap with cyclin D2(Powers et al. 2012).

In a previous study from our own lab, we uncovered a specific relationship between transcription and cell cycle during hematopoiesis, where we showed that in addition to cell cycle regulation, cyclin D3 had specific role in transcriptional regulation of genes in pro and small pre-B cells (Powers et al. 2012). Interestingly, this function was mediated by a large fraction of cyclin D3 that was bound to the nuclear matrix (see section 1.14). This was in contrast to non-hematopoietic MEFs (mouse embryonic fibroblasts), which had substantially small fraction bound to the nuclear matrix. Indeed, cyclin D3 was involved in repression of ~250 genes in *WT* pro-B cells, of which Vκ gene segments were shown to be repressed by nuclear matrix bound fraction of cyclin D3 (Powers et al. 2012).

Similar to transcription, cell cycle is also intimately tied to recombination, whereby, DNA replication (S phase) and recombination are strictly uncoupled. This uncoupling is required to maintain genomic stability of proliferating cells. For example, RAG-mediated DNA cleavage of V-J gene segments happens at G_0-G_1 transition, and the signal ends produced by this cleavage are already repaired by components of NHEJ DNA repair pathway, in G_1-S transition, prior to

DNA replication (Fig 1.7) (Li et al. 1996, Lin and Desiderio 1993, Jiang et al. 2005). Furthermore, during G_1-S transition, RAG2 is phosphorylated by cyclin A-Cdk2 complexes and targeted for polyubiquitilation by Skp2-SCF ubiquitin ligase and proteosomal degradation (Fig 1.9) to prevent further cleavage. In absence of NHEJ and its components such as XRCC4 and the proapoptotic protein p53 or Ku80, DNA breaks persist throughout recombination and replication stages, and manifest as frequent *Igh* and *Myc* chromosomal translocations (Gao et al. 2000, Zhu et al. 2002, Difilippantonio et al. 2000).

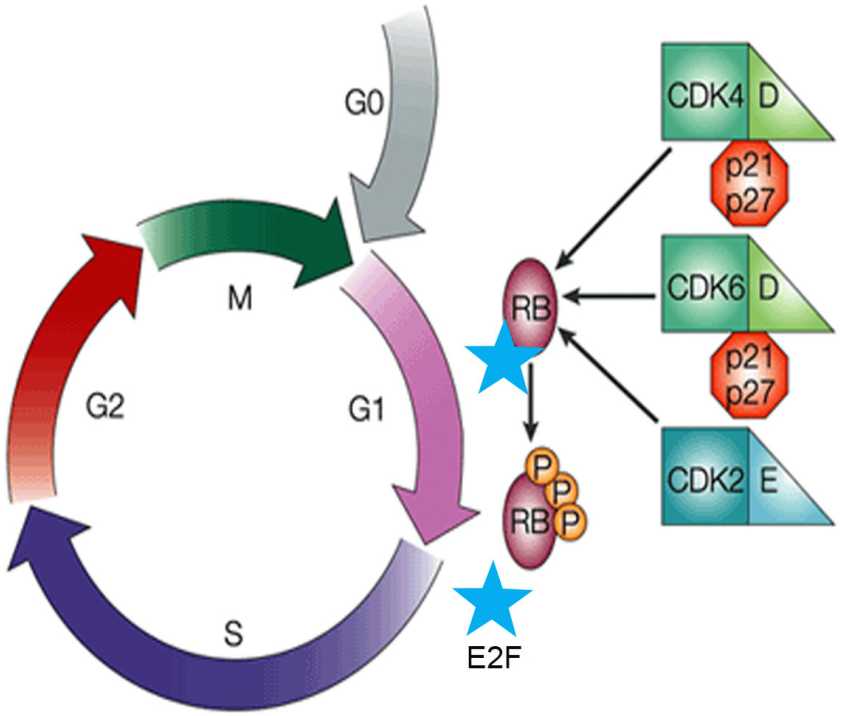


Figure 1.8 Cell cycle and cyclin D3.

Cell cycle is divided into four major stages, DNA synthesis (S) phase and a mitiotic (M) phase separated by two gap (G_1 and G_2) phases. G_1, D and G_2 phases are together called interphase. In mammalian cells different cyclin-cdk (cyclin dependent kinases) regulate progression through cell cycle. The concentration of cyclins oscillates during cell cycle, and whereas that of cdks remain constant and exceed cyclin levels. G_1/S associated D-type cyclins activate cdk4 and 6, which in turn phosphorylate retinoblastoma protein (Rb)-E2F transcription factor complex. Rb phosphorylation relives E2F, which then drives transcription of several cell cycle associated genes. In the meantime, activated cdks are bound by inhibitors such as p21 and p27 that inactivate them, thereby pushing cell cycle progression forward. Transition from G_1 to S phase is then mediated by activation of cdk2 by cyclin E. During S phase, all chromosomes are replicated. Following S phase, cells enter G2 phase, where proteins necessary for miosis such as microtubules are synthesized. Finally, replicated chromosomes separate in M phase, completing the cycle. Cell that do not enter cell cycle remain dormant in G_0 phase (adapted and modified from Coleman et al. 2004).

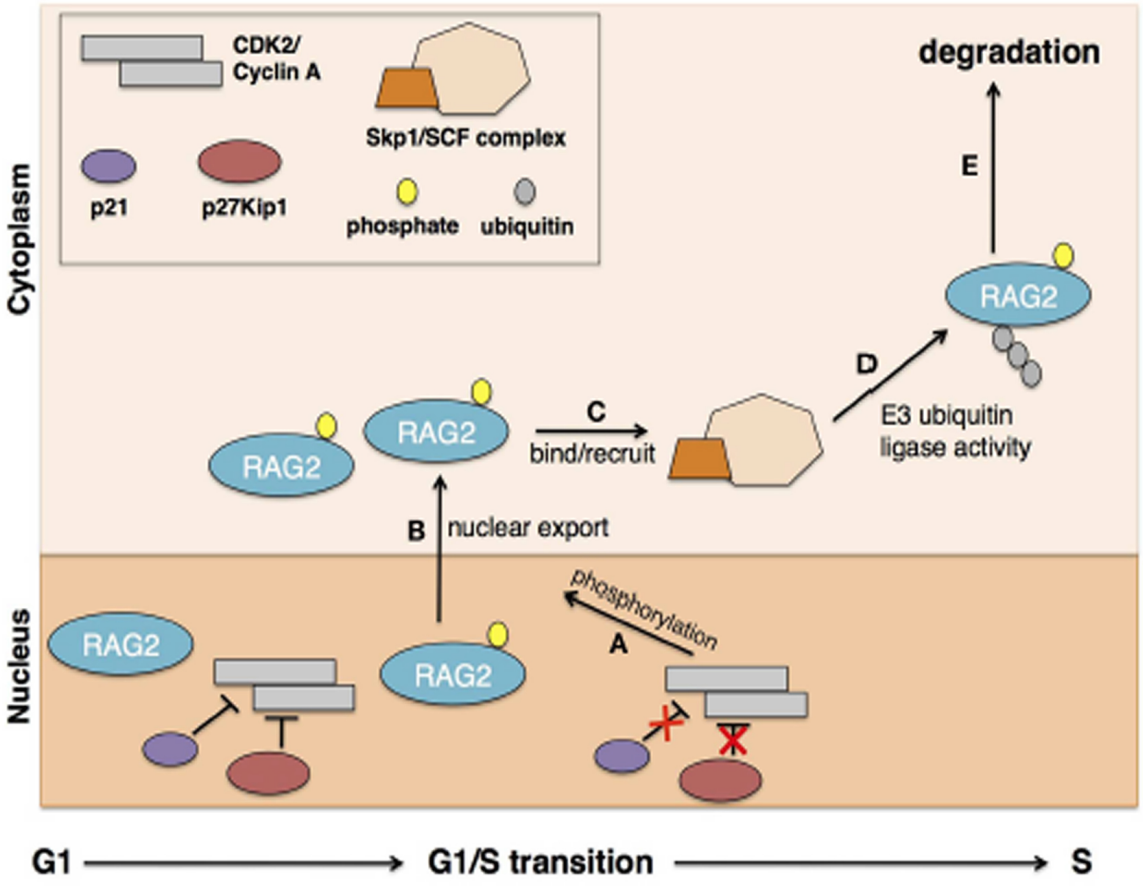


Figure 1.9 Cell cycle and recombination.

DNA replication and recombination are mutually excluded by regulating the levels of RAG2. When cell cycle transitions into synthesis phase, RAG2 proteins are phosphorylated by cyclin A-cdk2 complexes, which marks them for transport to cytoplasm. When in cytoplasms they get ubiquitinilated by Skp1/SCF complex and subsequently targeted for degradation (adapted from Jiang et al. 2005).

Nuclear matrix and associated proteins

As mentioned above, a large fraction of cyclin D3 is bound to the nuclear matrix (nm). Nuclear matrix is the structural framework of the nucleus that consists of peripheral lamins, protein complexes, an internal ribonucleic protein network and nucleoli (Fig 1.10) (Berezney and Coffey 1974, Zink, Fischer, and Nickerson 2004). It contains ~10% of nuclear proteins and is largely devoid of lipids, DNA, and histones (Fey et al. 1991). Most of the nuclear matrix-associated proteins (NMP) identified to date are common to all cell types, but several are also tissue and cell type specific (Getzenberg 1994) . Nuclear matrix is resistant to high –salt extraction and detergent washes (Wilson and Coverley 2013) and provides anchor to the DNA, as well as a platform for transcription, DNA replication, splicing and chromatin remodeling (Jackson and Cook 1986, 1985, Reyes, Muchardt, and Yaniv 1997, Zeitlin et al. 1987).

The role of nuclear matrix in DNA anchoring and organization has been well studied. Eukaryotic DNA in its linear form stretches across 2 meter distance, however, is repackaged into a 10nm fiber and further to accommodate within 6-10μm diameter nucleus. The 10nm fiber is further organized as domains as evident by frequent intra-chromosomal interactions demonstrated by FISH and chromosome conformation capture assays (Fudenberg and Mirny 2012). These chromosome loops/domains periodically attach to the matrix via attachment elements called matrix attachment regions (MARs) or scaffold attachment regions (S/MARs), and show overrepresentation of As and Ts(Fudenberg and Mirny 2012). These elements are estimated to be 100-1000 bp in size. S/MARs that anchor constitutively have been suggested to have a maintenance role in chromatin organization. Consistent with their constitutive nature, a

number of proteins that bind S/MARs called MAR binding proteins such as matrins, lamins and topoisomerase II and high-mobility group proteins make the core structural unit of the nuclear matrix(Albrethsen, Knol, and Jimenez 2009). Other MARs appear to be transient or facultative in their attachment to the matrix(Wang et al. 2010). Examples of facultative MAR binding proteins include Scaffold Attachment Factors A (SAF-A) and B and SATB1(Fudenberg and Mirny 2012).

The MAR near Eκi has been proposed to anchor *Igκ* locus to the nuclear matrix and facilitate Vκ and Jκ loop organization, thereby increasing efficiency of recombination. *Igκ* MAR also appears to control stage-specific recombination, since mice with specific deletion of MAR show pre-mature *Igκ* recombination in pro-B cells(Yi et al. 1999). Based on the role of nuclear matrix and associated proteins in providing scaffold for DNA anchor and chromatin organization, we speculate that nm-cyclin D3 in cycling cells is involved in creating nuclear scaffold or niches that are repressive.

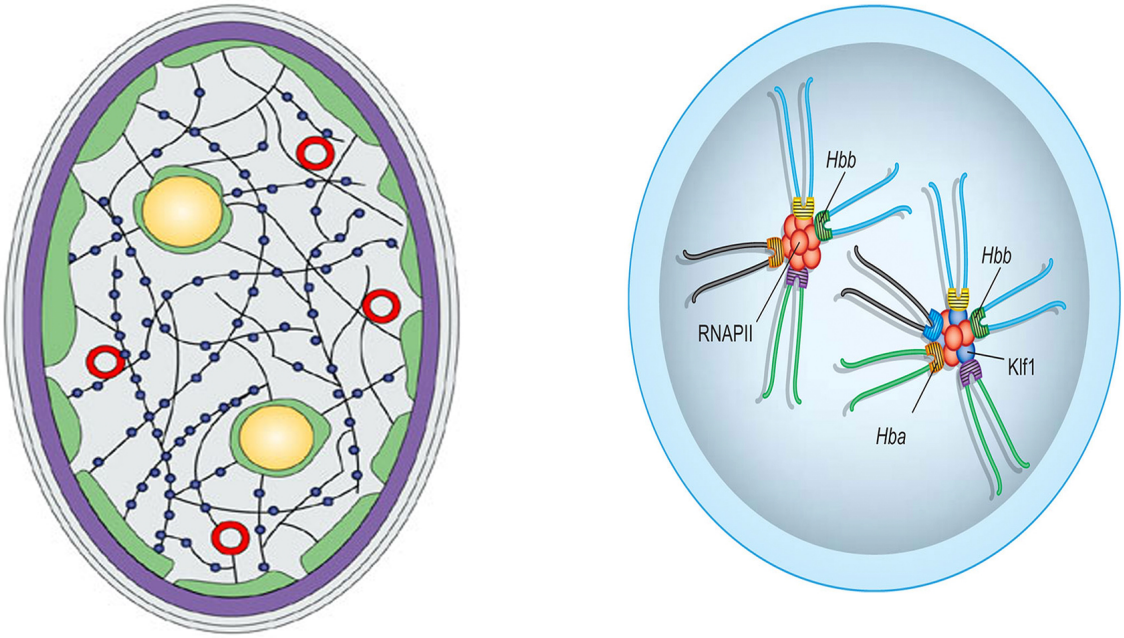


Figure 1.10 Nuclear matrix and transcription factory.

(A) Nuclear matrix (black network) binds matrix associated proteins (blue circles) and participates in spatial organization of chromatin. Nuclear matrix comprises of structural proteins such as matrins, lamins, hnRNA particles. Lamins (purple) and heterochromatin (green) in particular are directly associated with the matrix (adapted from Zink et al. 2004).

(B) Transcription factory comprises of cluster of RNA Polymerase II (orange) and transcription factors (dark green, purple, yellow), where genes (blue) are sequestered for transcription (adapted from Ragoczy et al. 2010).

Transcription factory model

There is evidence that like cyclin D3, RNA polymerases also directly associate with the nuclear matrix (Jackson et al. 1998). When agarose-embedded nuclei are subjected to restriction enzyme mediated chromatin digestion and removal, more that 90% of nascent transcripts, the DNA template and RNA polymerases still retain inside the nucleus, suggesting that the transcription machinery is attached to the matrix, and DNA is what moves and associates with these transcription foci(Papantonis and Cook 2011). These transcription machineries can be seen at 300-500 discrete sites in the nucleus, and are called "Transcription Factories" (Fig 1.8) (Jackson et al. 1993). By labeling both nascent transcripts and RNA Polymerases by different sized gold particles, transcription factories can be seen as organized clusters, typically made of 8 polymerases, which colocalize with nascent transcripts on a one-to-one basis (Iborra et al. 1996). Nuclear run-on assays using polymerases in the nuclear matrix-bound fraction show that they still retain activity after DNase I treatment. Components of the factories have been identified by subjecting this insoluble fraction to caspases, and performing mass-spectrometry of the released components. This method showed that transcription factories are more complex and comprise of co-activators and chromatin remodelers, transcription factors, histone modifying enzymes, RNPs, splicing and processing factors and RNA helicases (Melnik et al. 2011).

Transcription factories have also been proposed to organize nuclear and chromatin structures directly. As a nuclear organizer, transcription factories provide a site of transcription for co-regulated genes. Schoenfelder et al observed that in mouse erythroid cells, genes that are co-regulated associated with the same transcription factory (Schoenfelder et al. 2010, Ragoczy and Groudine 2010). For example, transcription factories with erythroid specific genes, such as

Hbb, Hba, and *Epb4.9* were enriched with erythroid transcription factor, KLF1. However, specialization of transcription factories does not appear to be general rule.

What is still debatable is whether transcription factories are stable structures or whether they assemble *de novo* in response to transcriptional needs of a cell. However, there is more evidence for stable or pre-assembled transcription factory model than *de novo* assembly model. For example, no additional transcription factories are formed after 15 min of pulse-labeling of nascent RNA, (Iborra et al. 1996). Similarly, poised (phosphorylated at Ser 5 only) factories assemble even prior to transcriptional activation of an inducible gene *uPA*, and continue to exist in the absence of transcription(Ferrai et al. 2010). However, there is also evidence for slow disintegration of factories over time, as seen by decline in the size of individual factories in the absence of transcription, which suggest sustained expression may be necessary for their matrix localization.

For a long time, transcription factories were dismissed as being fixation artifacts. However, recently "transcription factories" were reported in live cells by using fluorescently tagged CDK9, a kinase known to specifically phosphorylate and associate with elongating RNA Pol II (Ghamari et al. 2013), which has refuted these theories. In the context of *Igκ*, transcription factories in association with nm-cyclin D3 may have direct role in transcriptional regulation of Vκ gene segments (Powers et al. 2012). It is possible that nm- cyclin D3 regulates Vκ germline transcription by preventing access of Vκ loops to these organized centers designated for transcription. Immobile nature of nuclear matrix and associated nm-cyclin D3 and e-pol II

proteins, implies that these proteins are involved in creating nuclear niches that can affect transcription of mobile V loops, their contraction towards Jκ and recombination all together.

Ground State Depletion

Nuclear matrix bound proteins have been visualized by confocal imaging, and recently by super-resolution imaging. To visualize nm-cyclin D3 and RNA-Pol II at high resolution, we employed single molecule localization microscopy called ground state depletion (GSD). In this microscopy, fluorophores that appear overlapping by confocal microscopy are temporally "separated" to allow high precision detection of single molecules and beat the diffraction limit. Briefly, fluorophores are excited with high power lasers to transfer them into long-lived “off state”.. Single fluorophore molecule in the off state is then allowed to return to ground state, and emit photons. The photons are recorded, their position determined and final super-resolution image reconstructed. GSD provides 20nm resolution at XY and 50nm resolution in the Z axis. (Fig 1.11).

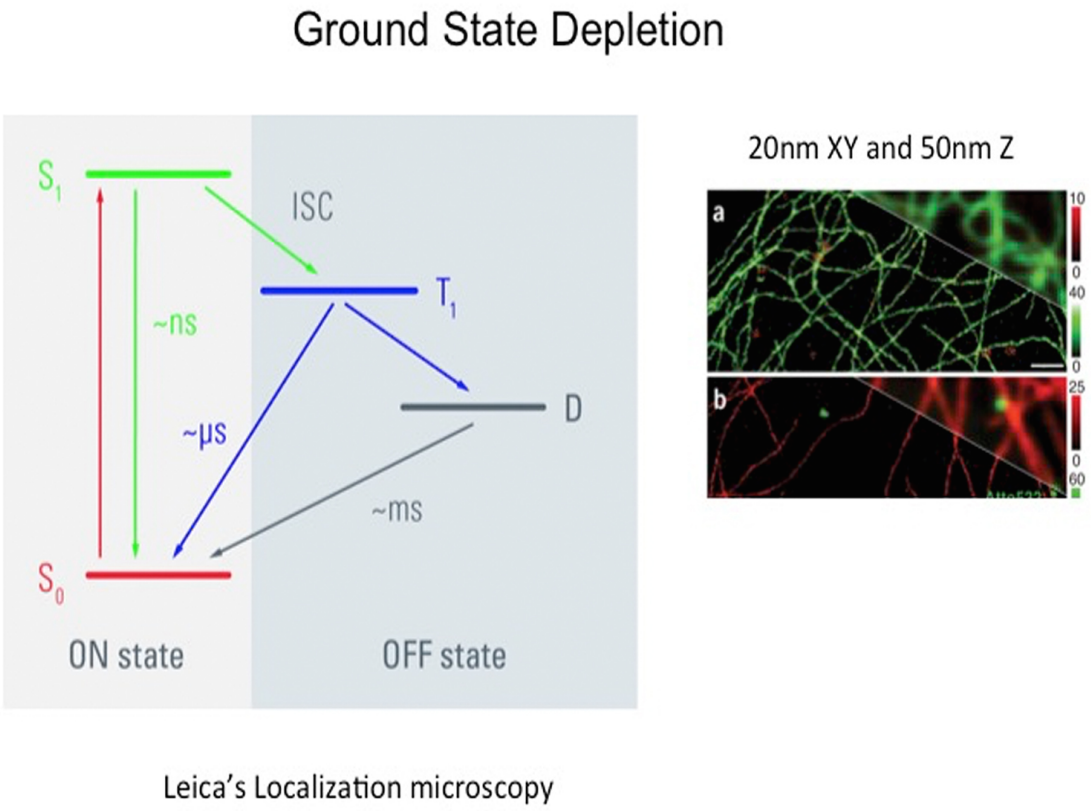


Figure 1.11 Schematic representation of Ground State Depletion.

When fluorophores in ground state S_0 (ON state) are excited they enter the excited state S_1 (ON) and oscillate between the two states. Alternatively, fluorophores from S_0 state may enter OFF states called triplet (T_1) or dark state (D). Unlike fluorophores in S_1 state, those in the triplet or dark state are unable to emit light. When low intensit light is used to excite the fluorophores, only few fluorophores enter the dark state, with majority entering the excited state S_1 and returning to ground state. However, when high intensity light is used to excite the fluorophores, more number of fluorophores reach the dark state. These fluorophores in dark state can be forced to return to ground state and emit photons, by using right embedding medium and choice of fluorophores. When enough molecules are in the OFF states, it is possible to detect individual molecules return that can be captured as an image (adapted http://www.leica-microsystems.com/science-lab/super-resolution-gsdim-microscopy/).

ViewRNA *in situ* Hybridization

To delve into how nm-cyclin D3 affects Vκ transcription, we used Affymetrix based RNA detection assay called ViewRNA ISH Cell Assay. This assay uses a direct fluorescence RNA *in situ* hybridization method in combination with branch DNA signal amplification (bDNA) for single molecule detection of RNA signal. Target specific RNA probes are designed with specific oligonucleotides that act as adaptor for branch amplification, The pre-amplifier oligonucleotide pairs with its cognate RNA probe-adapter complex, which then becomes adapter for the amplifier oligonucleotide. The branched chain formed by probe-pre-amplifier and amplifier is then bound by label probe oligonucleotide attached to fluorophore, that then allows specific detection of the RNA target. In a given cell, upto 4 RNA targets can be detected, made possible by uniqueness of each RNA probe and the attached adapter conjugated to different fluorophores. This branching creates 400 binding sites for each label probe, thereby providing 8000 fold amplification for one transcript. Hybridized RNA can be visualized using a standard fluorescence microscope with the corresponding filter sets (EX: 488, 550, 650 and 750nm). This technique is quite sensitive, providing single transcript sensitivity. Therefore, any visible dot corresponds to one target molecule.

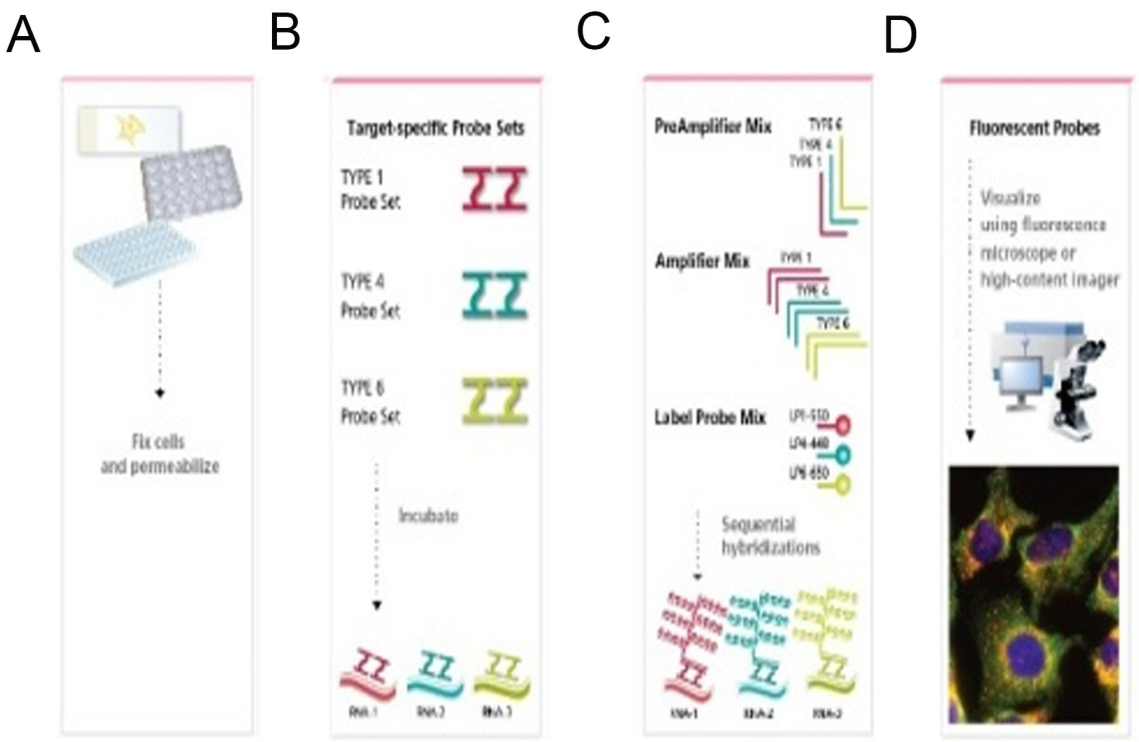


Figure 1.12 ViewRNA *in situ* Hybridization.

(A) Sample preparation: Cells are fixed and permeabilized to allow target accessibility.
(B) Target Hybridization: Probe-sets are designed to bind target RNA. Probe sets come as oligonucleotide pairs where one pair targets one RNA. Signal amplication happens individually for each target.
(C) Signal Amplification: Signal is amplified using branched chain (bDNA) technology, which is achieved via a series of sequential hybridization steps. Probe set oligonucleotides are first allowed to hybridize the target. Then PreAmplifier and Amplifier mix are added to make branched chains for each target. Label Probe mix conjugated to the fluorescent dyes are finally added to allow visualization.
(D) Fluorescent Detection: 4 filters can be used to detect the transcripts, which include 488, 550, 650 and 750, suggesting that 4 different RNA molecules can be targeted per cell.

Primary *Igκ* recombination happens in one allele, whereby a single V segment out of >100 recombines with a single J segment from the chosen allele, forming the basis of its monoallelic and monogenic expression. However, the precise mechanism of how this is achieved is not known. So far, our knowledge of *Igκ* accessibility, epigenetic modification and allelic choice was based on our understanding of J-Cκ cluster, and not Vκ that constitutes majority of the locus. Therefore, recent finding that Vκ accessibility is regulated by nm-cyclin D3, whereby downregulation or loss of cyclin D3 led to their germline expression has led us to investigate the role of Vκ accessibility in monoallelic and monogenic usage of *Igκ*.

A number of outcomes are likely from this study. Based on monoallelic *Igκ* recombination, it is likely that Vκ transcription/accessibility occurs on a single allele in which case Vκ accessibility would determine allelic choice for recombination. In addition, it is likely that a single V segment is accessible per cell, and constitutes that specific V segment that undergoes recombination. Alternatively, it is possible that multiple Vs are accessible per cell, which would imply that recombination happens as a result of competition between accessible Vs. Similarly, single or multiple Vκs may juxtapose with Jκ by locus contraction prior to recombination. At the end, all of these findings must integrate to explain pre-selection Vκ repertoire in the bone marrow. For my dissertation, I sought to test these possible scenarios on a step-by-step basis.

In this study, we first used confocal imaging as well as Leica' GSD to visualize nuclear matrix-bound cyclin D3 and e-Pol II or transcription factory complexes in *WT* pro-B and small pre-B cells to understand how positioning of nm-cyclin D3 on the matrix may affect Vκ

transcription during proliferation (pro-B) and differentiation (small pre-B) stages. We validated cyclin D3 as Vκ repressor using RNA-seq and combined DNA-RNA FISH approaches on *Ccnd3*$^{-/-}$ pro-B cells. We then investigated how Vκ germline transcription seen in *Ccnd3*$^{-/-}$ pro-B cells relates to Vκ transcription in *WT* C57BL/6 (B6) mice, using DNA FISH in small pre-B cells. We examined mono versus biallelic nature of Vκ transcription by combined DNA-RNA FISH in *Ccnd3*$^{-/-}$ pro-B and *WT* small pre-B cells, and validated the finding using single cell RNA-seq analysis of small pre-B cells from divergent F1 cross (B6X CAST/EiJ) in collaboration with Dr. Singh's lab in Cincinnati Children's Hospital. Due to analysis done at single cell level, we also tested whether single or multiple Vκ gene segments are active per cell.

We then measured spatial niche of fixed e-Pol II surrounding both *Igκ* alleles on 3D confocal images of 2-color Vκ and Jκ DNA FISH to test if nm-cyclin D3 and e-pol II created differential niche around the chosen and excluded allele. Based on the acquired measurements, we built a 2D model that takes into account spatial niche, Vκ transcription and contraction to explain allelic choice versus exclusion of *Igκ*. We then collaborated with Dinner laboratory at Uchicago and performed polymer chain simulation of *Igκ* using measured niche dimensions along with high probability CTCF sites to model *Igκ* positioning and topology that best predicts observed Vκ usage(Aoki-Ota et al. 2012).Lastly, we investigated a database of monoallelic genes, to test whether other genes dysregulated in *Ccnd3*$^{-/-}$ pro-B cells were also monoallelic, and that downregulation of cyclin D3 and entry into differentiation stage was common mechanism for monoallelic and monogenic choice.

2. Results

Nuclear matrix-bound cyclin D3 assembled with elongating RNA Pol II

To understand how matrix-bound cyclin D3 is involved in transcriptional repression of Vκ gene segments in cycling pro-B cells, we sought to visualize the positioning of matrix bound cyclin D3 relative to matrix bound transcription factories, which are enriched in e-Pol II. To enable us to visualize nuclear matrix and chromatin bound proteins while washing away soluble proteins, we used a published washing protocol that uses cytoskeletal stabilizing buffer (CSK) with 0.5% Triton (Sawasdichai et al. 2010). We first validated the protocol by washing away α-tubulin, which only has soluble fraction (Fig S2.1A). We then took *WT* pro-B cells, washed, fixed and stained for cyclin D3, pSer2 e-Pol II and known matrix bound protein special AT-rich sequence-binding protein 1 (SATB1). At 250nm resolution conferred by confocal microscopy, we found that in pro-B cells, nm- cyclin D3 either colocalized or laid in close apposition to e-Pol II, assembling over ~ 50% of active e-Pol II foci consistently throughout the nucleus (Fig 2.1A and D). Both cyclin D3 (35%) and e-Pol II (50%) also associated with SATB1, validating that both proteins are assembled on the nuclear matrix (Fig 2.1B-D, Fig S2.1B).

To be able to see structural details of cyclin D3-e-Pol II complexes assembled on the matrix, we further imaged pro-B cells using Leica's super-resolution microscopy (Leica GSD-IM), and found that even at 20-40nm resolution, e-Pol II and cyclin D3 colocalized at multiple sites throughout the nucleus, almost intercalating with each other (Fig 2.1A, Fig S2.1B, last panel). This was also true for cyclin D3-SATB1 and SATB1-Pol II staining (Fig 2.1B-C, last panel, Fig S2.1B). In both confocal and super-resolution images, we noticed that cyclin D3-e-Pol II complexes were primarily assembled on the nuclear body, and excluded the lamina, which is

usually devoid of e-Pol II foci.

We then visualized the positioning of matrix-bound D3-Pol II complexes in *WT* small pre-B cells, where Vκ is derepressed, and found that most of the cyclin D3 in the nuclear body was depleted (Fig 2.1E), and resembled *Ccnd3*$^{-/-}$ pro-B cells. Overall, these results suggest that during B cell development, cyclin D3 acts as a transcriptional repressor by assembling on a subset of e-Pol II foci on the nuclear matrix, and derepression of these genes occur when cyclin D3 is downregulated and cells exit the cell cycle.

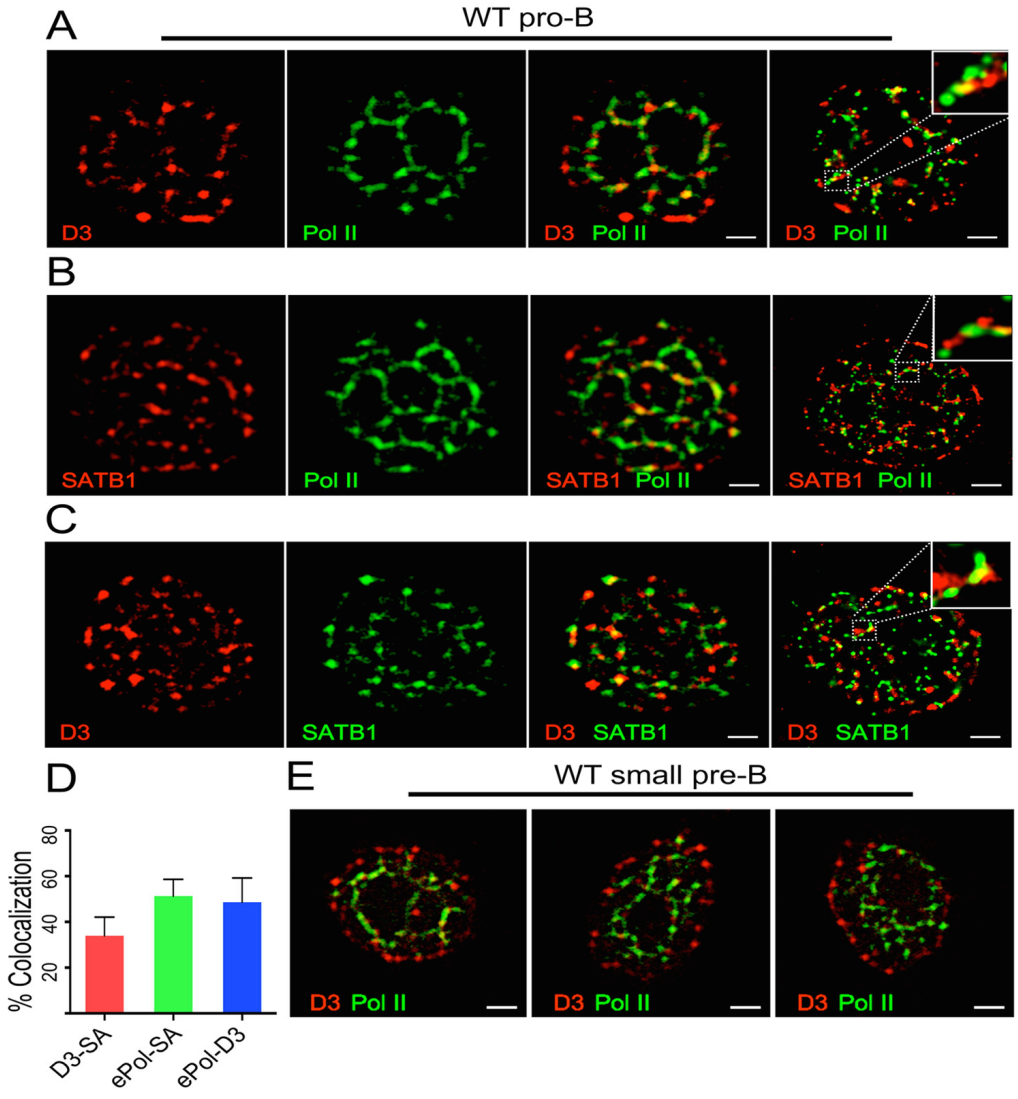


Figure 2.1 Cyclin D3 is assembled on a fraction of e-Pol II on the nuclear matrix.

(A-C) Representative confocal images (from 40, cells n=2 experiments) of *WT* pro-B cells washed 10 times (3 minutes/wash) with CSK+0.5%Triton to remove soluble nuclear proteins and then fixed and stain with antibodies specific for cyclin D3 and e-Pol II, SATB1 and e-Pol II and SATB1 and cyclin D3 (panels 1-3). Super-resolution image of similarly stained WT pro-B cells (panel 4).

(D) Percent colocalization of e-Pol II-D3, D3-SATB1 (SA) and Pol II-SATB1 (SA) stains calculated by using Manders on 30 2D confocal images per samples (n=2 experiments).

(E). Representative confocal images (40 cells, n=2 experiments) of *WT* small pre-B cells washed and stained with antibodies specific for cyclin D3 and e-Pol II (panels1-3).

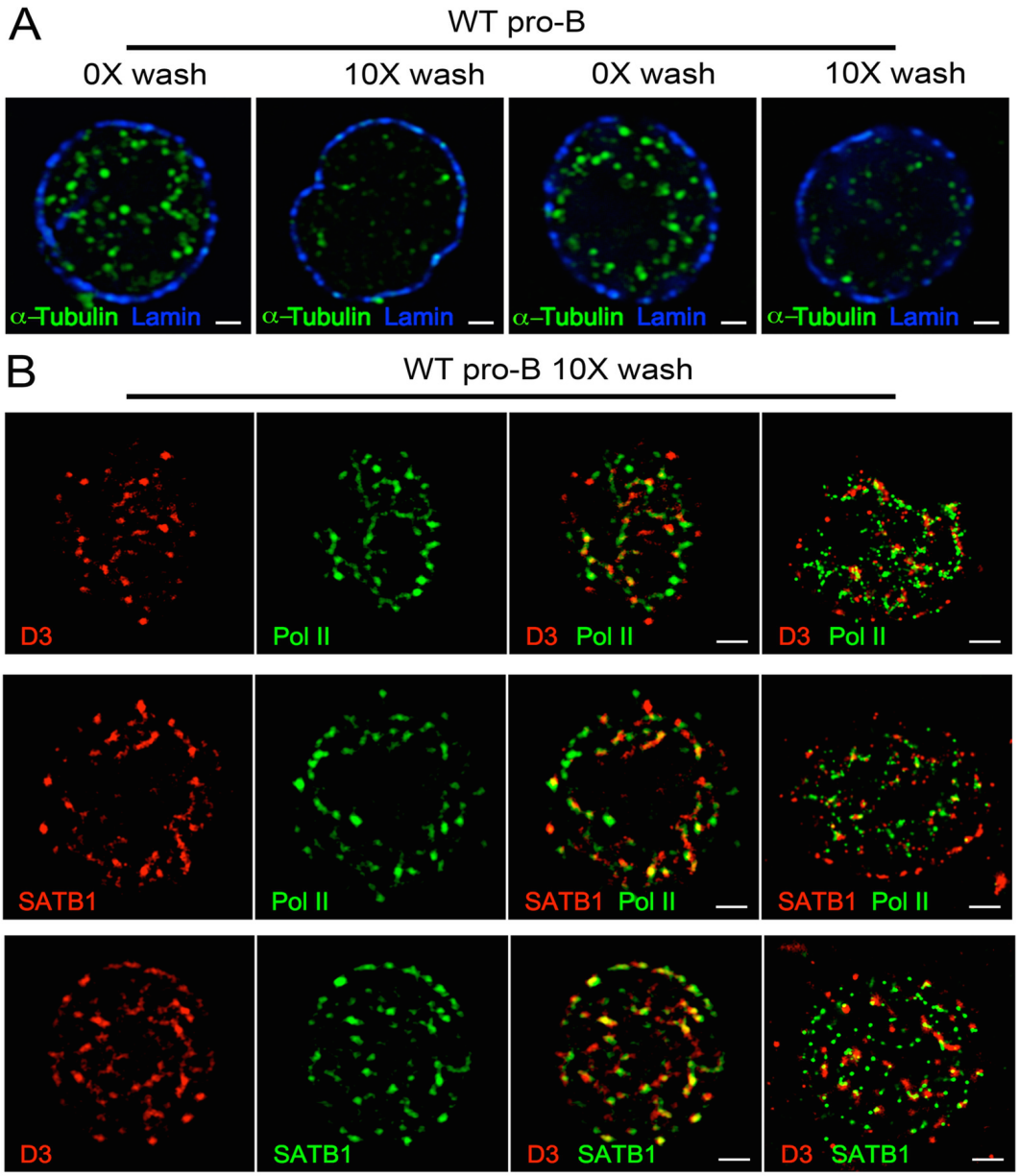


Figure S2.1 Assembly of cyclin D3 and e-Pol II on SATB1^{+} nuclear matrix.

(A) Representative confocal images of *WT* pro-B cells (40 cells, n=2 experiments) without CSK+0.5% triton washing, fixed and stained for the soluble nuclear protein a-tubulin and lamin (panel 1 and 3). Confocal images of WT pro-B cells washed 10 times (3 minutes/wash) with CSK+0.5% triton (panel 2 and 4).

(B) Representative confocal images of *WT* pro-B cells (40 cells, n=2 experiments) washed 10 times (3 minutes/wash) with CSK+0.5% triton buffer and fixed and stained for cyclin D3/ e-Pol II, SATB1/ e-Pol II and SATB1 and cyclin D3 (panel 1-3). Super-resolution image of WT pro-B cells washed and stained with cyclin D3/ e-Pol II, SATB1/ e-Pol II and SATB1 and cyclin D3 (panel 4).

Monoallelic Vκ association with e-Pol II foci regulated by cyclin D3

We then sought to directly visualize how Vκ gene segments are localized with respect to cyclin D3-Pol II repressive complexes in pro-B cells and active e-Pol II foci in small pre-B cells. To do so, we first washed *WT* or *Ccnd3*$^{-/-}$ pro-B cells with CSK+0.5% Triton buffer ten times, fixed with 2% PFA and performed Fluorescence *in situ* Hybridization (FISH) with a 488-labeled BAC probe, RP-23 182E6, that binds to a 0.2Mb region spanning ~ 10 distal V gene segments (Vκ 2-113 to 1-122), followed by staining for cyclin D3 and e-Pol II (combined immunoFISH). We found that in *WT* pro-B cells and *Rag2*$^{-/-}$ pro-B cells, both Vκ alleles were surrounded by Pol II-cyclin D3 repressive complexes and rarely colocalized (no localization >95% of times) with e-Pol II, suggesting that in presence of matrix bound cyclin D3, Vκ loops are inaccessible to the transcription machinery (Fig 2.2A,G, Fig S2.2.1A, Fig S2.2.2A).

To test earlier observation that Vκ gene segments are transcriptionally accessible in *Ccnd3*$^{-/-}$ pro-B cells, we performed combined immunoFISH in these cells, and found that indeed, in *Ccnd3*$^{-/-}$ pro-B cells, Vκ colocalized with active Pol II sites (35-40%) at a very high frequency (Fig 2.2B,G, Fig S2.2.1B). We saw transcriptional activation of multiple other V segments by bulk RNA-seq on *Ccnd3*$^{-/-}$ pro-B cells (Fig S2.2.3). Cyclin D3 absence had no effect of Jκ or Jλ transcription (Fig S2.2.3). As expected, Vκ was highly accessible in *WT* small pre-B cells, showing ~40% colocalization with e-Pol II similar to *Ccnd3*$^{-/-}$ pro-B cells (Fig 2.2C,G, Fig S2.2.1C). Remarkably, in both *Ccnd3*$^{-/-}$ pro-B cells and *WT* small pre-B cells, only single Vκ allele was found to colocalize with e-Pol II (>95%), suggesting monoallelic nature of Vκ transcription. Importantly, co-localization of Vκ to e-Pol II was specific to Vκ as Tcr Vβ was

rarely found to associate with e-Pol II in *WT* small pre-B cells (Fig S2.2.4). Interestingly, Vκ colocalization to e-Pol II occurred in early replicating *Igκ* allele (Fig S2.2.5), which is consistent with early replicating allele being more accessible and preferred for recombination (Farago et al. 2012).

In order to validate that monoallelic localization with e-Pol II is reflective of true monoallelic expression, we decided to look at transcription of germline Jκ, which has been reported to be biallelic, using a BAC probe that spans distal Jκ promoter to Cκ (RP24 387E13). We found that Jκ was transcriptionally inaccessible in *WT* pro-B, *Rag2*$^{-/-}$ pro-B and *Ccnd3*$^{-/-}$ pro-B cells (Fig 2.2D,E,H, Fig S2.2.1A-B, lower panel, Fig S2.2.2B), which goes along earlier reports, which showed that Jκ transcription is not regulated by cyclin D3, but by STAT5-E2A axis. As expected, Jκ alleles did colocalize with e-Pol II in *WT* small pre-B cells (Fig 2.2F,H, Fig S2.2.1C), which reiterates E2A-mediated activation of germline Jκ transcription in small-pre cells. Remarkably, Jκ colocalization to e-Pol II was biallelic. These results show that cyclin D3-Pol II complexes only act to regulate Vκ transcription, but not Jκ.

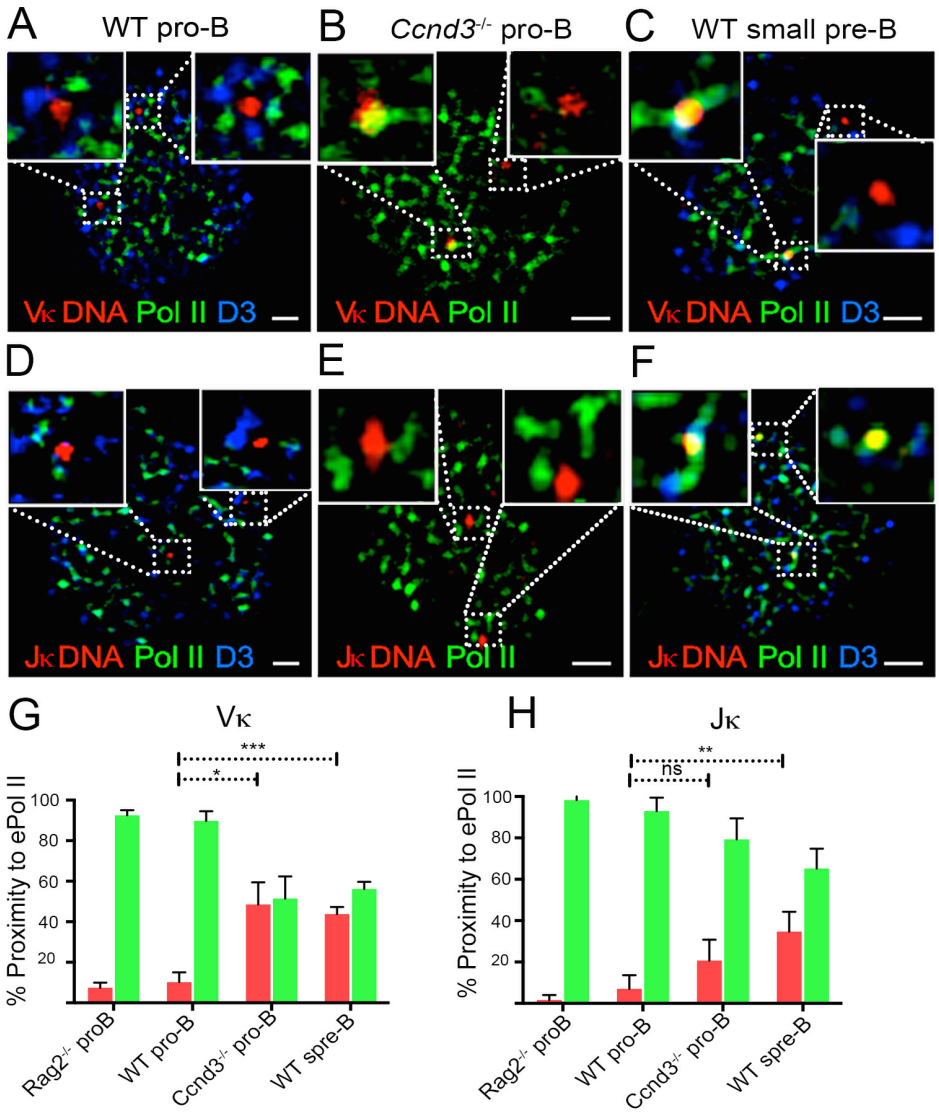


Figure 2.2 *Igκ* germline transcription depends upon loss of cyclin D3 and STAT5.

(A-C) Representative confocal image (50 cells, n=3 experiments) of *WT* pro-B cells, *Ccnd3*$^{-/-}$ pro-B cells and *WT* small pre-B washed extensively to remove soluble nuclear proteins then hybridized with Vκ DNA probe (RP23-182E6) spanning 10 distal Vκ genes (Vκ 2-113 to 1-122), and stained with antibodies specific for cyclin D3 and e-Pol II .

(D-F) Representative confocal image (50 cells, n=3 experiments) of *WT* pro-B cells, Ccnd3-/- pro-B and *WT* small pre-B washed, hybridized with Jκ DNA probe (RP24-387E13) spanning Jκ - Cκ and stained with antibodies specific for cyclin D3 and e-Pol II.

(G-H) Percent Vκ or Jκ colocalized (red) and not colocalized (green) to e-Pol II scored on confocal images of 50 nuclei per sample (n=3 experiments), and plotted for each sample. Colocalization scored when at least one Vκ allele engaged e-Pol II. Statistical significance calculated by unpaired Student's t test ($p<0.05$ *, $p<0.01$ ** and $p<0.001$ ***).

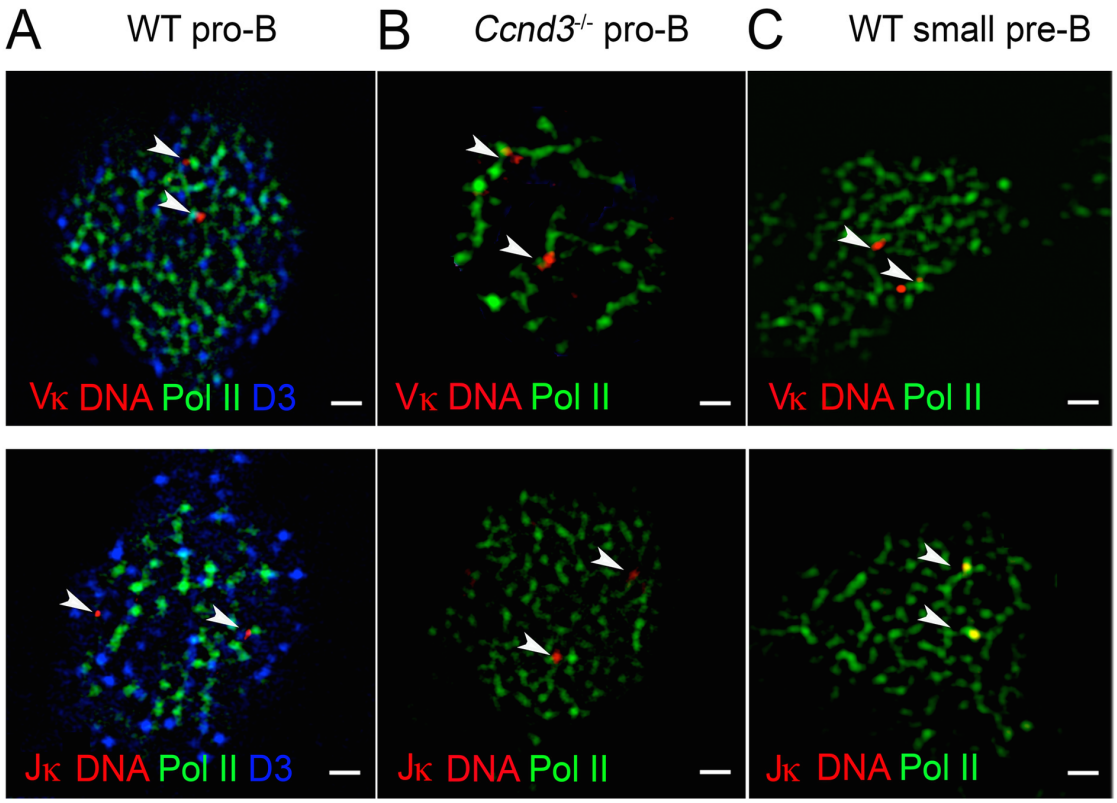


Figure S2.2.1 Mono-allelic association of Vκ with elongating RNA polymerase II rich transcription centers.

(A) Representative confocal images of *WT* pro-B cells (50 cells, n=3 experiments) washed with CSK+0.5% Triton buffer, hybridized with Vκ DNA probe (RP23-182E6: top) and Jk DNA probe (RP24- 387E13: bottom) spanning 10 distal Vκ genes, and stained with cyclin D3 and e-Pol II. Arrows show individual alleles.

(B) Representative confocal images of *Ccnd3*$^{-/-}$ pro-B cells (50 cells, n=3 experiments) washed with CSK+0.5% Triton buffer, hybridized with Vκ DNA probe (RP23-182E6: top) and Jκ DNA probe (RP24- 387E13: bottom) spanning 10 distal Vκ genes, and stained with e-Pol II. Arrows show individual alleles. Top panel demonstrates one allele has duplicated and that one of these is associated with e-Pol II.

(C) Representative confocal images of *WT* small pre-B cells (50 cells, n=3 experiments) washed with CSK+0.5% Triton buffer, hybridized with Vκ DNA probe (RP23-182E6: top) and Jκ DNA probe (RP24- 387E13: bottom) spanning 10 distal Vκ genes, and stained with e-Pol II. Arrows show individual alleles. Top panel demonstrates one allele has duplicated and that one of these is associated with e-Pol II.

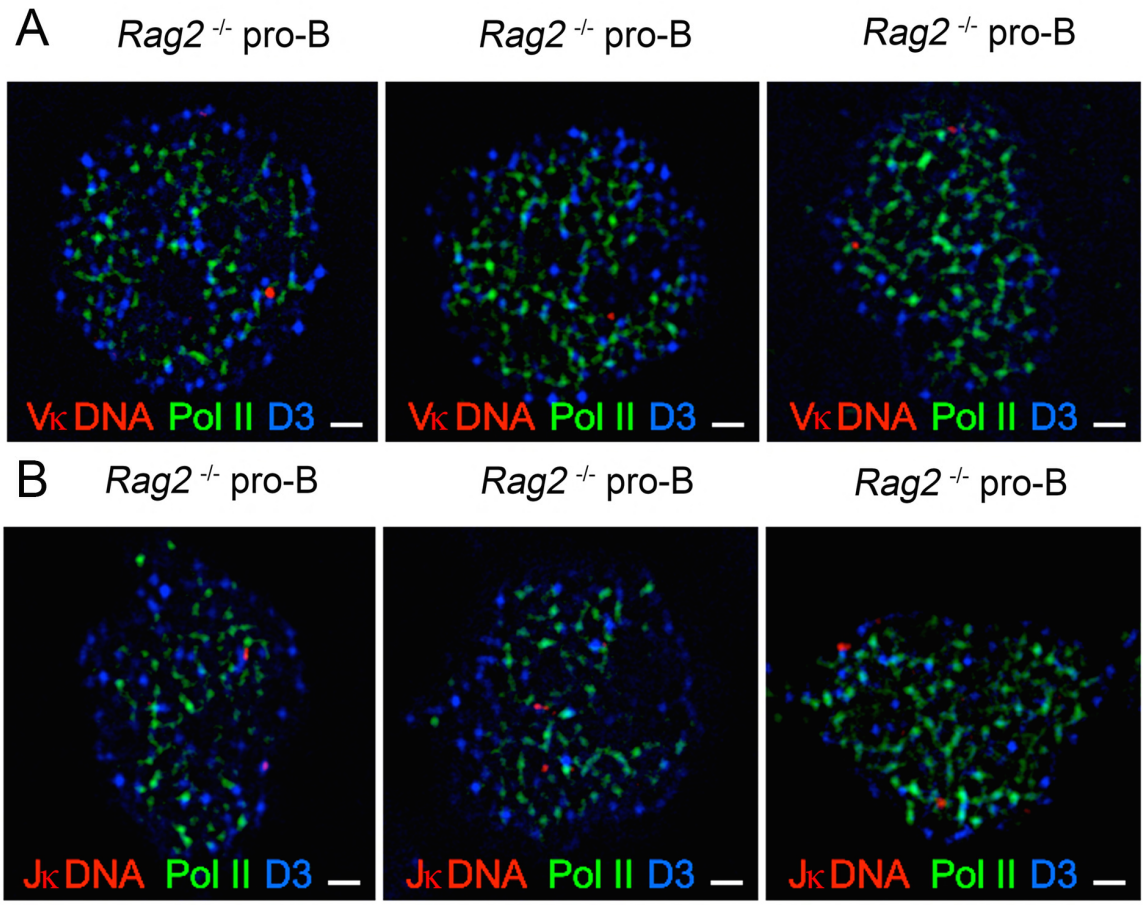


Figure S2.2.2 Cyclin D3 mediated Vκ repression is rag-independent.

(A) Representative confocal images of *WT* pro-B cells (50 cells, n=3 experiments) washed with CSK+0.5% Triton buffer, hybridized with Vκ DNA probe (RP23-182E6) spanning 10 distal Vκ genes, and stained with cyclin D3 and e-Pol II. Arrows show individual alleles.
(B) Representative confocal images of *Ccnd3*$^{-/-}$ pro-B cells (50 cells, n=3 experiments) washed with CSK+0.5% Triton buffer, hybridized with Jκ DNA probe (RP24- 387E13), and stained with e-Pol II. Arrows show individual alleles. Top panel demonstrates one allele has duplicated and that one of these is associated with e-Pol II..

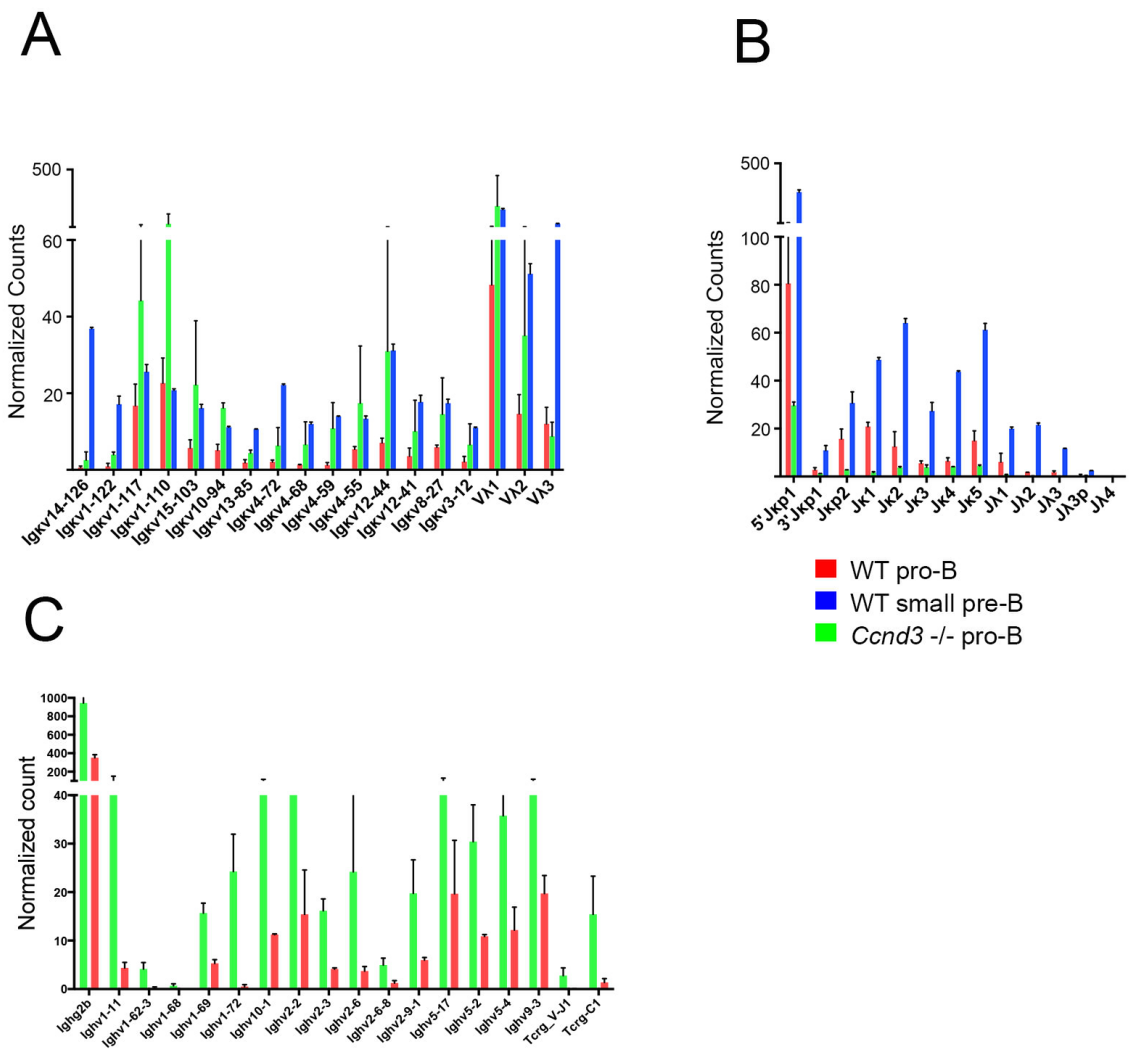


Figure S2.2.3 Vκ genes dysregulated in *Ccnd3*$^{-/-}$ pro-B by bulk RNA-seq.

(A) *Ccnd3*$^{-/-}$ pro-B (green), *WT* pro-B (red) and *WT* small pre-B (blue) pro-B RNA-seq normalized read counts, showing expression of multiple Vκ gene segments.
(B) *Ccnd3*$^{-/-}$ pro-B (green), *WT* pro-B (red) and *WT* small pre-B (blue) pro-B RNA-seq normalized read counts, showing expression of multiple Jκ-Cκ gene segments.
(C) *Ccnd3*$^{-/-}$ pro-B (green) and *WT* pro-B (red) RNA-seq normalized read counts, showing VH genes and TcrγV genes.

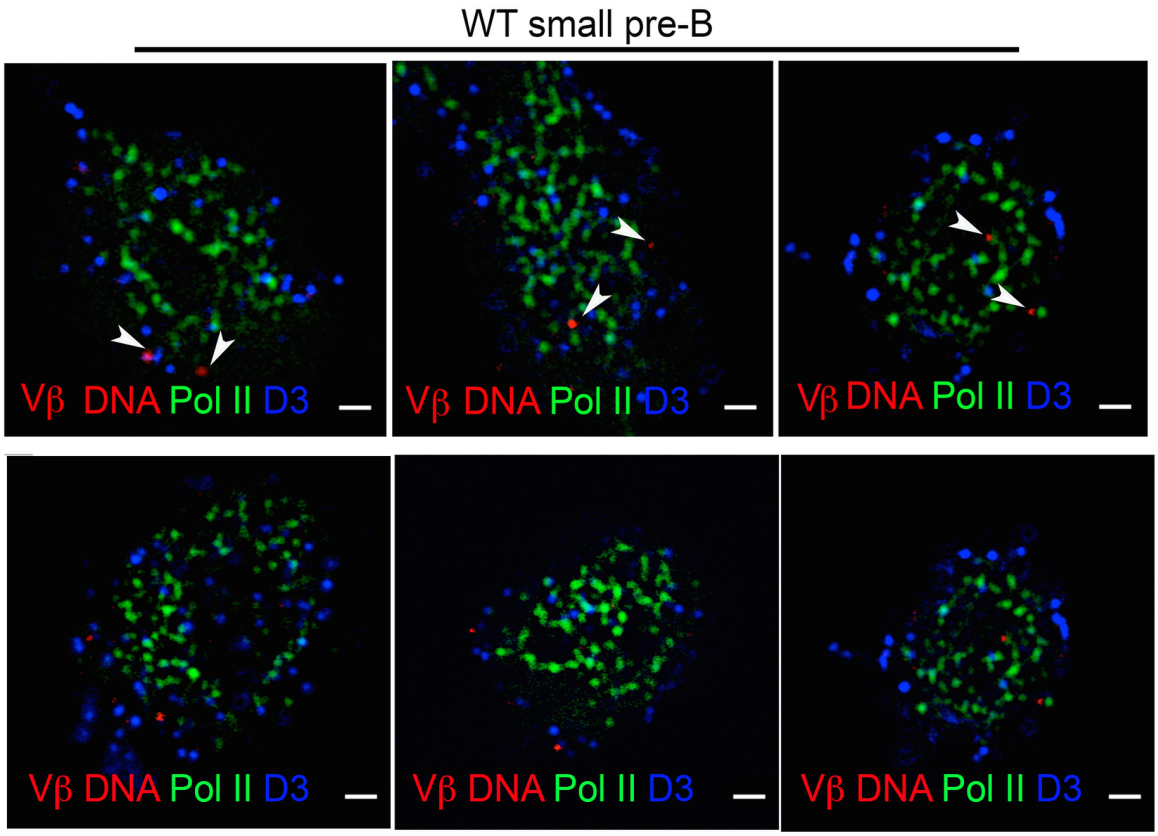


Figure S2.2.4 Vκ localization to nm e-Pol II is specific to *Igκ*.

Representative confocal images of *WT* small pre-B cells (50 cells, n=3 experiments) washed with CSK+0.5% Triton buffer, hybridized with Tcr Vκ DNA probe (RP23-184C1) and stained with cyclin D3 and e-Pol II. Arrows show individual alleles.

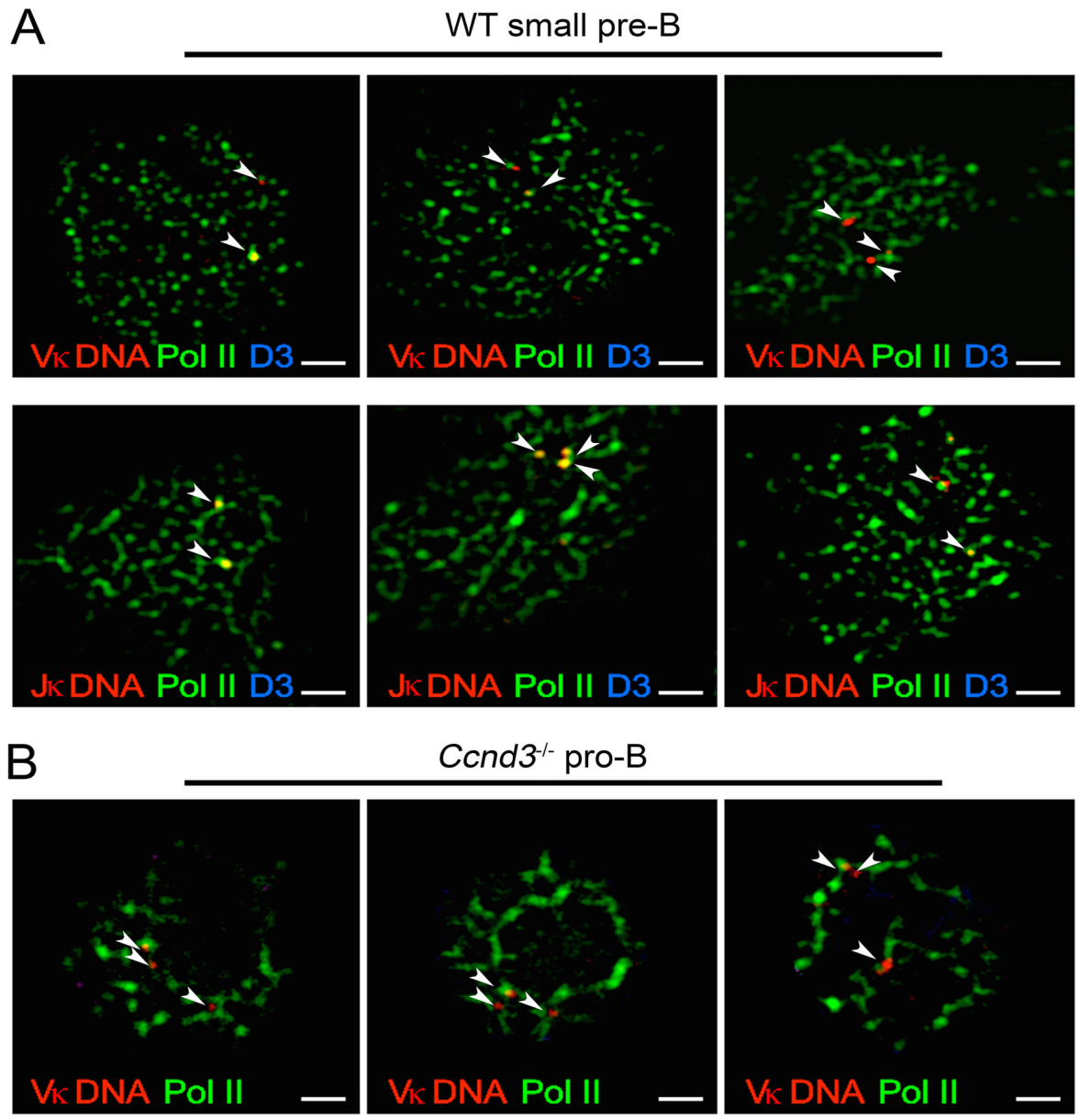


Figure S2.2.5 Early replicating allele is preferentially chosen for germline activation.

(A) Representative confocal images of *WT* small pre-B cells (50 cells, n=3 experiments) washed with CSK+0.5% Triton buffer, hybridized with Vκ DNA probe (RP23-182E6: top) and Jκ DNA probe (RP24- 387E13: bottom) spanning 10 distal Vκ genes, and stained with cyclin D3 and e-Pol II. Arrows show individual alleles.

(B) Representative confocal images of *Ccnd3*$^{-/-}$ pro-B cells (50 cells, n=3 experiments) washed with CSK+0.5% Triton buffer, hybridized with Vκ DNA probe (RP23-182E6) spanning 10 distal Vκ genes, and stained with e-Pol II. Arrows show individual alleles. Top panel demonstrates one allele has duplicated and that one of these is associated with e-Pol II.

Vκ germline transcription is monoallelic

Based on this observation, we decided to verify monoallelic transcription by using combined DNA-RNA FISH probing for a single Vκ (Vκ 1-117) transcript that lies within the region of the BAC probe. We specifically designed RNA probe (spanning RSS+3'UTR) such that it targets germline transcripts and excludes rearranged transcripts (Fig S2.3.1, Table S1). In both *Ccnd3*$^{-/-}$ pro-B and *WT* small pre-B cells, we detected a single transcript either coming right of e-Pol II or localized nearby Vκ gene/Pol II complexes (Fig 2.3A,B, Fig S2.3.2A,B). Monoallelic transcription at Vκ occurred in ~95-98% of *WT* small pre-B cells and *Ccnd3*$^{-/-}$ pro-B cells, reiterating their monoallelic nature (Fig 2.4A). We then verified germline J-Cκ transcription in *WT* small pre-B cells, using an RNA probe that binds 5' of Jκ 1 and 3' of distal promoter, therefore only targeting germline transcript (Fig 2.3C, Fig S2.3.2C). Biallelic transcription at Jκ was noted in nearly 40% of the cells that had transcripts (Fig 2.4A). The rest of the cells had transcripts but not in the proximity of the Jκ gene, therefore, were unassignable as mono-versus biallelic. These data confirm that at least at the level of single Vκ segment, germline transcription is monoallelic, as opposed to biallelic for Jκ.

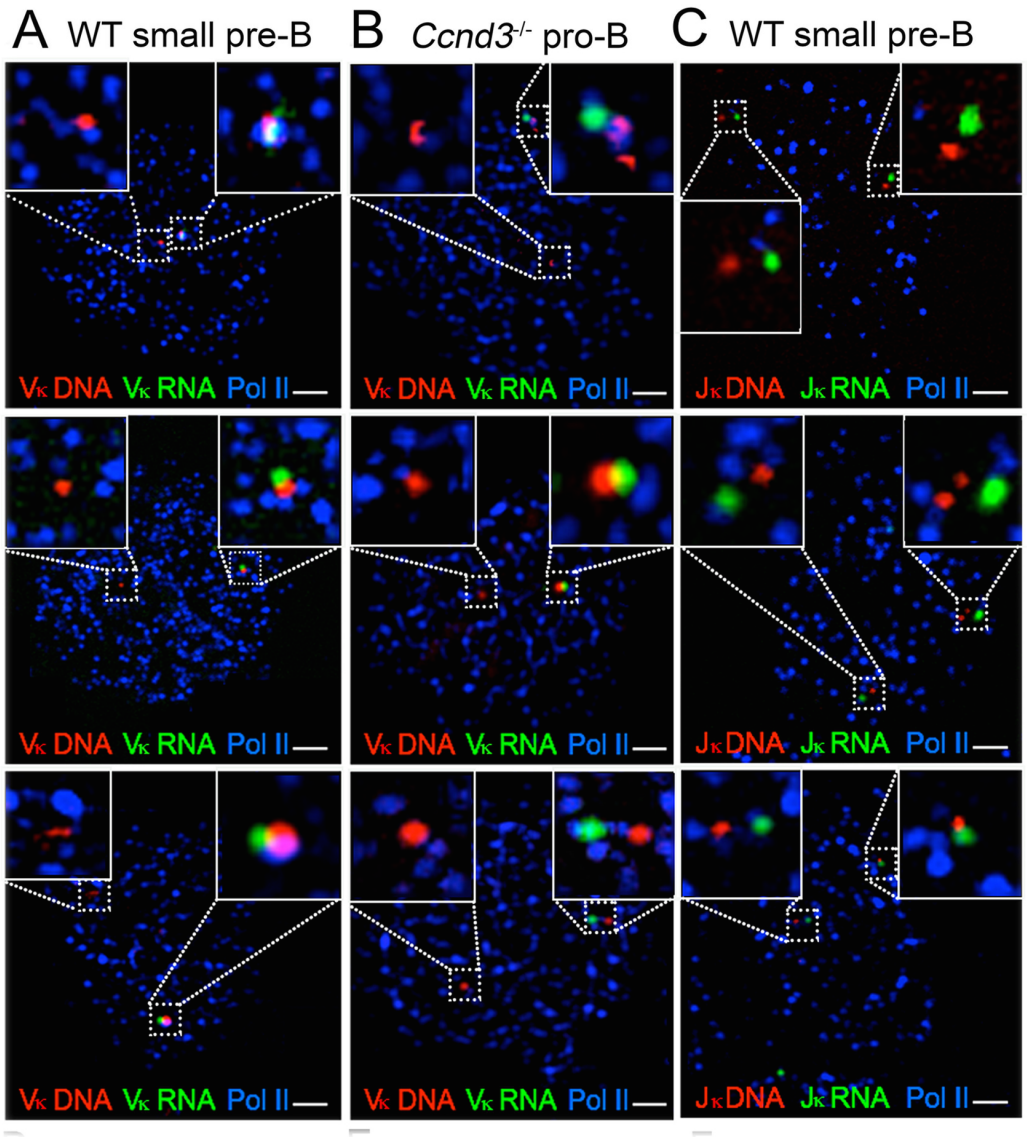


Figure 2.3 Vκ transcription is mono-allelic and dependent upon down regulation of cyclin D3.

(A) Representative confocal images (40 cells, n=2 experiments) of *WT* small pre-B cells, washed extensively to remove soluble nuclear proteins then hybridized to DNA probe RP23-182E6, stained with antibodies specific for e-Pol II followed by hybridization with RNA probe targeting Vκ 1-117. Shown are three representative examples, arranged vertically.

(B) Representative confocal images (40 cells, n=2 experiments) of *Ccnd3*$^{-/-}$ pro-B cells, washed, hybridized and stained as in (A). Shown are three representative examples, arranged vertically.

(C) Representative confocal images (40 cells, n=2 experiments) of *WT* small pre-B cells, washed, hybridized with DNA probe RP24-387E13, stained with antibodies specific for e-Pol II followed by hybridization with RNA probe targeting Jκ-Cκ. Shown are three representative examples, arranged vertically.

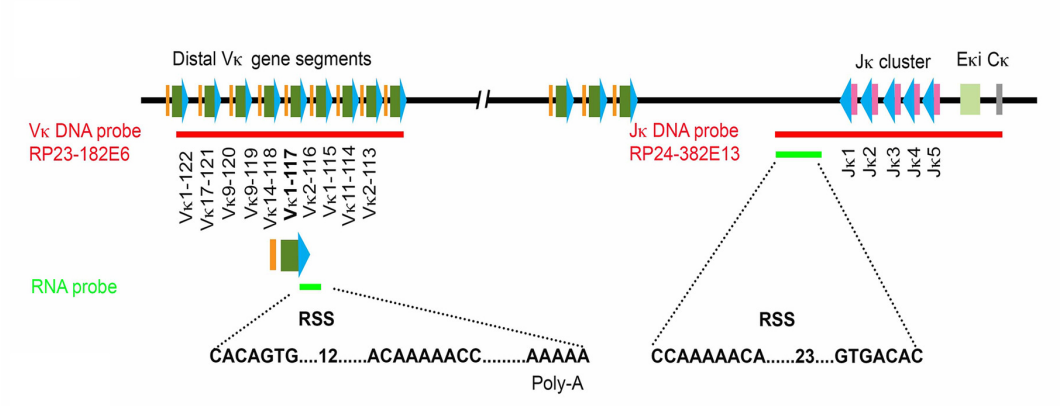


Figure S2.3.1 Schematic of DNA and RNA probes used in the study.

Vκ DNA probe targets distal Vκ gene segments, whereas RNA probe only targets a single Vκ segment (1-117). The RNA probe binds to region 3' of the RSS site, and therefore targets unrearranged transcripts. Jκ DNA probe targets Jκ-Cκ region, whereas the Jκ RNA probe targets region between distal promoter 3.5kb 5' of Jκ1 and Jκ. Vκ RNA was always found close to the gene, whereas Jκ RNA was seen at some distance. This spacing between Jκ DNA and RNA could be due to RNA probe targeting 5' of the transcript, which can be kilobases away from the Jκ.

Vκ 1-117 RNA probe	Affymetrix Product # VB6-17816
	ggtctgtatcactgtgggaggaa gatacccaaagtaagactatttttgttag aacatgaataacatatttagcagctgt ggagtccctcagaacctctctca ccctcagctggaaccaataaact ctgtagccaggctgcacttg gccaaaaatgaatgaccagtgag ttcatattgttattgtggcaatgaa
Jκ RNA probe	Affymetrix Product # VB6-18544
	cctgtctaaactcccacataaatactc acggtctctaggctcctgacc tcttttaaagtacttttatttacttctatagttc cctcccctacctcctcacctct ttccccttttcatactgcatatca ctctgcttgtacttttagtattccact ctcatctccctgccttccac tggacccaccaggctacaac gaatgtagagattggctgaggtcta tggggactgaagaaattgaagaa tgtataactactgccagtgaagcc gctctgggcaaaggttggt aaacacttctgtccacacgcat ggctgtggaaaccaccagag gcatcaaggtattccagcgag taacaaggcatggagcctaact aaatatttttttccacgcaggaa cgatctgtacataagcttattaagctaaa catgcaccctagcataagagaca cactggattgtacaaaaatacaactg aatctacagggagctctagcagtagt gctaagtagaggaattaaatatgaagatc ttgtccatagacttgtaaggttattgt gctagctagtttgaagatgttttgtc gatcattgaggaagaaggtggtaa aattgaatggccttctttggata atgtaatttggcatatgtgtgctag ccagttttgttccccaggtg tgtgcttggttgtttttaccatt ccccatgtgtctaccactgacatat tgaattctgagctctctgccttc ccagcttactctatacgtgtgataaac ttacatattggctgttgtgggtt gaggggctaagcctctgaaat gtggatcagggtttactaacctca cacttcacatgaaagagaaaacagtaa aggacattaggctacagccattt gtgaagatgtgtctcttgatttcaga aagtctggtcccagtctggaaa gcatgtcacaacgtggcacaa cctttcctgtgtgtgcatgtgt catgtgggcacaagccaagt cccacttgtccctgaagaaagg tgagacgctccttatgtccatt caaaggaggccacgtaagga

Table S1 RNA probe sequences.

RNA probes were designed on given target sequence by Affymetrix. Eight or more small RNA probes were designed to decorate the target Vκ and Jκ RNA seqences. Individual small RNA probes are shown. These probes are provided as cocktail and used in conjunction with ISH RNA kit from Affymetrix. Since Jκ-Cκ cluster is >3.5kb, and is larger than Vκ (~300-500bp) more RNA probes could be designed for Jκ versus Vκ (shown above).

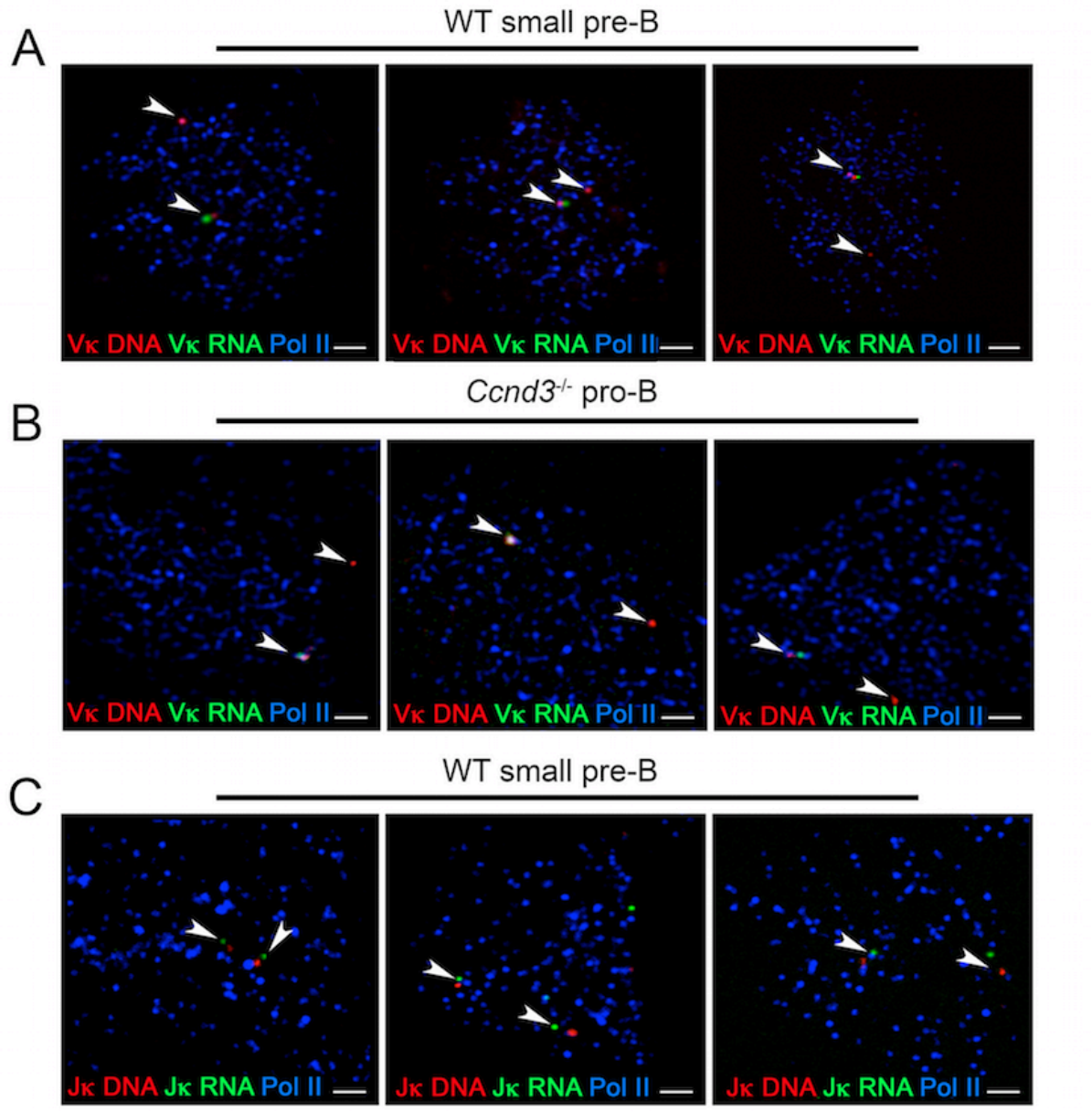


Figure S2.3.2 Monoallelic Vκ transcription upon downregulation of cyclin D3.

(A) Representative confocal images (40 cells, n=2 experiments) of *WT* small pre-B cells, washed extensively to remove soluble nuclear proteins then hybridized to DNA probe RP23-182E6, stained with antibodies specific for e-Pol II followed by hybridization with RNA probe targeting Vκ 1-117. Shown are three representative examples.

(B) Representative confocal images (40 cells, n=2 experiments) of *Ccnd3*$^{-/-}$ pro-B cells, washed, hybridized and stained as in (A). Shown are three representative examples.

(C) Representative confocal images (40 cells, n=2 experiments) of *WT* small pre-B cells, washed, hybridized with DNA probe RP24-387E13, stained with antibodies specific for e-Pol II followed by hybridization with RNA probe targeting Jκ-Cκ. Shown are three representative examples.

Monoallelic Vκ expression predestined by asymmetric e-Pol II niches

We decided to further delve into how allelic choice for monoallelic Vκ expression was made in small pre-B cells, and hypothesized that the two Vκ alleles are localized in asymmetric niche of e-Pol II. In order to address this question, we performed distance mapping on 3D DNA-RNA FISH images between the transcribing or the non-transcribing allele to surrounding e-Pol II using Imaris. By plotting mean distances from e-Pol II to the transcribing versus the non-transcribing allele, we found that the transcribing allele was almost always placed in a tighter e-Pol II niche compared to the non-transcribing allele (Fig 2.4B). This was in contrast to biallelically transcribed Jκ, where both alleles were localized in similar e-Pol II niches (Fig 2.4C). We then performed Vκ and Jκ DNA-FISH on *WT* pro-B, *Ccnd3*$^{-/-}$ pro-B and WT small pre-B cells to investigate if asymmetric e-Pol II niche was associated with asymmetric Vκ-Jκ contraction, which would poise allele in tighter e-Pol II niche for recombination. Asymmetric e-Pol II niche surrounding *Igκ* was apparent in all three-cell types (Fig 2.5A-B, Fig 2.6 A-B, Fig 2.7 A-B, Fig S2.7A-B). Asymmetry seen in *WT* pro-B cells implies that the niche is already set up earlier during development. In *WT* pro-B cells (in the absence of *Igκ* transcription), asymmetric e-Pol II niche translated to large distances between Vκ to Jκ gene segments (200-1000nm) (Fig 2.5 A, C, Fig S2.7A), which were not significantly different between the two *Igκ* alleles. In *Ccnd3*$^{-/-}$ pro-B cells, however, where only germline V segments are active, asymmetric e-Pol II niche associated with increased contraction between Vκ and Jκ, compared to *WT* pro-B in both alleles (Fig 2.6 A,C), although contraction was not complete. In small pre-B cells, asymmetric niche co-related with clear asymmetric contraction, where the allele located in large e-pol II niche contracted to 200-400nm, whereas the allele located in tighter niche of e-Pol II

contracted to 0-200nm (Fig 2.7A, C, Fig S2.7B). Asymmetric niche associated with transcribing as well as fully contracted allele, suggests that physical constraint imposed by a tight e-Pol II niche may favor both transcription and locus contraction, necessary for recombination.

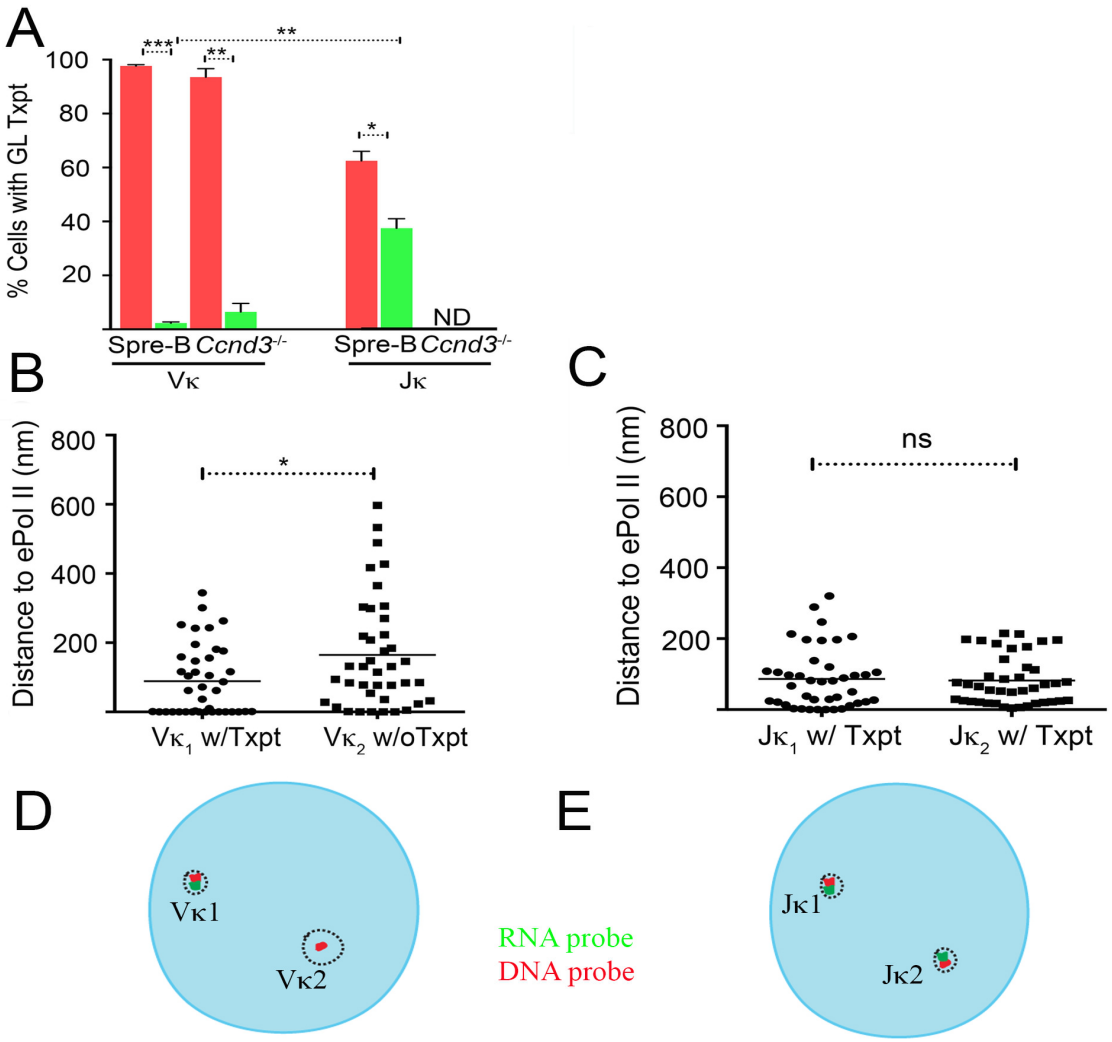


Figure 2.4 Monoallelic Vκ expression associated with asymmetric e-Pol II niche.

(A) Percent cells with Vκ or Jκ germline transcripts scored on 50-60 nuclei per sample (n=2 experiments), and further scored for monoallelic (red) versus biallelic (green) transcription. ND: Not detected. Statistical significance calculated by unpaired Student's t test ($p<0.05$ *, $p<0.01$ ** and $p<0.001$ ***).

(B) Mean distances between Vκ allele with or without transcript and surrounding e-Pol II calculated by Euclidean Distance transformation on Imaris, and plotted for each allele. Statistical significance calculated on 40 images combined from two independent experiments by unpaired Student's t test ($p<0.05$ *, $p<0.01$ ** and $p<0.001$ ***). Schematic shown in (D).

(C) Mean distances between Jκ allele with transcript and surrounding e-Pol II calculated by Euclidean Distance transformation on Imaris, and plotted for each allele. Statistical significance calculated on 40 images combined from two independent experiments by unpaired Student's t test ($p<0.05$ *, $p<0.01$ ** and $p<0.001$ ***). Schematic shown in (E).

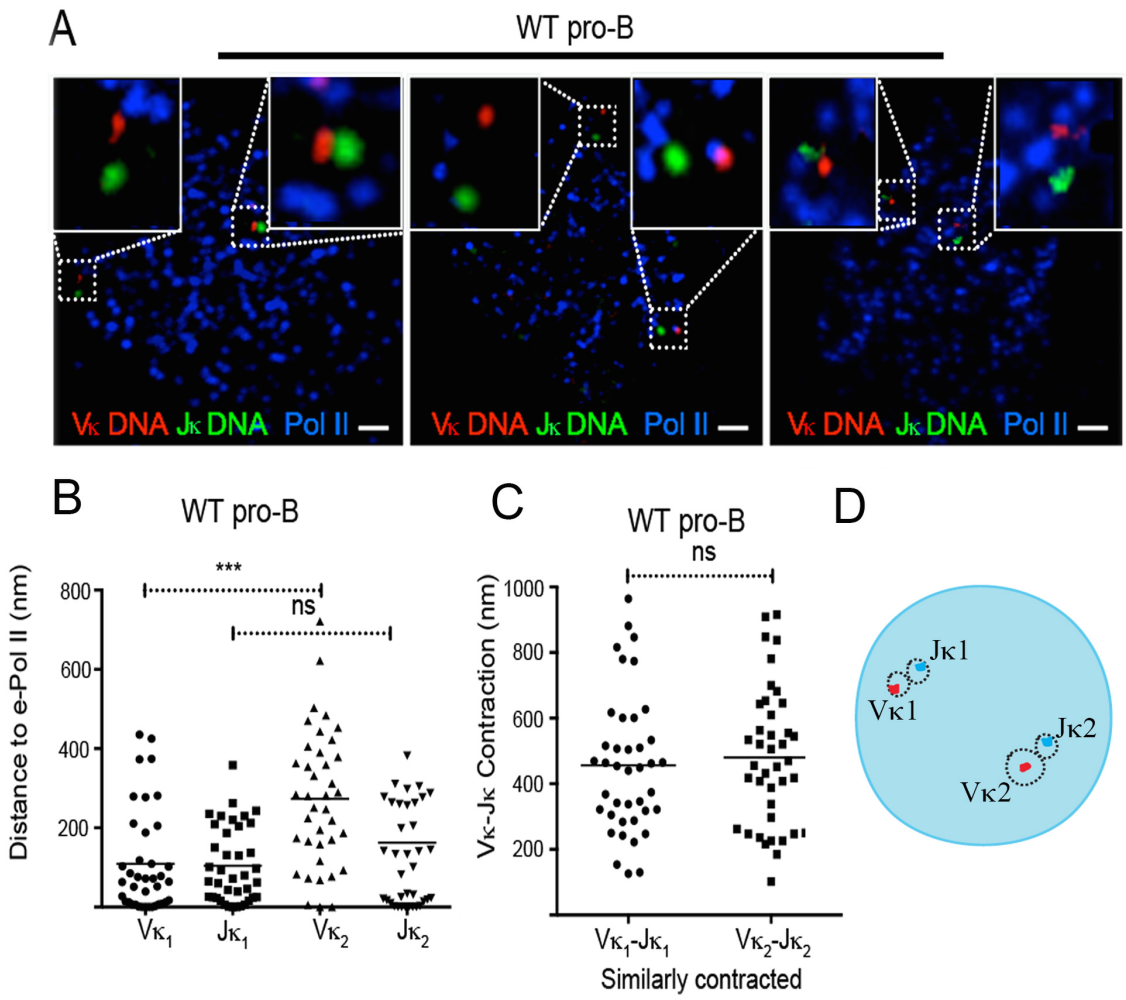


Figure 2.5 Asymmetric e-Pol II niche is set up in pro-B stage associated with varying degree of locus contraction.

(A) Representative confocal images (40 cells, n=2 experiments) of *WT* small pro-B cells, washed to remove soluble nuclear proteins (CSK+0.5%Triton), hybridized to Vκ DNA probe RP23-182E6 (Red) and Jκ RP24-387E13 (Green) and stained with antibodies specific for e-Pol II (Blue).

(B) Mean distances from each Vκ/Jκ allele to surrounding e-Pol II, calculated using Euclidean Distance Transformation on Imaris. Analysis performed on 3D confocal images in (A). Statistical significance calculated on 46 images combined from two independent experiments by paired Student's t test ($p<0.05$ *, $p<0.01$ ** and $p<0.001$ ***). Schematic in (D).

(C) Minimum distances between Vκ and Jκ, plotted for both alleles using Euclidean Distance Transformation. Analysis performed on same images used in (A) and therefore paired for each cell. Statistical significance calculated on 46 images combined from two independent experiments by paired Student's t test ($p<0.05$ *, $p<0.01$ ** and $p<0.001$ ***). Schematics in (D).

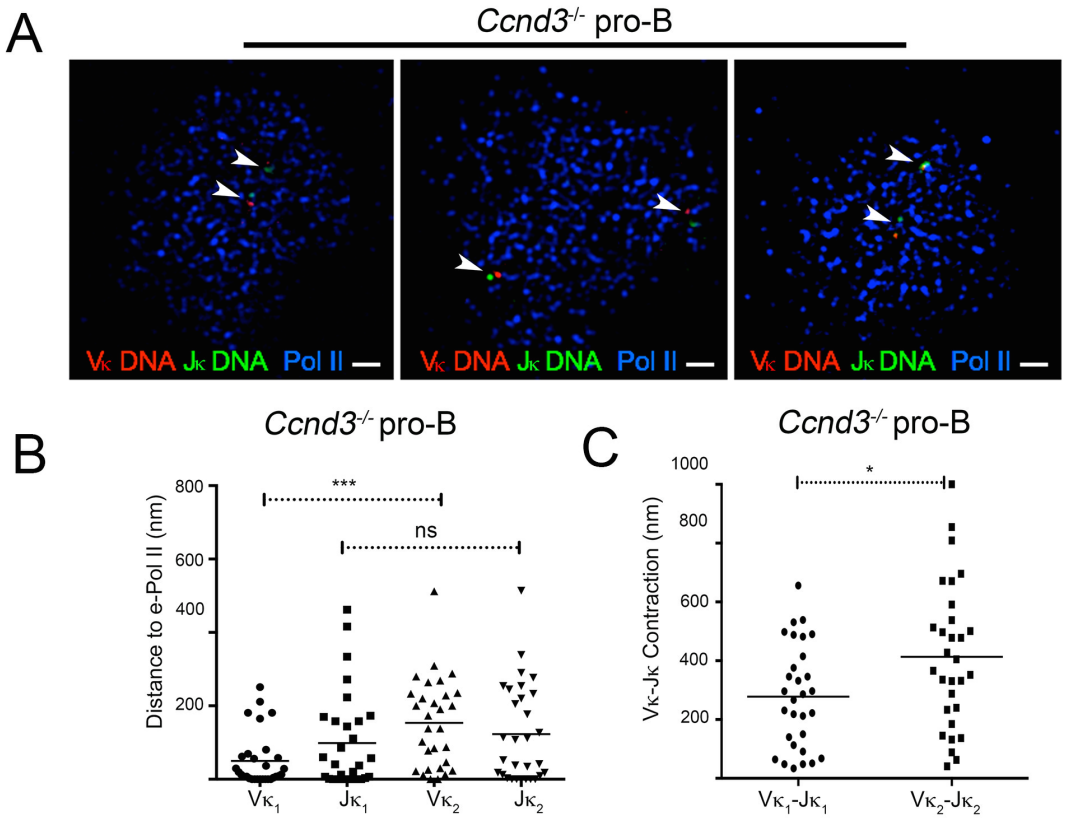


Figure 2.6 Asymmetric niche associated with Vκ transcription and partial contraction.

(A) Representative confocal images of *Ccnd3*$^{-/-}$ pro-B cells, (40 cells, n=2 experiments) washed with CSK+0.5%Triton, hybridized to Vκ DNA probe RP23-182E6 (Vκ:Red) and RP24-387E13 (Jκ:Green) and stained for e-Pol II (Blue). Three representative examples shown horizontally from two independent experiments. Arrows show individual alleles.

(B) Mean distances from each Vκ/Jκ allele to surrounding e-Pol II, calculated using Euclidean Distance Transformation on Imaris. Analysis performed on 3D confocal images (A). Statistical significance calculated on n=37 by paired Student's t test ($p<0.05$ *, $p<0.01$ ** and $p<0.001$ ***).

(C) Minimum distances between Vκ and Jκ, plotted for both alleles using Euclidean Distance Transformation on Imaris. Analysis performed on same images used in (C) and therefore paired for each cell. Statistical significance calculated on=37 by paired Student's t test ($p<0.05$ *, $p<0.01$ ** and $p<0.001$ ***).

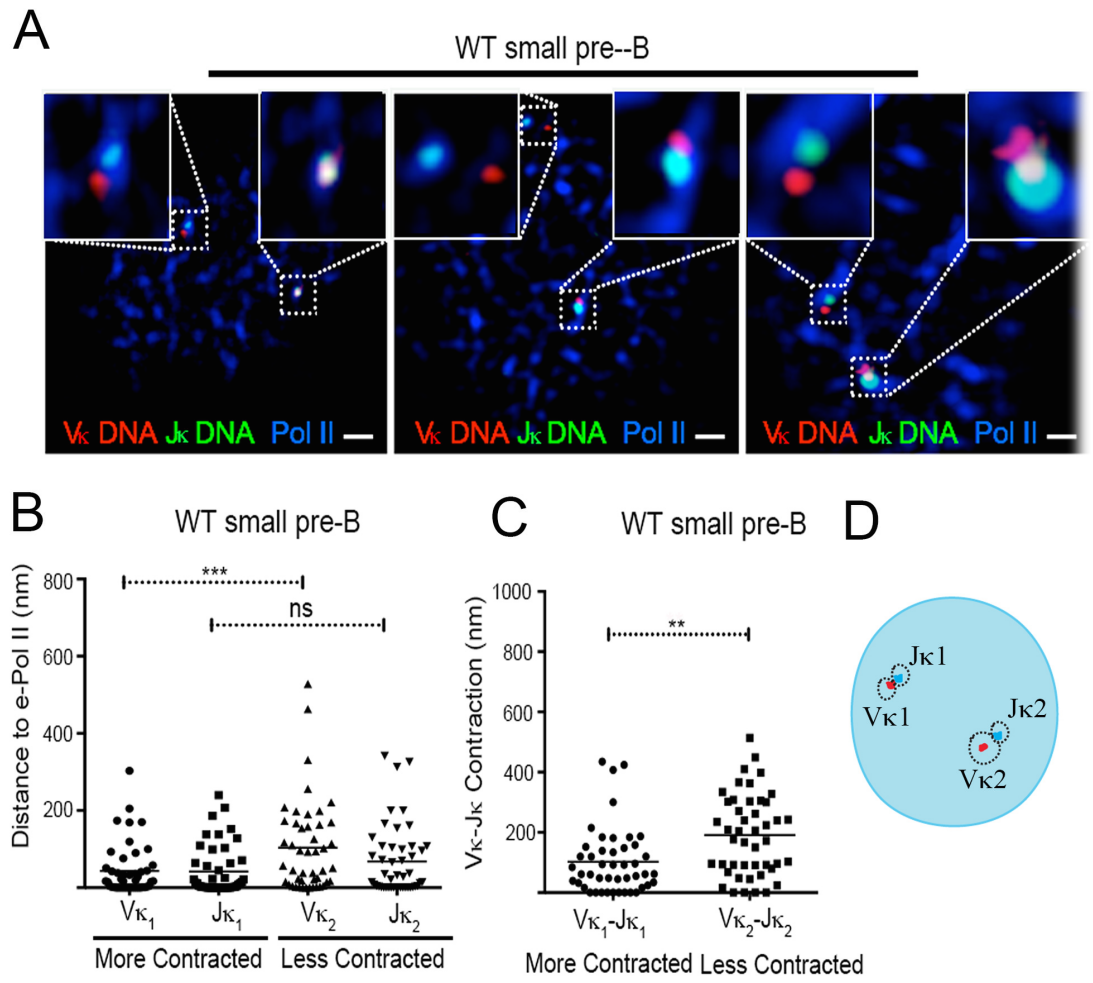


Figure 2.7 Asymmetric e-Pol II niche is maintained in small pre-B cells associated with asymmetric contraction.

(A) Representative confocal images (40 cells, n=2 experiments) of *WT* small pre-B cells, washed to remove soluble nuclear proteins, hybridized to Vκ DNA probe RP23-182E6 (Red) and Jκ RP24-387E13 (Green) and stained with antibodies specific for e-Pol II (Blue).

(B) Mean distances from each Vκ/Jκ allele to surrounding e-Pol II, calculated as in (B). Analysis performed on 3D confocal images (A). Statistical significance calculated on 40 images combined from two independent experiments by paired Student's t test ($p<0.05$ *, $p<0.01$ ** and $p<0.001$ ***).

(C) Minimum distances between Vκ and Jκ, plotted for both alleles. Analysis per formed on same images used in (A). Statistical significance calculated on=40 by paired Student's t test ($p<0.05$ *, $p<0.01$ ** and $p<0.001$ ***). Schematic shown in (D).

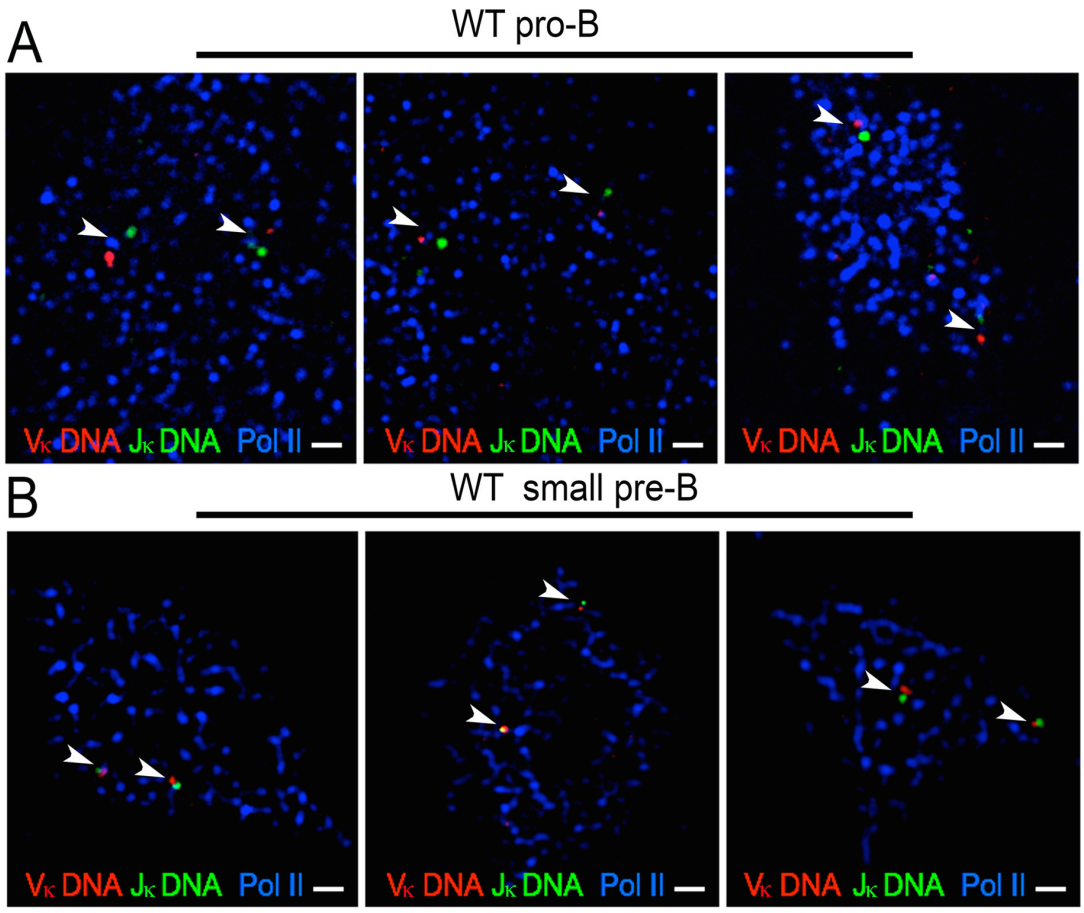


Figure S2.7 Asymmetric niche predicts transcription and locus contraction.

(A) Representative confocal images of *WT* pro-B cells (40 cells, n=2 experiments), washed with CSK+0.5%Triton, hybridized to DNA probe RP23-182E6 (Vκ:Red) and RP24-387E13 (Jκ:Green) and stained for e-Pol II (Blue), shown horizontally from two independent experiments. Arrows show individual alleles.

(B) Representative confocal images of *WT* small pre-B cells (40 cells, n=2 experiments), washed with CSK+0.5%Triton, hybridized to DNA probe RP23-182E6 (Vκ:Red) and RP24-387E13 (Jκ:Green) and stained for e-Pol II (Blue), shown horizontally from two independent experiments. Arrows show individual alleles.

Transcription of captured Vκ loops by nm e-Pol II in single cells

Monoallelic GLT of Vκ 1-117 prompted us to investigate if GLT is predominantly mono-allelic at all Vs, and to further examine if monoallelic activation occurs in single or multiple gene segments in a singe cell. In order to address these questions, we sought to perform single cell RNA-seq on small pre-B cells from a divergent F1 cross that would allow us to discriminate expression from both alleles. Before single cell RNA-seq, we performed bulk RNA-seq of $CD43^{low}$ small pre-B cells from divergent cross (B6 x CAST F1) (Fig S2.8.1), and assigned B6 and CAST specific SNPs to all expressed Vs and tested whether Vκ expression in the cross captured both alleles (Fig S2.8.2). We found that majority of expressed Vs captured expression from both alleles therefore assuring us that allelic assignment in single cell would be result of true monoallelic expression and not technical artifact (Fig S2.8.2). This population expressed Vκ GLT but had not rearranged to a high degree (Fig S2.8.3), therefore confirming that single cell RNA seq on this specific population would give us Vκ expression profile during initial allelic choice. We finally performed single cell RNA seq on small pre-B cells sorted from the cross, gated specifically on $CD43^{low}$ population and assigned allele to each expressed Vκ using B6 and CAST-specific SNPs (Fig S2.8).

We were able to capture different stages of *Igκ* transcription and recombination. From two experiments, we obtained 144 single cell libraries (Fig 2.8, Fig S2.8.4), with an average of $5.2X10^6$ 75bp paired-end reads/cell and 83% concordant alignment rate. Of these, 73 cells did not express Vκ or Jκ genes. Of the remaining 71 cells (Fig 2.8, Fig S2.8.4), 3 expressed only Jκ GLT (Fig 2.8B, Fig S2.8.4: cell numbers 2-60,1-47, 1-46) while 35 cells expressed Vκ and Jκ

GLT (Fig 2.8A, C-D, Fig S2.8.4). These latter 35 cells were poised for *Igκ* recombination but recombination had not occurred. Eighteen cells had started recombination at one allele (Fig 2.8A,E Fig S2.8.4) and 15 cells had recombined on both alleles (Fig 2.8, Fig S2.8.4) with 14 of these manifesting *Igλ* transcription. For all cells with Vκ transcripts, specific Vκ gene expression was always monoallelic. This result is consistent with monoallelic Vκ 1-117 transcription observed by DNA-RNA FISH. Of 35 poised cells, 30 cells expressed Vκ s that could be assigned to just one allele with 5 cells expressing Vκs from both alleles (Fig S2.8.4: cell numbers 1-19,1-21,1-37,1-4 and 2-5). In contrast, and consistent with previous reports (Amin et al. 2009), Jκ expression was always biallelic. These data indicate that Vκ expression is primarily monoallelic. We next included Vκ genes that could not be assigned to one particular allele (Fig S2.8.4). These data revealed that each cell could express multiple Vκ genes with up to 12 in a single cell poised for *Igκ* recombination (cell number 1-21). In general, distal Vκ genes were expressed in almost every cell with a subset of cells expressing both distal and proximal Vκ genes (eg: cell numbers 2-36,1-19,1-21,1-7,1-4,1-39, 2-5). Intermediate Vκ genes were infrequently expressed.

The distributions of Vκ expression could also be grouped based on topological domains predicted from known CTCF binding sites and *Rag 2*$^{-/-}$ pro-B cell Hi-C data (summarized Fig S2.8.4 top, Fig S2.8.5)(Aoki-Ota et al. 2012, Choi et al. 2013, Lin et al. 2012, Ribeiro de Almeida et al. 2011). Only high probability CTCF-mediated loops are shown (Fig S2.8.4). Distal Vκs fell within three loops (L1-L3). The L1 and L2 loops are relatively small with frequent Vκ expression throughout these topological domains. In contrast, expressed Vκs in the larger L3 loop tended to cluster towards both CTCF sites. The L4 region is large and lies between two

more clearly defined loops. Expression through this region was sparse and clustered closer to the 3' CTCF. Expression through the smaller L5 and L6 loops was more evenly distributed. In single cells, Vκs from up to four separate loops were expressed (eg: cell numbers 1-75: L1/2/3,1-19: L1/2/3/4,1-21: L1/2/3/5).

Examination of Vκ orientation and expression in each loop revealed specific patterns. For the larger L3 loop and L4 domain, Vκ genes were primarily oriented away from proximal CTCF sites, and tended to be transcribed in this orientation (eg: cell numbers 1-33: L4, 2-5: L4,1-15: L4, 2-4: L3, 4, 2-14: L5, 1-75: L3, 1-18: L3). In the smaller loop L2, Vκ genes were oriented in both directions. The Vκs in L1, L5 and L6 were primarily oriented in the forward, reverse and forward orientations respectively. These data suggest that e-Poll II captures loops at or near CTCF sites (Chernukhin et al. 2007, Pena-Hernandez et al. 2015) and then can read through several Vκ genes. This was evident by Vκ -Vκ splicing observed for all loops (frequently in L3), which is suggestive of read through transcription and pre-mRNA processing. Both intra-loop (eg: cell numbers 1-37,1-75, 2-11, 2-4, 2-14) and inter-loop (eg: cell numbers 1-75, 1-11, 2-31) read-through transcription and pre-mRNA processing was observed (Fig S2.8.5).

These data indicate broad Vκ expression. However, only relatively few Vκ alleles, 9 of 96 functional Vκs belonging to seven different Vκ gene families, are highly used in the initial repertoire (Aoki-Ota et al. 2012). Highly used Vκ s were distributed throughout the loops with L1 containing 1, L2 with 1, L3 with 4 and L5 with 3 highly used Vκ genes (Fig S2.8.4). All 9 highly used Vκs were expressed in the 35 poised cells transcribing Vκs. Thirty-four cells

expressed one assignable highly used Vκ or less while only one cell expressed two assignable highly used Vs (cell number 1-21). These data suggest that a combination of Vκ expression, and availability for recombination, contributes to *Igκ* monoallelic choice. All poorly transcribed Vκs are also under represented in the initial *Igκ* repertoire (Aoki-Ota et al. 2012). Interestingly, the majority of these Vκs are in L4 (Fig S2.8.4). We did observe expression of some L4 Vκ alleles in a relatively few single cells. These expressed L4 Vκ alleles were primarily located in the 3' end of L4 but removed from the flanking CTCF site. The 5' flanking CTCF sites likely provides a strong e-Pol II anchor (Fig S2.8.4)(Choi et al. 2013). However, the Vκ alleles 3' to these sites are in the wrong orientation for read-through transcription. These data suggest that L4 Vκ allele transcription primarily occurs through stochastic loop capture and not e-Pol II anchoring to flanking CTCF sites.

Of 33 cells that had undergone *Igκ* rearrangement, 18 cells showed rearrangement at a single allele, as evident by monoallelic distal Jκ promoter expression with biallelic Cκ expression. Fifteen cells showed rearrangement at both alleles (Vκ + Cκ without distal Jκ p expression)(Fig S2.8.4: white arrows showing V-J-C splicing/recombination). Rearrangement events were limited to proximal Vκs confined within L5 and L6. These data indicate that proximal Vκ recombination precedes distal Vκ recombination. Recombination appeared inefficient as Vκ -Jκ recombination products were detected in only three cells (cell numbers 2-12, 2-14 and 1-23) and all were out of frame. Fourteen cells had *Igλ* expression with at least Jλ to Cλ splicing suggesting that *Igλ* was poised for recombination in these cells.

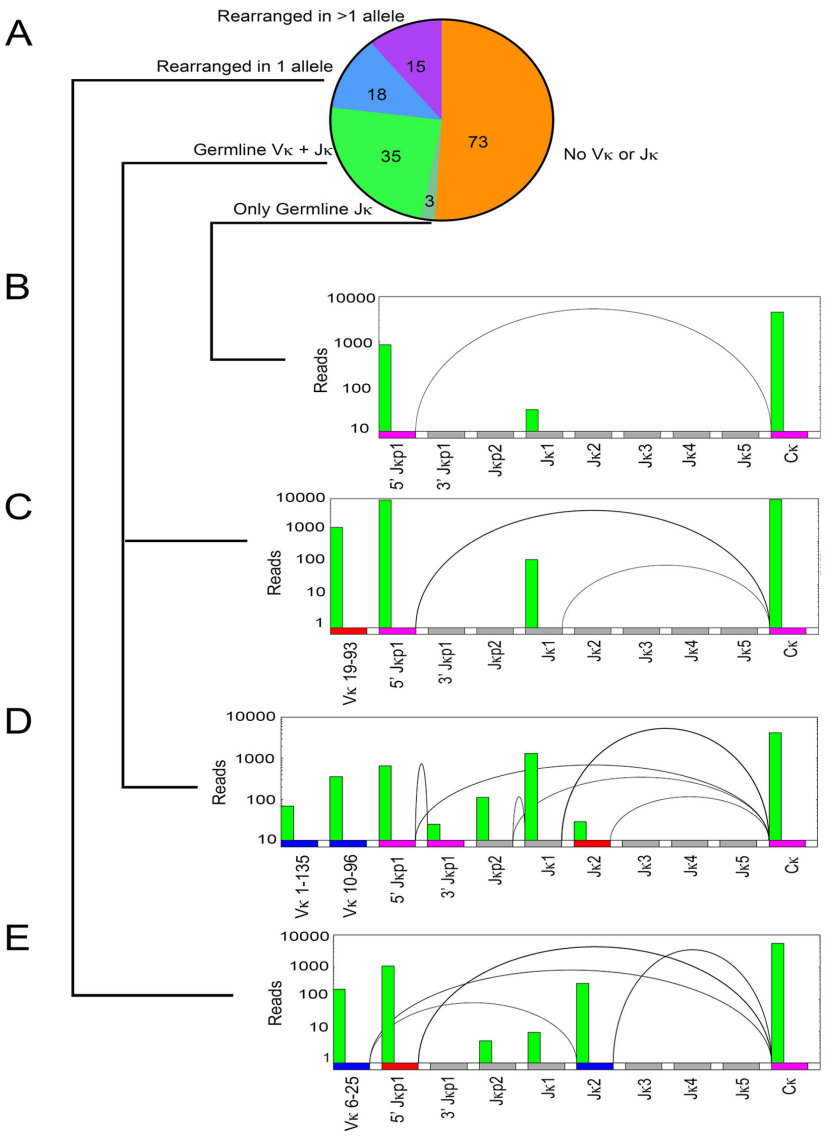


Figure 2.8 Monoallelic activation of Vκ.

(A) Single cell RNA-seq *Igκ* expression profile of $CD19^{+}B220^{+}CD43^{low}IgM^{-}$ small pre-B cells. Of 144 cells captured, number of cells expressing only germline Jκ (dark green), germline Vκ-Jκ (green), rearrangement in one allele (blue), rearrangement in more than one allele (purple) and no Vκ or Jκp(orange) are shown.

(B) Representative example of biallelic germline transcription of 5'Jκ1 promoter (magenta) without Vκ expression.

(C) Representative example of biallelic germline transcription of 5'Jκ1 promoter (magenta) with single Vκ expression from B6 allele (red).

(D) Representative example of biallelic germline transcription of 5'Jκ1 promoter (magenta) with multiple Vκ expression from CAST allele (blue).

(E) Representative example of proximal Vκ recombination on CAST allele(blue).

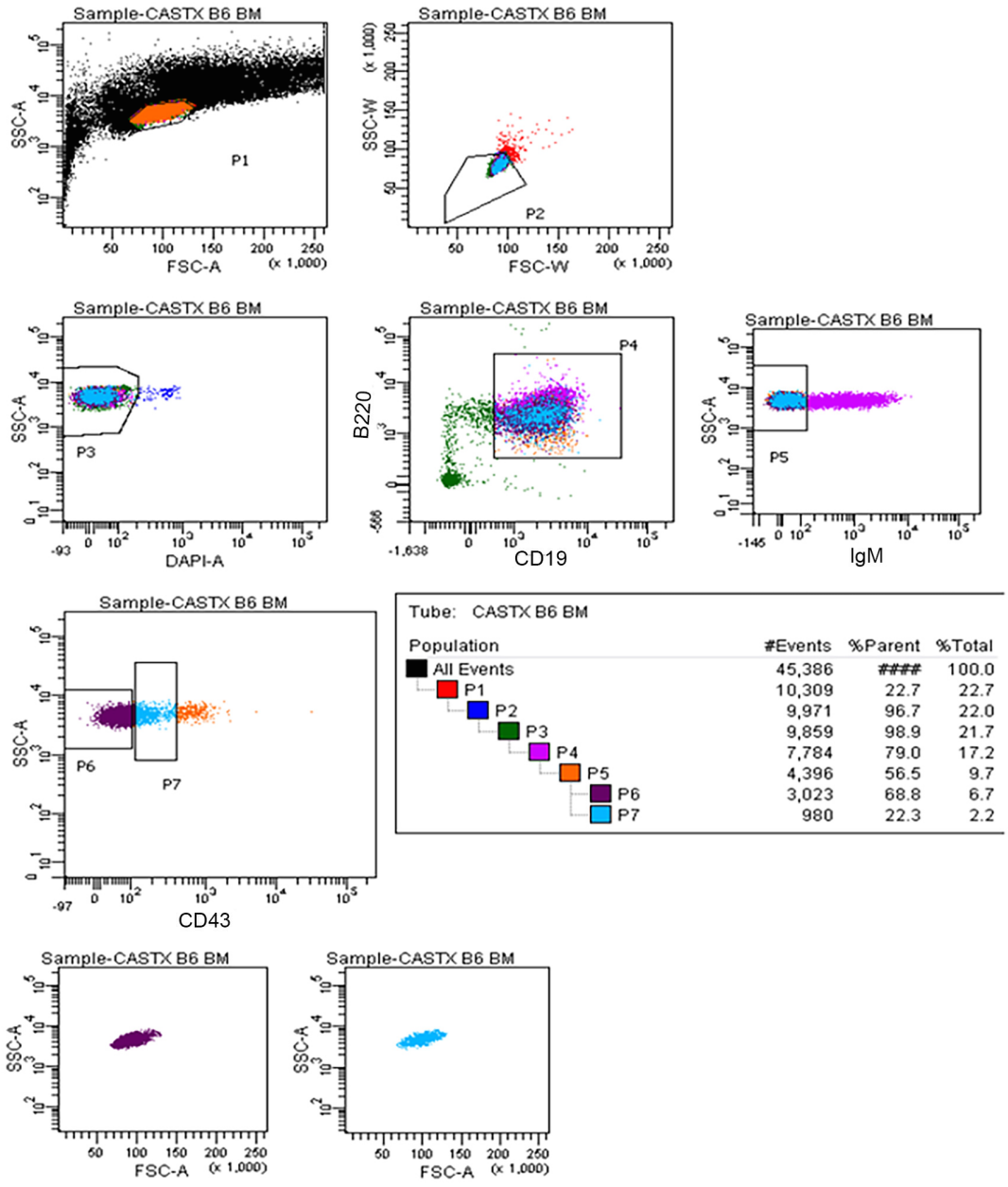
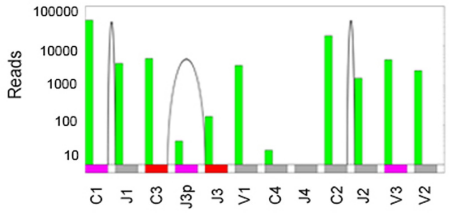


Figure S2.8.1 Bulk RNA-seq of B6X CAST showing sorting gate.

BM extracted from B6X CAST F1 were sorted for $CD19^{+}B220^{+}IgM^{-}CD43^{low}$ cells. This gating strategy was employed to test if $CD43^{low}$ cells would capture Vκ and Jκ transcription in germline configuration. $CD43^{+}$ cells are pro-B cells, and expected to not express Vκ and Jκ, whereas $CD43^{-}$ cells are small pre-B cells and expected to express only rearranged Vκ-Jκ transcripts.

A



B

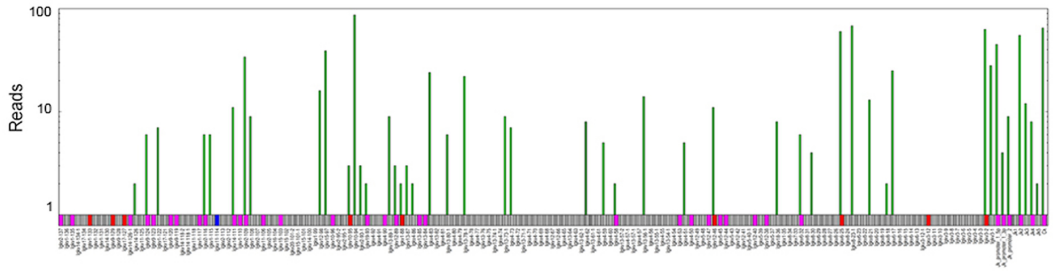


Figure S2.8.2 Validation of capture of B6 and CAST alleles of *Igκ* and *Igλ* loci by bulk-RNAseq.

RNA-sequence reads from 1X10^6 CD43^{low} small pre-B cells from B6 X CAST F1 cross were aligned to mm10 genome build of B6 reference genome. Alleles were assigned to B6 or CAST genome using CAST-specific variants, from Sanger Institute (see methods). For most assignable Vs, we could detect expression from both B6 and CAST (magenta) alleles, suggesting that single cell RNA seq would be able to capture both B6 and CAST alleles . Some Vs that were assigned as B6 (red) or CAST (blue) alone, were usually the ones with low read count (below 10).

.

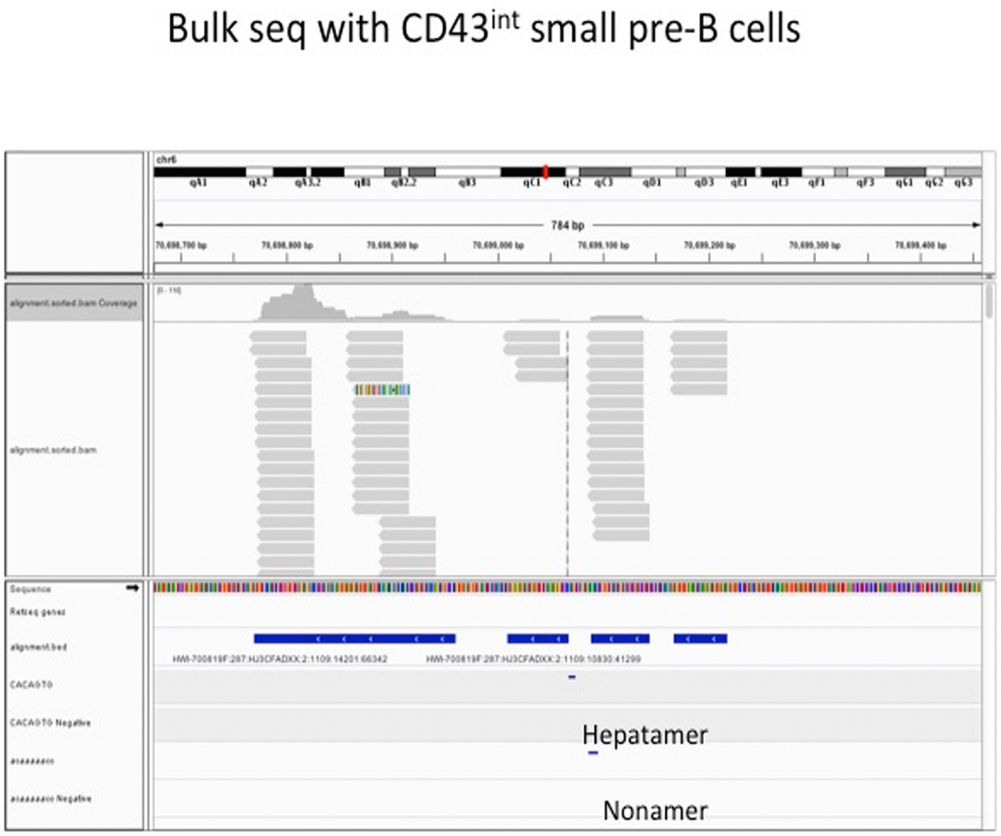


Figure S2.8.3 Validation of bulk RNA-seq of B6X CAST sorting gate.

B6xCAST F1 CD43^{low} small pre-B cells RNA sequences aligned to B6 genome. Hepatamer and nonamer motifs were confirmed for a fraction of reads as shown, suggesting their germline configuration. This gate was used for single cell RNA-sequencing.

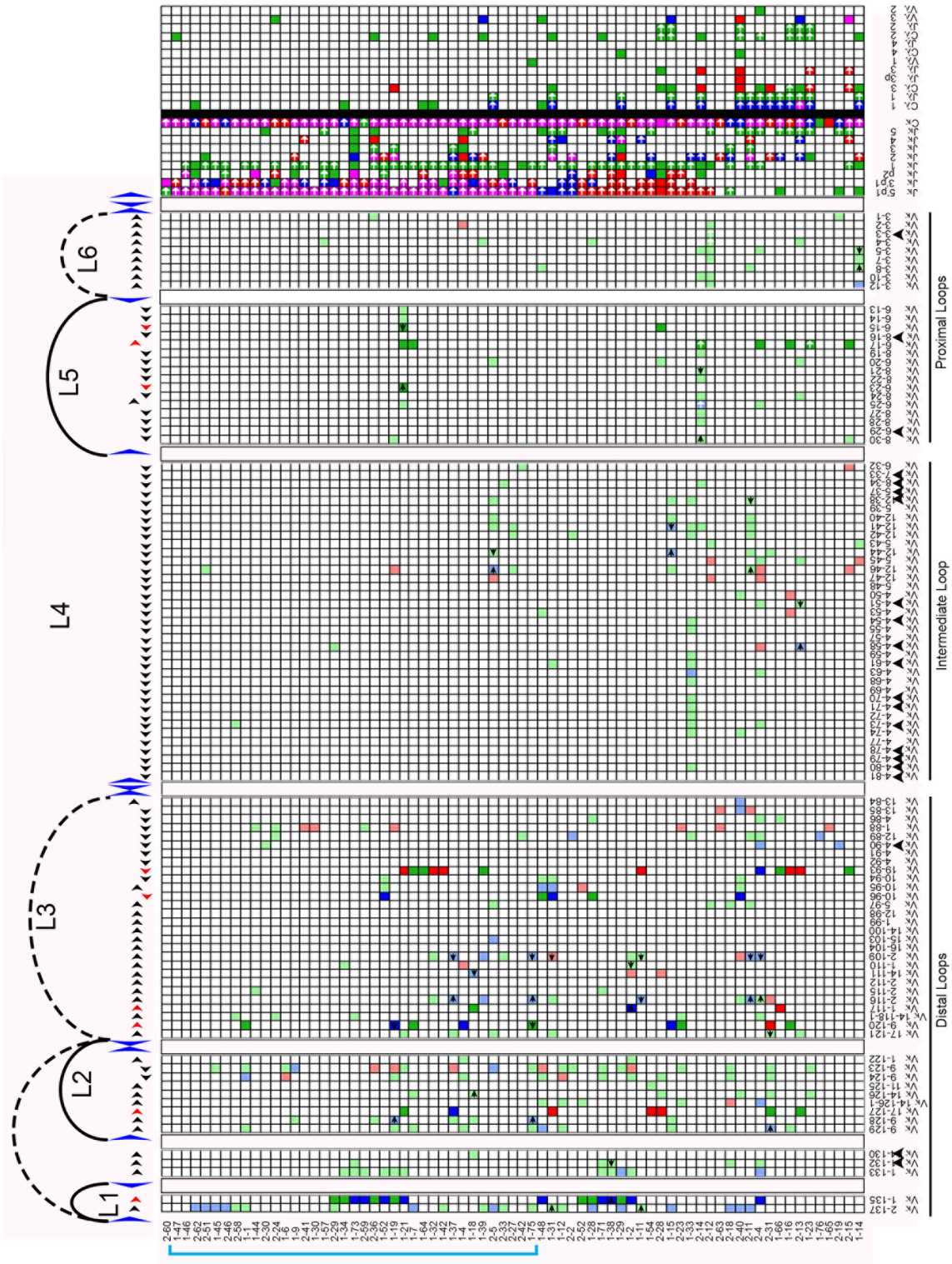


Figure S2.8.4 Single cell RNA-seq heatmap of *Igκ* and *Igλ* expression.

Figure S2.8.4 continued.

Sequence reads from 144 single cells from B6 X CAST F1 cross were aligned to mm10 genome build of the B6 reference genome. Alleles were assigned to B6 or CAST genome using CAST-specific variants, from Sanger Institute (see methods).The heatmap shows 71 analyzable cells, with Vκs expressed from B6 shown in red, CAST shown in blue and unassigned shown in green. High usage Vκs (Aoki-Ota et al. 2012) are shown as dark red (B6), dark blue (CAST) and dark green (unassigned). Biallelic expression of Vκ or Jκ segments is shown as magenta. All Jκ and Cκ segments are shown as dark red, blue or green for B6, CAST or unassigned respectively. Jκ promoter to Cκ pre-mRNA splicing and Vκ to Jκ recombination events are shown as white arrows, whereas Vκ-Jκ splicing is shown as black arrows within each box. Unrearanged cells are marked by blue line on the left. Total of 105 Vκs are shown, which include 93 that have been evaluated to be functional (Martinez-Jean, Folch, and Lefranc 2001) and 12 that showed expression in 3 or more cells. CTCF location and orientation are shown as blue arrows on the top. Only peaks with strong binding (Choi et al. 2013) and high contact probability (see methods) are shown. Black loops represent loops made by convergent (forward (F) and reverse (R) : solid) and divergent (F+F, R+R and R+F : dotted) CTCF sites. Smaller black arrows on the top represent transcriptional orientation of each Vκ, with red highlighting orientation of high used Vs.

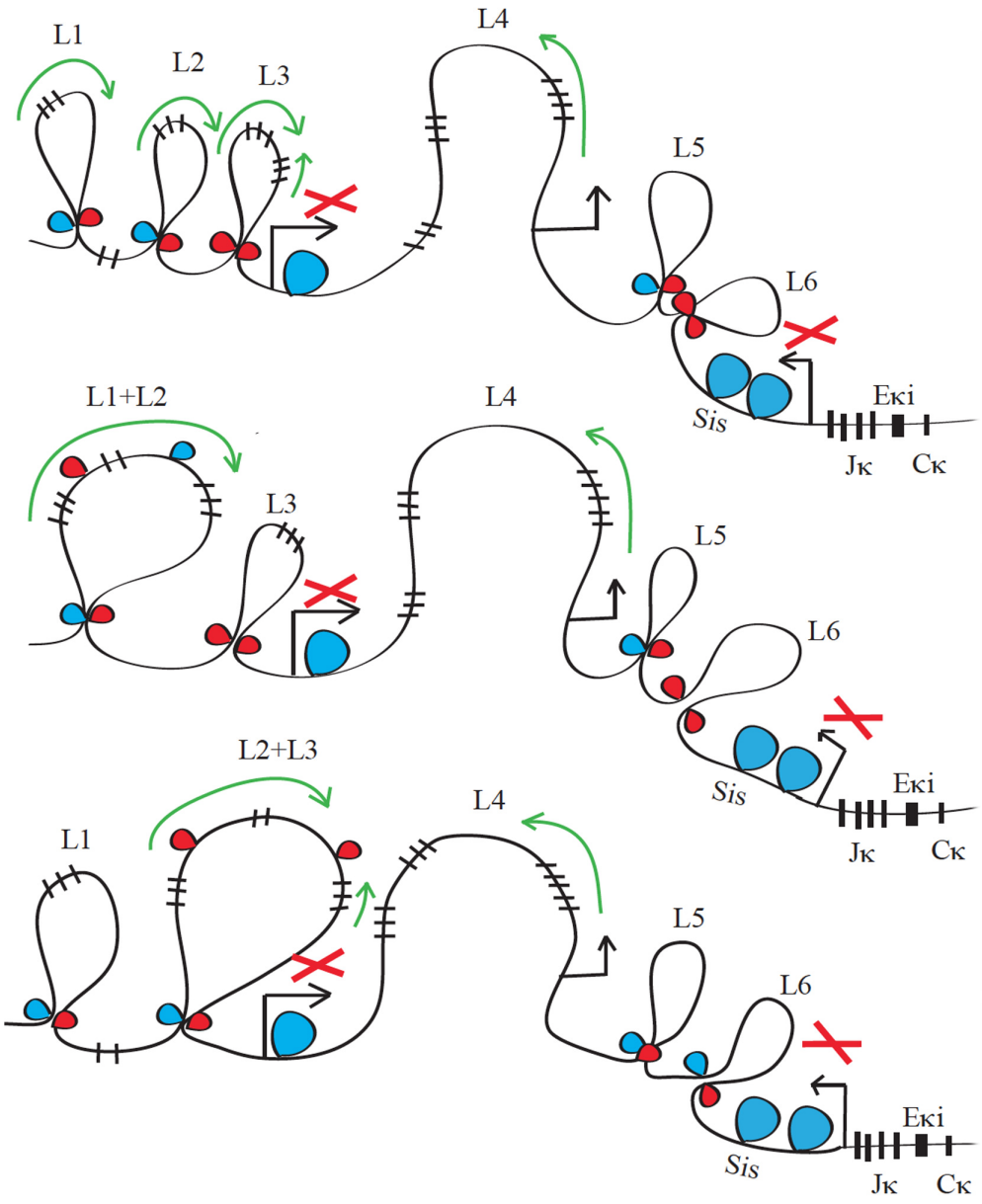


Figure S2.8.5 Alternative CTCF pairing to make Vκ loops.

Shown here is a diagram where convergent CTCF pairs (Blue: forward and Red: Reverse) form 5 defined loops. L4 is less structured. Read-through transcription happens within a CTCF-contained loop spanning multiple Vs (A). Alternative CTCF pairing may happen for distal CTCF sites, where L1+L2 or L2+L3 may form single loop (B-C). The CTCF site between L3 and L4 is a true TAD boundary as no alternative pairing or read-through transcription occurs between these loops. Similarly, in the 3' end, Sis element bound by CTCF defines another boundary. Size of CTCF reflects their binding strength.

Nuclear niche predicted to shape repertoire of transcribed Vκ genes

We then sought to combine our imaging data showing asymmetric e-Pol II niche surrounding Vκ with single-cell RNA-seq data into a 2D model. In this model, e-Pol II niche dimension surrounding Jκ at both alleles is same whereas that surrounding distal Vκ is variable between the two alleles (Fig 2.9 A). We focused primarily on modeling distal Vκ expression and usage as we used one of the distal Vs for DNA-RNA FISH (Vκ 1-117) and contraction analysis, and also because most highly transcribed and highly used Vs are distally located. We obtained niche dimensions surrounding both *Igκ* alleles from distance mapping in WT pro-B cells (Fig 2.5 B,C), where the niche is set-up in the absence of transcription, and therefore precludes any transcription-associated skewing of DNA to e-Pol II distances. Jκ niche diameter was similar between the two alleles (Allele 1 and Allele 2) and was estimated to be ~200-250nm (i.e twice the mean distance in Fig 2.5 B), whereas Vκ niche diameter was estimated to be 250nm for Allele 1, and 600nm for Allele 2. Based on these measurements we predict a cylindrical shaped niche for the allele fated for recombination or chosen allele (Allele 1), and a larger conical niche for the excluded allele (Allele 2). In *WT* pro-B cells, both alleles are in similar configuration extending between 200-1000nm from distal Vκ to Jκ (Fig 2.5C, Fig 2.9A). Upon entering small pre-B stage, Allele 1 undergoes complete contraction, allowing distal Vκ to interact with Jκ, whereas Allele 2 undergoes some level of contraction but does not undergo complete contraction to interact with Jκ (Fig 2.6C, Fig 2.9B). Based on Vκ -Jκ contraction data and single cell RNAseq data, Allele 2 or excluded allele can undergo some level of transcription within the cone in small pre-B cells, however, does not contract as far as chosen alleles to allow entry into the

"recombination center". Allele 1 or the chosen allele on the other hand contracts enough to enter the "recombination center", where increase in local concentration of distal Vs, increases the likelihood of synapse formation with Jκ. We also measured asymmetric niche around Vκ in *Ccnd3*$^{-/-}$ pro-B cells, where we found partial Vκ -Jκ contraction, suggesting that there might be an initial transcription associated contraction before Vκ completely contracts to Jκ (Fig 2.6 A-C).

In the model, distal loops 1,2 and 3 and proximal loops 5 and 6 are aptly positioned to make favorable interactions with surrounding e-Pol II sites and with Jκ by contraction, whereas intermediate loop L4 is positioned to face away from e-Pol II sites and Jκ. We postulate that confinement of *Igκ* within tight e-Pol II niche not only exposes highly transcribed and used Vs and enable them to interact with e-Pol II, but also allow favorable contraction with Jκ.

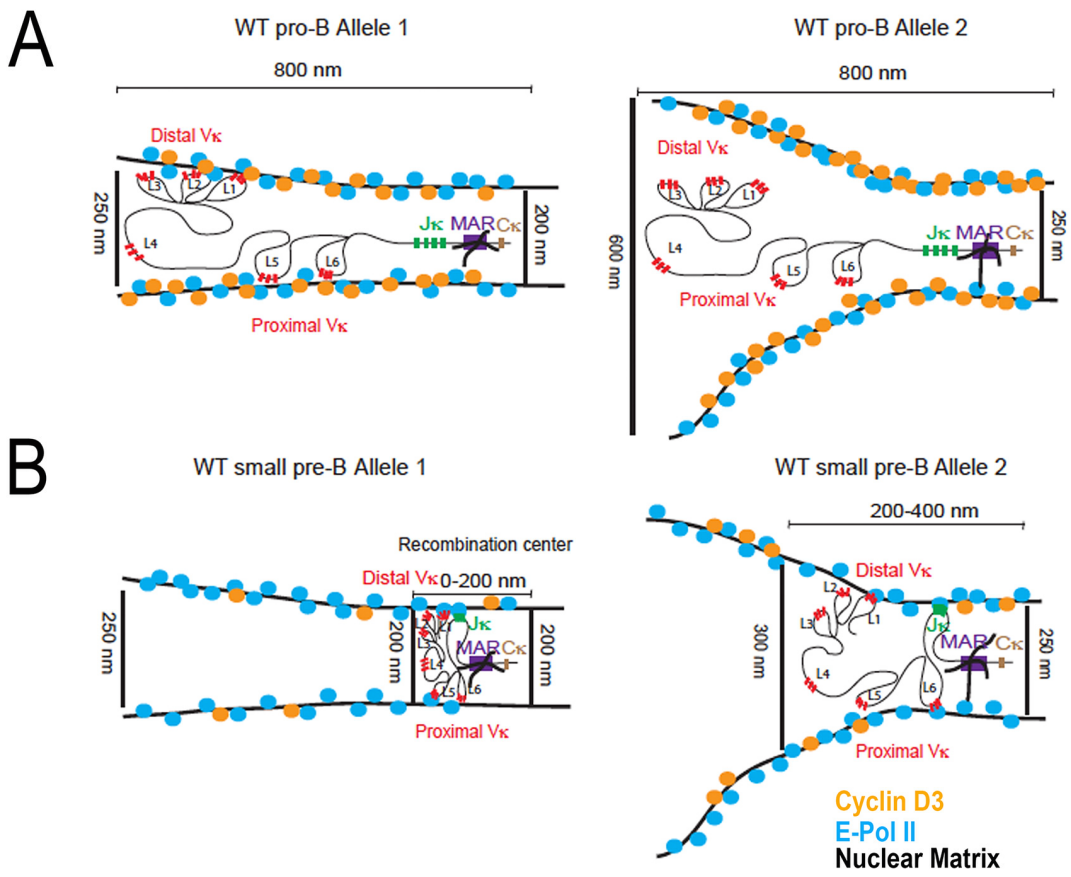


Figure 2.9 Nuclear niche predicted to shape Vκ transcription repertoire.

(A) In *WT* pro-B cells, where level of matrix-bound cyclin D3 (orange) is high, Vκ gene segments (red) are inaccessible to e-Pol II transcription centers (blue). E-Pol II niche around *Igκ* alleles is asymmetric for Vκ (Vκ to e-Pol II ranges from 250-600nm) but symmetric for Jκ (~Jκ to e-Pol II 200nm-250nm for both alleles). E-Pol II niche surrounding Allele 1 creates a cyclinder, whereas that for Allele 2 forms a truncated cone. In both niches, *Igκ* extends ~500nm on average.

(B) In *WT* small pre-B cells, where levels of cyclin D3 decline, multiple Vκ gene segments randomly associate with e-Pol II transcription centers. Allele 1 residing within the cylinder contracts along the length of the cylinder and interacts with Jκ, whereas Allele 2 contracts partially but fails to reach Jκ.

3D simulation of *Igκ* polymer confined within fixed e-Pol II niche

We used polymer chain simulation of *Igκ* to investigate how Vκ domains might organize within the cylindrical niche defined by confocal distance measurements. Simulation of Vκ structure was performed either without spatial constraints or constrained within a 0.8 μmX0.2 μm cylinder, which approximates nuclear niche size when Vκ transcription is first initiated (Fig 2.5). When unconfined, the *Igκ* polymer folded into a globular meshwork of DNA, without prominent loop structures (Fig 2.10A). Furthermore, most CTCF sites were embedded within the meshwork (Fig 2.10B). Interestingly, those CTCF sites predicted to form L2 (light blue) are not closely apposed suggesting that this loop would not form in a spatially unconstrained environment. Predicted loop structure changed dramatically when restricted to a cylinder (Fig 2.10C). The proximal loops (blue) interacted extensively with Jκ (3' green) consistent with preferential initial recombination. Loops containing highly transcribed Vκs (the distal loops L1, 2, 3 shown in purple and L5, blue) occupied defined regions within the cylinder. In contrast L4 (white), a large, unstructured domain flanked by the more well formed L3 and L5 loops, extended across much of a surface of the other loops. Most CTCF sites were paired to form loops and were exposed along the cylinder (Fig 2.10D). This is predicted to make them relatively accessible to nm e-Pol II. However, the CTCF pairs flanking L4 5' (light green) and 3' (yellow) are widely separated, which likely contributes to the unstructured nature of the L4 domain. These results suggest that restriction within nm cylindrical niche positions transcriptionally permissive Vκ containing loops for capture by nm-ePol II.

We displayed highly transcribed/highly used and poorly transcribed/poorly used Vs along *Igκ* polymer (Fig 2.11 B-D) and measured perpendicular distances from each V gene to the cylinder (Fig S2.11), with the hypothesis that highly transcribed/used Vs would be exposed and located closer to the e-Pol II cylinder compared to the poorly transcribed/used Vs. We saw some differential clustering of highly transcribed and poorly transcribed Vs across the cylinder, however, there was no obvious correlation between frequency of highly transcribed Vs and their distances to e-Pol II (Fig 2.11 B-D, Fig S2.11). This result suggests that although our simulated model captured loop formation and orientation of loops well, it was unable to predict why some Vs are transcribed and used more than others.

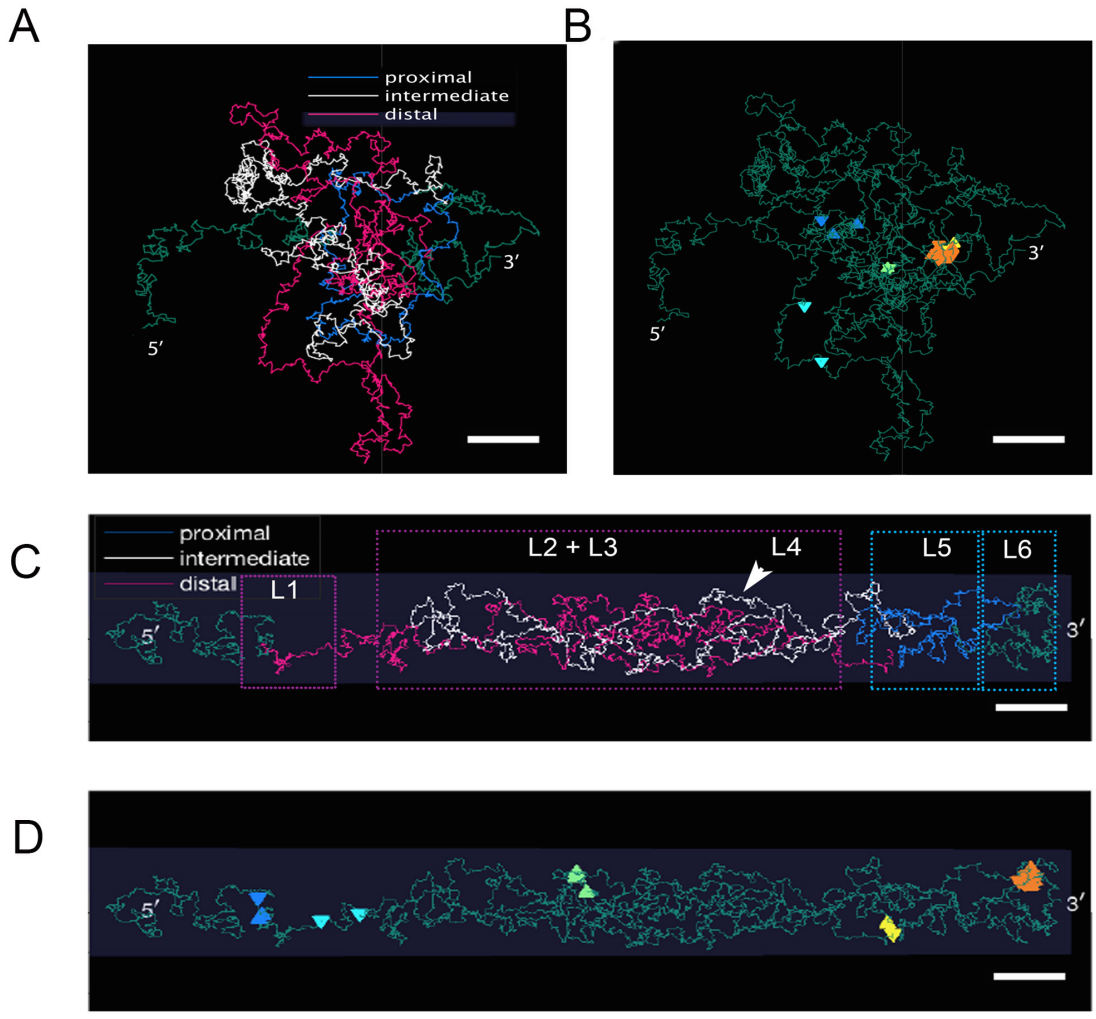


Figure 2.10 Polymer chain simulation of *Igκ* within free or cylindrical confinement and comparison of CTCF-formed loops.

(A) Snapshot of *Igκ* when unconfined showing distal (purple), intermediate(white) and proximal (blue) Vκ region and 3' (Jκ) and 5' end shown as green.
(B) Snapshot of *Igκ* when unconfined showing CTCF sites with high contact probability.
(C) Snapshot of *Igκ* when confined within 0.8μm x 0.2μm cylinder showing distal (purple), intermediate(white) and proximal (blue) Vκ region and 3' (Jκ) and 5' end shown as green. Dotted boxes show location of Loop 1, 2, 3, 5 and 6 as they compare to Fig S2.8.4.
(D) Snapshot of *Igκ* when confined within 0.8μm x 0.2μm cylinder showing CTCF sites with high contact probability.

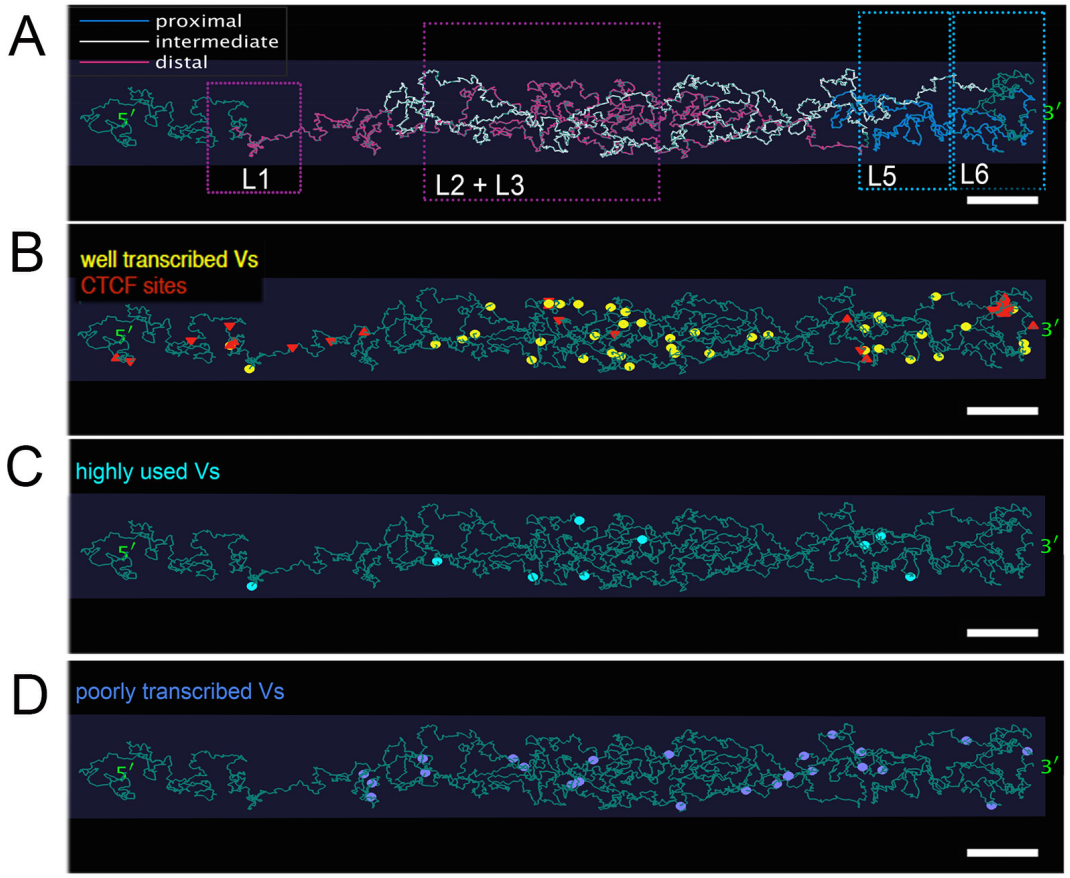


Figure 2.11 Polymer chain simulation of *Igκ* within free or cylindrical confinement and comparison of distance of highly and poorly transcribed Vs from the cylinder.

(A) Snapshot of *Igκ* when confined within 0.8μm x 0.2μm cylinder showing distal (purple), intermediate(white) and proximal (blue) Vκ region and 3' (Jκ) and 5' end shown as green. Dotted boxes show location of Loop 1, 2, 3, 5 and 6 as they compare to Fig S2.8.4.
(B) Snapshot of *Igκ* when confined within 0.8μm x 0.2μm cylinder showing well expressed Vκs and CTCF sites with high contact probability.
(C) Snapshot of *Igκ* when confined within cylinder showing highly used Vκs (cyan).
(D) Snapshot of *Igκ* when confined within cylinder showing poorly used Vs (blue).

A

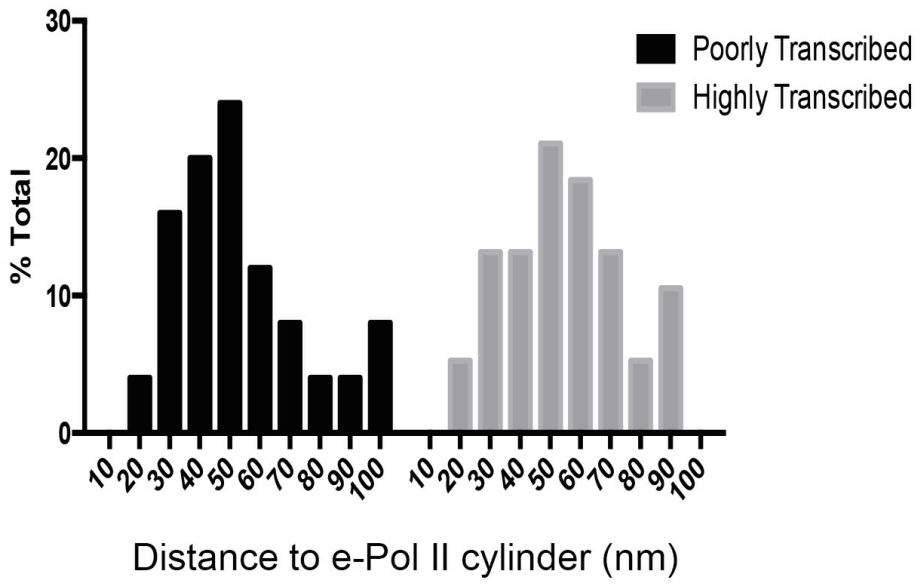


Figure S2.11 Distribution of highly and poorly transcribed Vκs as a function of distance from e-Pol II.

(A) Perpendicular distance form each highly expressed and poorly expressed Vκ to e-Pol II cylinder measured on Imaris and plotted as percent of highly and poorly expressed Vκ at varying distance from e-Pol II.

Cyclin D3 is a general repressor of monogenic genes

In addition to repressing Vκs, cyclin D3 repressed approximately 200 other genes in pro-B cells (Powers et al. 2012). Upregulated genes in *Ccnd3*$^{-/-}$ pro-B cells were hierarchically ranked by fold increase over expression in *WT* pro-B cells and those with equal or greater than a three-fold increase were plotted (Fig 2.12**)**. Surprisingly, multiple well-known monoallelic genes were over-expressed including protocadherins (*Pcdh9*, *Pcdhb20*: shown, *Pcdhb16* : 2.5 fold, not shown)(Chess 2005), OR genes (*olfr65* shown as filled black arrows)(Monahan and Lomvardas 2015), and imprinted genes such as *Ppp1r9a* (unfilled black arrow) *Tnfrsf22*, *Htra3*, *Grb10*, *Ddc* and *Kcnq1ot1* (2.5-fold higher, not shown)(Barlow and Bartolomei 2014, Glaser, Ramsay, and Morison 2006). Furthermore, we also found *IgH* and *Tcrγ-V4* variable genes dysregulated in *Ccnd3*$^{-/-}$ pro-B cells (unfilled bars). Interrogation of the Database of Autosomal Monoallelic Expression (dbMAE) revealed that 92% of the 57 over-expressed genes were randomly monoallelically expressed in multiple tissues (Savova et al. 2016). Of these, 66% were randomly monoallelically expressed in both B cells (CD43^{-}) and other tissues (Fig 2.12, blue and red bars)(Reinius and Sandberg 2015, Nag et al. 2015). Another 28% were measured to be randomly monoallelic in other tissues (green), but were inferred as biallelic in B cells. There were only 4 genes (7%) which were biallelic in both B cells and multiple other tissues (grey). These data suggest that cyclin D3 preferentially represses monoallelically expressed genes.

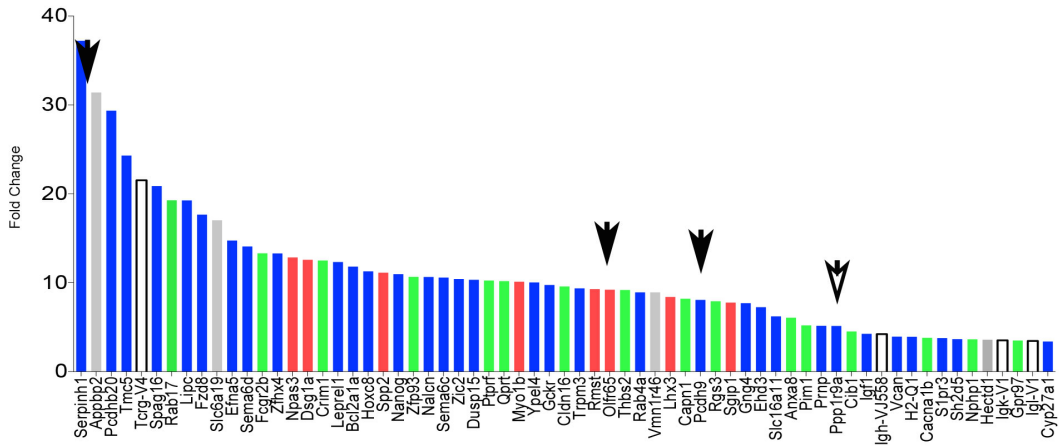


Figure 2.12 Cyclin D3 preferentially represses monoallelically expressed genes.

Upregulated genes in *Ccnd3*$^{-/-}$ pro-B cells on microarray were hierarchically ranked by fold increase over WT pro-B cells and those with equal or greater than a three-fold increase were plotted. Genes demonstrating some degree of mono-allelic expression including those demonstrated to be monoallelic in multiple tissues (dark blue), those inferred by epigenetic marks to be monoallelic in CD43^{-}B cells (red), those measured monoallelic in other tissues, but inferred biallelic in CD43^{-} B cells (green), known mono-allelically expressed genes (filled black arrows), imprinted genes (unfilled black arrow) and antigen receptor variable genes (unfilled bars). Monoallelic expression in B cells was inferred based on co-occurrence of H3K27me3 and H3K36me3 marks integrated specifically over the gene body (Nag et al. 2015). Those genes known to be biallelically expressed in multiple tissues are shown as grey bars. Gene ids were entered on dbMAE database (https://mae.hms.harvard.edu/) and tested for random monoallelic expression in multiple tissues or by chromatin signature in CD43- B cells**.**

3. Materials and Methods

Mice

WT (C57BL/6), *Ccnd3*$^{-/-}$ (C57BL/6), *Rag2*$^{-/-}$ (Balb/C) and C57BL/6 X CAST/EiJ mice were housed in clean animal facility at University of Chicago, and used at 6-12 weeks of age under IACUC protocol.

Isolation and culture/sorting of cells

Pro-B cells were isolated by positive selection ($CD19^+$) for *Rag2*$^{-/-}$ pro-B cells or by negative selection ($CD3^-CD4^-CD8^-Ter^-IgM^-CD11b^-CD11c^-Gr\text{-}1^-NK1.1^-$) for *WT* and *Ccnd3*$^{-/-}$ mice using magnetic separation (MACs) columns, and culturing them in 12ng/uL of IL-7 for 2 days. Alternatively, cells were directly FACs sorted for pro-B cells ($CD19^+B220^+IgM^-CD43^+$) and small pre-B cells ($CD19^+B220^+IgM^-CD43^-$small). Following antibodies were used for negative selection: anti-CD3e(145-2C11), anti-CD11b (M1/70),anti-CD11c(HL3), anti-NK1.1(PK136), anti-Ter119(Ter-119), anti-Gr-1(RB-8C5),anti-CD4(RM4-5),anti-CD8a(53-6.7) and anti-IgM(Fab2 Jackson. Following antibodies were used for sorting: PE- anti-CD43 (S7: BD), PE-Cy7-anti-B220 (RA3-6B2:BD), FITC-anti-IgM (II/41:BD) and APC-anti-CD19(1D3:BD).

Immunofluorescence

Cultured or sorted cells were coated on coverslip (300,000 cells per coverslip) in 1X PBS at 37^0C for 5 min. Cells were washed with CSK (100mL**:** 2mL 5M NaCl, 30mL 1M Sucrose, 2mL 0.5M PIPES (pH 6.8)) with 0.5% Triton 10 times 3 min each and fixed in 2% PFA for 10 min. Cells were blocked and stained with anti-cyclin D3 (DCS22 mAb 2936: Cell Signaling),

anti-SATB1 (ab92307: Abcam) and anti-RNA Pol II (3E10:Millipore) antibodies.

Bacterial artificial chromosomes (BAC)

Vκ RP-23 182E6), Cκ (RP-24 387E13) and Vβ (RP-23 184C1) bacterial artificial chromosomes (BACs) clones were ordered from Children's Hospital Oakland Research Institute (CHORI), as agar slab and grown in 2X YT media. The BACs were extracted and used for nick-translation reaction.

Nick Translation Reaction to make DNA probes

Nick translation protocol was derived from Jane Skok's lab. Briefly,1ug (15 uL volume) of BAC DNA was RNase-treated (0.2ug/mL) at 37^0C for 1 hr and was mixed with 5ul of 0.1M β-ME, 5uL of 0.5mM mix of dATP, dGTP and dCTP, 5uL of 10X Nick Translation Buffer (0.5M Tris-HCl pH8 (100ml of 1M Tris-HCl pH8, 50mM $MgCl_2$ (10ml of 1M $MgCl_2$), 0.5mg/ml BSA (0.1g BSA + 90ml H_2O), 0.6uL of 1mM dTTP, 1.2uL of Alexa Fluor 488-dUTP (Vκ) or Cy5-dUTP (J-Cκ) with 3uL of DNase (15U/mL) and 1uL DNA polymerase I with 14.2uL dH_20 in a total of 50uL volume, and incubated at 16^0C for 2hrs. Incubation time was optimized to get a labeled fragment of length 500bp-1kb The mix was then filtered through (Millipore, cat# VSWP02500) to get rid of unlabeled probes. 300ng (15uL from 50uL reaction) was then precipitated with 1.5uL of (10mg/uL) of salmon sperm DNA (Ambion, cat# 9680), 6uL of (1ug/uL) of mouse Cot-1 (Invitrogen, cat# 18440-16), 3uL of 0.3mM NaAc and 60uL of 100% EtOH, stored at -80^0C for 30-45 min, spinned and the pellet resuspended in 5-10 uL of Hybridization mix (200µl 50x Denhardt's solution (100ml: 1g Ficoll 400 (Fisher Bioreagents,

Biotech grade, cat# 525-5), 1g Polyvinylpyrrolidone-360 (Fisher Bioreagents, Mol Biol grade, cat# BP431-100),1g BSA), 800µl 25% Dextran sulfate in 12.5x SSPE (Sigma, Mol Biol Tested, Mw 500,000, cat# D8906-10G), 1000µl Formamide (Fisher Bioreagents, Mol Biol grade, cat# BP227-500)).

RNA probe

Vκ RNA probe (Vκ1-117) was custom designed to bind between heptamer recombination signal sequence (RSS) and 3'UTR of Vκ1-117, to specifically target germline transcript. Similarly, Jκ RNA probe was custom designed to bind 5' of Jκ1, to target germline transcript. The probes were designed by Affymetrix, and were used together with ViewRNA ISH Cell Assay Kit 24 assays (Affymetrix: Cat# QCV0001). Vκ RNA probe was purchased as 647-labeled (Type 6) and Jκ probe was purchased as 488 labeled (Type 4).

Single or combined DNA and RNA fluorescent *in situ* hybridization with immunofluorescence

Cultured or FACs sorted pro-B and small pre-B cells were plated on poly-L lycine coated coverslips at 300,000 cells per coverslip in 50uL volume of 1X PBS for 5 min at 37^0C. To get rid of soluble proteins, cells plated on the coverslip were washed with CSK(100mL**:** 2mL 5M NaCl, 30mL 1M Sucrose, 2mL 0.5M PIPES (pH 6.8)) +0.5mL of Triton (0.5%) 10x 3 min each on ice. Cells were then fixed with 2% PFA for 10min at RT. For DNA FISH+ IF, fixed cells were treated with 0.2ug/mL of RNAse for 30 min in 37^0C before treating them with 0.7%Triton/0.1M HCL for 10 min on ice. Cells were denatured in 50% formamide in 2X SSC buffer for 10 min in

80^0C, and reaction terminated by placing them on ice for 2-3 min. 5-10uL of precipitated probe (depending upon the dimension of coverslip) was denatured separately in 94^0C for 5 min and pre-annealed for 30-45 min before hybridization. Hybridization was setup on a glass slide where the coverslip was mounted over a drop of probe, and sealed along the edged with rubber cement. The slide was placed in a humidified chamber, and placed in 37^0C incubator overnight (16 hr minimum). The following day, the rubber cement was peeled off and the coverslip was washed 3X with 50% formamide/2X SSC 5 min each, and final 5 min wash with 2X SSC at 37^0C. The cells were then blocked and stained for pSer2 RNA Pol II (clone 3E10, EMD Millipore cat# 04-1571) and cycln D3 (Cell signaling, cat# DCS22 mAb 2936). If not proceeding with RNA FISH, cells were mounted in prolong gold and used for imaging. If proceeding with RNA FISH, the RNAse step was omitted. To combine RNA FISH, DNA+IF cells were incubated with 1:100 of RNA probe for 3 hrs, 1:25 pre-amplification mix for 30 min, 1:25 of amplification mix for 30 min, and finally 3 1:25 of labeled probe for another 30 min. All the incubations were done at 40^0C and each step was followed with 3X washes with the wash buffer provided in the kit. Cells were then mounted in prolong gold and used for imaging.

Imaging

Images were captured with a Leica TCS SP5 II STED laser scanning confocal microscope (Leica Microsystems, Inc.) and image processing was performed using ImageJ (Schneider, Rasband, and Eliceiri 2012). For super-resolution imaging, samples were mounted in Prolong Gold, and images captured with a Leica SR GSD 3D Ground State Depletion Microscope 3 days post mounting. Alexa-647 was depleted with 15% laser (approximately 23mW), and acquired at threshold of 10 events with 5% laser and 10% back-pump

(approximately 2mW of 405nm laser power). Alexa-532 was depleted with 100% laser and acquired with threshold of 20 with 20% laser. Distance between e-Poll and Vκ/Jκ or between Vκ and Jκ were measured using Bitplane Imaris software v. 7.3.1 (Andor Technology PLC) using Euclidean distance transformation. Measured mean distances from e-Pol II to Vκ1 and Vκ2 were paired with distances from e-Pol II to Jκ1 and Jκ2 respectively for analysis.

Distance mapping using Imaris

In order to estimate the size of e-Pol II niche around monoallelically transcribed Vκ alleles, 3D confocal images of Vκ DNA (488), Vκ 1-117 RNA (647) with e-Pol II (546) were opened with Imaris. E-Poll and Vκ DNA were displayed as volumes and spots, respectively, and Euclidean distance transformation was used to obtain mean distances from e-Pol II to Vκ alleles with and without transcript in 40-50 cells. E-Pol II-Vκ mean distances surrounding the transcribing allele and non-transcribing allele were plotted in separate columns and t test was run to evaluate statistical significance. Similar analysis was done for bi-allelically transcribed Jκ, on 3D confocal images of Jκ DNA (647), Jκ RNA (488) and e-Pol II (546). E-Pol II niche around Vκ was found to be asymmetric with transcribing Vκ allele in a tight niche of e-Pol II compared to the non-transcribing allele, whereas e-Pol II niche around Jκ was found to be symmetric.

Vκ-Jκ contraction was measured in 3D images of Vκ DNA (488) and Jκ DNA (647) and e-Pol II (546) acquired from pro-B, small pre-B and *CcnD3*$^{-/-}$ pro-B cells. On Imaris, Vκ and Jκ were displayed as spots and e-Pol II was displayed as volume. First, e-Pol II niche around *Igκ* alleles was defined based on distance of e-Pol II to Vκ alleles, not Jκ, as shown by Vκ-

DNA-RNA FISH analysis. Measured mean distances from e-Pol II to Vκ1 and Vκ2 were plotted, segregating smaller mean distance (tight e-Pol II niche) and larger mean distance (larger e-Pol II niche) into two columns, respectively. Mean distances from e-Pol II to Jκ1 and Jκ2 were paired with respective Vκ allele distances. Vκ to Jκ contraction was finally measured for both alleles (Vκ1/Jκ1 and Vκ2/Jκ2) using spot to spot distance mapping, and minimum distance was used for analysis.

Random Monoallelically expressed genes comparison to database

Genes up-regulated in *Ccnd3*$^{-/-}$ pro-B cells (Powers et al. 2012) were sorted by fold change and top 60 genes were tested for random monoallelic expression in available database for random monoallelic genes (Savova et al. 2016) and tested for random monoallelic expression in multiple tissues or chromatin signature based monoallelic expression in CD43^{-} B cells.

Single Cell RNA-seq of B6XCAST F1

First, CD19^{+}B220^{+}IgM^{-}CD43^{low} small pre-B cells were bulk sorted. This CD43^{low} gate was earlier tested and shown to have *Igκ* expression mostly in germline configuration (Fig S2.8.3). Then single cell suspension of this gate was prepared using C1TM Single-Cell Auto Prep System (Fluidigm, San Fransisco, CA), according to the manufacturer's instructions. Briefly, flow-sorted cells were resuspended at a concentration of 3×10^4 cells/ml then loaded onto a pre-primed C1 Fluidic Chip (5-10 mm). Cell separation was visually scored. Between 70 to 80 single cells was normally captured. Single cell capture on 96-well C1 was done in duplicates. Cells were lysed on chip and reverse transcription was performed using Clontech SMARTer Kit

using the mRNA Seq:RT +Amp (1771x) according to the manual. After the steps of reverse transcriptase, cDNAs were transferred to a 96 well plate and diluted with C1TM DNA Dilution Reagent. Quant-iTTM PicoGreen dsDNA Assay Kit (LifeTechnologies, Grand Island, NY) and Agilent High Sensitivity DNA Kit (Agilent Technologies (Santa Clara, CA) were used to quantify cDNAs. Libraries were prepared using Nextera XT DNA Library Preparation Kit (Illumina Inc, Santa Clara, CA) on cDNAs with an initial concentration>200 pg/µl that were then diluted to 100pg/µl. In each single-cell library preparation, a total of 125pg cDNA was tagmented at 55 °C for 20 minutes. Libraries were pooled and purified on AMPure® bead-based magnetic separation before a final quality control using Qubit® dsDNA HS Assay Kit (Life Tecnologies, Grand Island, NY) and Agilent High Sensitivity DNA Kit. scRNA Seq libraries were sequenced per HiSeq 2500 gel with 75bp paired-end sequencing.

Raw data QC and genome alignment

Raw fastq data was quality trimmed to a minimum Phred score of 20 using trimmomatic (Bolger, Lohse, and Usadel 2014). Reads were then filtered against mouse ribosomal sequences using bowtie2 (Langmead and Salzberg 2012), followed by full genome and transcriptome alignment to mouse reference mm10 using STAR (Dobin et al. 2013). Apparent PCR duplicates were removed using Picard MarkDuplicates (http://broadinstitute.github.io/picard) as were reads that mapped to multiple locations in the genome.

Allele assignments

Variant for CAST/EiJ mice were obtained from release-1505-GRCm38 from Sanger Institute's ftp site (ftp://ftp-mouse.sanger.ac.uk/). The SNPs are based on published sequence-based characterization of multiple mouse genome (Keane et al. 2011, Yalcin et al. 2011). Allelic assignment was limited to only *Igκ* locus. SNPs with quality score of at least 100, reported by Sanger, were retained for further analysis. A coverage threshold of 30 per SNP per expressed Vκ segment was used for allelic assignment. A particular SNP was given more confidence if the SNP could be used to clearly assign same Vκ gene in a different cell. Reads that matched the reference sequence were assigned as B6, reads that matched the alternate sequence were assigned CAST, and reads that matched neither were considered unassigned. If the SNPs varied in assignment, that segment was considered ambiguous. After confirming allelic assignment for a particular V segment with overall coverage threshold of 30 in one or more cells, a lower threshold of 10 per cell was used to incorporate all low-medium expressed Vs. Any SNP with an allele distribution that was at least 92% B6 or CAST was considered monoallelic for that strain otherwise they were considered biallelic. To ensure monoallelic expression of gene at single cell level and not due to technical issues in detecting the alternate allele, we performed bulk RNA-seq on B6X CAST small pre-B, which confirmed Vκ and Jκ expression from both B6 and CAST allele (Fig S2.8.2).

Distal Jκ promoter (3.5kb upstream of Jκ1) was split into 5'Jκ promoter 1 (5'Jκp1) and 3'Jκ promoter 2 (3'Jκp1), to avoid ambiguity in allelic assignment. In most instances, 5' end of the distal promoter was found to be biallelic whereas 3' end was found to be either monoallelic

or showed no expression. Monoallelic or no expression of 3' end of the distal promoter is possible if the splice sites are in between 5' and 3' end of the promoter. Proximal promoter (100bp upstream of Jκ1) was called Jκp2 for allelic assignment.

RSS Motif analysis

Unrearranged cells were discriminated by biallelic expression of distal 5'Jκ1 promoter and/or proximal promoter (Jκp2), presence of heptamer/nonamer RSS motif in the J segment reads and absence of recombination products. Reads were searched for 5' [GA]GTTTTTGT (nonamer) and 3' CA[CG]TGTG (heptamer) with 23 bp spacer in the sense strand.

Rearrangement analysis

Read alignments were split into separate bed regions based on their CIGAR assignment for splicing (N in CIGAR strings) using bedtools bamtobed (Quinlan and Hall 2010). V and J segments were annotated using bedtools intersect to record all of the reads that overlapped with each V and J segment. For each read ID, different annotated V and J junctions were counted to get V-V, V-J, and J-J splicing/recombination. For visualization purposes, only segment-to-segment splicing/recombination that was observed at least 10 times were reported. This method does not discriminate between V-J recombination at the DNA level with V-V, J-J or J-C splicing occurring at the mRNA level. In-frame or out-of frame rearrangement were tested by using NCBI's IgBLAST.

Polymer chain simulations

Coarse-grained Langevin dynamics was used to simulate chains of identical monomers of mass $m = 1$, and radius $\sigma = 1$. $Ig\kappa$ polymer of total contour length $l = 3.74\ Mb$ was simulated with each monomer representing $dl = 1\ kb$ of DNA. Successive monomers in the chain interacted via stiff harmonic potential described by: $E_{bond}(r) = K\ (r - r_0)^2$, with a rest length $r_0 = 1/\sqrt{2}$ (distance units) and stiffness $K = 500$ (energy units). Excluded volume interactions between monomers, separated by a distance r, were implemented using a Weeks-Chandler-Anderson (WCA) potential (Weeks, Chandler, and Andersen 1971).

$$E_{WCA}(r) = \begin{cases} 4\varepsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} + 1/4\right), & r < r_c \\ 0, & otherwise \end{cases}$$

with $\sigma = 1$ (distance units), $\varepsilon = 1$ (energy units) and cut-off distance of $r_c = 1.12$. Any 3-body bending forces or 4-body torsion-angle forces were not incorporated. In addition, the monomers interacted via pairwise CTCF-based looping interactions, as described below.

Looping interactions. To induce the formation of CTCF-based loops a pairwise Morse potential (Morse 1929) weighted by a probability of looping, p_{ij}, between monomer positions *i* and *j* was used.

$$E_{loop} = \sum_{|i-j|>1} p_{ij}\ \varepsilon_{loop}\left[e^{-2\alpha(r_{ij}-r_0)} - 2e^{-\alpha(r_{ij}-r_0)}\right]$$

where $\varepsilon_{loop} = 10$ (energy units), $\alpha = 2$ (1/distance units), and a cut-off distance of 10. The looping probability depends on the orientation of CTCF motifs (Guo et al. 2015) as well as their binding strengths determined from CTCF ChIP-seq signal (Choi et al. 2013). Following Ref.

(Sanborn et al. 2015), the CTCF ChIP-seq signal S_i (binned at 1 kb) was converted into a probability of binding defined as: $b_i = 1 - \chi/(S_i - S_m)$, for $S_i > S_m$ and $b_i = 0$ otherwise, where S_m is the median ChIP-seq signal and χ is a normalization constant. Oriented looping probabilities, p_{ij}, between two loci *i* and *j*, is given by the product of binding probabilities times a constant, κ_{ij} that is dependent on the site orientations: $p_{ij} = \kappa_{ij} b_i b_j$, with [3]

$$\kappa_{ij} = \begin{cases} 0.8\,, & Forward(F) - Reverse(R) \\ 0.1\,, & F - F\ or\ R - R \\ 0, & otherwise \end{cases}$$

Wall-monomer interactions. In the confined chain simulations, ePol-II was treated as a cylindrical bounding wall, which interacts with nearby monomers via a repulsive 12/6 Lennard-Jones potential with a cut-off distance of 1.

Simulation runs and parameters. The polymers were initialized in their fully extended configurations and the equations of motion were integrated using Langevin dynamics (NVT ensemble), with a time step of 0.005 in units of τ, where $\tau = \sqrt{m\sigma^2/\varepsilon}$. We fix the number of monomers $N = l/dl$, temperature $T = 1$ (in units of ε/k_B), and a damping coefficient: $\gamma = 1/\tau_d$, with τ_d=5 (in units of τ). Each simulation run consisted of 2,000,000 time steps to ensure full collapse. The simulations were run using LAMMPS (Plimpton 1995) and the data were post-processed in Matlab.

Collaborator's contributions

Imaging and distance mapping was done by S. Karki, with *in situ* Hybridization techniques learned from E.Davis from Le Beau's laboratory using Jane Skok protocol. Bulk sorting for single cell sorting was done by S.Karki and H.Xu, and single cell sorting on C1 was performed by Cincinanti Children Hospital Genomic core in collaboration with Singh laboratory. Single cell RNA seq data processing including variant assignment and V-J splice junction identification was done by M-Cline. 3D simulation was done by S.Banerjee from Dinner's laboratory using distance and Vκ expression data provided by M. Mandal and S.Karki from Clark laboratory.

Funding and resources

Work was supported by US National Institutes of Health (AI120715, AI082724 and GM101090). Imaging work was done at the Integrated Microscopy Core Facility supported by Cancer Center Support Grant P30 CA014599. Single cell Sequencing data analysis was supported in part by the National Center for Advancing Translational Sciences, National Institutes of Health, through grant UL1TR000050, affiliated with Center for Research Information (CRI) at University of Illinois at Chicago (UIC). Polymer chain simulation was completed in part with resources provided by the University of Chicago Research Computing Center.

4. Discussion

Our confocal and super-resolution imaging of *WT* pro-B cells support earlier findings that a large number of RNA Pol II molecules are assembled as immobile foci (transcription factories) on the nuclear matrix (Kimura et al. 1999). However, since we visualized these factories in fixed pro-B and small pre-B cells, we do not know if they are pre-assembled or dynamically assembled between stages. Dynamic clustering of RNA-Pol II has been shown by live-cell imaging for heat-shock protein expression (Mitchell and Fraser 2008, Ghamari et al. 2013), whereby transcription initiation was necessary for transcription factory assembly. Factories were prone to disintegration during transcription initiation, however, once they were elongating were stable. In the context of *Igκ* , dynamic assembly model predicts that asymmetric e-Pol II niche is established after entry into small pre-B stage, where *Igκ* transcription is initiated. Our data favors pre-assembled model, as e-Pol II niche was already apparent in *WT* pro-B cells in the absence of transcription. However, this finding does not rule out the possibility that new e-Pol II foci are assembled around *Igκ* in small pre-B cells.

Discriminating between pre-assembled (stable) and dynamic assembly of transcription factories around *Igκ* in small pre-B cells is important as it will have consequences on how we model monoallelic Vκ transcription. If e-Pol II assembly were dynamic between stages, it would suggest that allelic choice is dependent upon stochastic transcription initiation events that are more likely to happen in one allele at a time. One way to address whether these factories are pre-assembled or dynamic is to use an inducible system such as Abelson transformed pre-B or

IRF4/8 pre-B cell line, where a fluorescent reporter can track e-Pol II elongation after induction. Red Fluorescent Protein (RFP) fused to CDK9, a kinase that specifically associates with and phosphorylates elongating RNAPII, has been successfully used in live-cell imaging to mark e-Pol II foci or transcription factories (Ghamari et al. 2013). As published for *IgH* (Lucas et al. 2014), Vκ alleles can be marked by introducing Tet-operator binding sites downstream of some of the highly expressed Vκ segments in a pre-B cell line and then transducing these cells with Tet-repressor-EYFP construct, which will bind and mark the V segments associated with the operator sites. Using this system will not only allow us to control the transition from proliferative pro-B to small pre-B cells, but also track assembly of factories around marked Vκ genes in real-time. Furthermore, inducing entry into small pre-B cells in the presence or absence of transcription initiation or elongation inhibitors, will provide a broader understanding of how transcription initiation and elongation events affect transcription factory assembly.

Stable versus dynamic assembly of e-Pol II will also have implication on receptor editing. In our single cell RNA-seq data, we saw multiple instances of secondary rearrangement in the second (excluded) *Igκ* allele and initiation of germline transcription of *Igλ* alleles, which is indicative of receptor editing. Dynamic assembly of RNA polymerase II factories in small pre-B cells would mean that upon out-of-frame rearrangement of first (chosen) *Igκ* allele or expression of an autoreactive BCR, e-Pol II assembles around second (excluded) *Igκ* or *Igλ* alleles to initiate transcription. On the other hand, stable pre-assembled factories would mean that the alleles are already poised to associate with these factories. These two models can be tested by using a mouse model that is prone to high levels of receptor editing, expresses RFP fused to

CDK9 as marker of transcription factories, and uses TetO-R-EYFP-like system to mark *Igκ* and *Igλ* alleles. Mice that have B cells expressing *IgH* and *IgL* transgenes encoding an anti-H-2K^k or H-2K^b undergo high levels of receptor editing when bred with mice expressing the corresponding MHC haplotype H-2K^k or the H-2K^b, and therefore can be used for this purpose (Tiegs, Russell, and Nemazee 1993). Culturing pro-B cells from these mice in presence of IL-7, e-Pol II assembly around *Igκ* and *Igλ* alleles can be tracked upon IL-7 withdrawal, which will enforce cells to undergo *IgL* rearrangement and receptor editing.

Our asymmetric niche model predicts how initial monoallelic and monogenic choice for *Igκ* is made, however, does not address how the model fits with receptor editing. From our model, the allele that is not chosen for initial rearrangement is located in an e-Pol II depleted niche, which is not permissive for efficient transcription and rearrangement. However, the fact that the second (excluded) *Igκ* allele is targeted for secondary rearrangement by receptor editing mechanism is not consistent with assumptions of our model. It is likely that we evaluated e-Pol II niche in pro-B and small pre-B where receptor editing is not substantial. Since secondary rearrangement happens in late small pre-B and immature B cell stage, perhaps niche dimensions should be measured in these stages, where we may not detect substantial asymmetry between e-Pol II niche surrounding the two Vκ alleles or *Igλ* alleles. It is likely that pre-existing asymmetric e-Pol II niche has a role in initial allelic choice, but there is requirement for dynamic assembly of transcription factories around other *Igκ* and *Igλ* alleles to enable receptor editing. These ideas need to be tested with more analyses of niche dimensions in different developmental stages either in fixed cells or in dynamic system mentioned above.

We have shown that nm-cyclin D3 has a repressive role in *WT* pro-B cells. But how cyclin D3 imparts its repression in pro-B cells is still unclear. Earlier, we reported that cyclin D3 does not bind chromatin (Powers et al. 2012), which suggests that cyclin D3 mediated repression of Vκ and neuronal genes is indirect. We visualized cyclin D3 intertwined with e-Pol II and SATB1 in *WT* pro-B cells, implying that cyclin D3 acts in concert with these and other unknown proteins to mediate repression. SATB1 appears to be a particularly good candidate for this role. Unlike nm-cyclin D3, SATB1 directly binds DNA, in particular specialized ATC sequences and tethers them to the nuclear matrix. Furthermore, similar to *WT* pro-B cells, SATB1 is distributed in thymocyte nuclei forming cage-like structure, where it not only organizes chromatin folding but also establishes specific histone modifications over the region where it binds. For example, in the *Il2ra* locus, SATB1 recruits histone deacetylase to SATB1 binding sites (SBS) (Yasui et al. 2002), and thereby contributes to repression of the locus. Interestingly and consistent with our observation in *Ccnd3*$^{-/-}$ pro-B cells, SATB1 target genes also include neuron-specific genes. These genes are properly repressed in wild-type thymocytes but ectopically transcribed in *Satb1*$^{-/-}$ thymocytes. Comparing expression profile of *Satb1*$^{-/-}$ thymocytes and *Ccnd3*$^{-/-}$ pro-B cells may identify genes that are co-repressed by cyclin D3 and SATB1, and may provide insight into their cooperative mode of repression.

E-Pol II also binds cyclin D3 in *WT* pro-B cells, and may be indirectly involved in gene repression. SATB1, which is part of cyclin D3-e-Pol II complexes in pro-B cells has been shown to bind a variant of RNA pol II, subunit 11 (RPB11) (Durrin and Krontiris 2002). However, whether this variant of RNA-pol II renders it repressive in not known. Since e-Pol II is common to both repressive complexes in *WT* pro-B and activated complex in *WT* small pre-B cells or

Ccnd3$^{-/-}$ pro-B cells, e-Pol II immunoprecipitation (IP) of nuclear matrix extract from these cells, using published protocol (Staufenbiel and Deppert 1984), may highlight the differences between these presumed repressive and active complexes. Mass-spectrometry analysis of fraction that is unique to *WT* pro-B cells may help identify proteins that are involved in repression. Identified proteins can then be tested for *Igκ* repression by transducing them in *Ccnd3*$^{-/-}$ pro-B cells.

Earlier study from our lab showed that nuclear matrix-localization of cyclin D3 was specific to pro-B and small pre-B cells but not fibroblasts, although fibroblasts did show similar expression of the protein (Powers et al. 2012). This result suggests that expression of cyclin D3 alone cannot be correlated to nuclear matrix-localization and therefore its ability to repress transcription. It is likely that cyclin D3 mediated repression is cell type or lineage specific. A thorough analysis of different cell types may allow us to understand the nature of these cells or lineage and the significance of matrix localization of cyclin D3 in those cells. Neuronal cells are likely candidates, as a significant number of neuronal genes were dysregulated in *Ccnd3*$^{-/-}$ pro-B cells(Powers et al. 2012). Cyclin D3 has been shown to be expressed in differentiating motor neurons and dorsal interneurons (Salles et al. 2007), however, their matrix localization and transcriptional regulation of neuron specific genes is yet to be tested.

No post translational modifications of cyclin D3 has been reported to date that would help explain compartmentalization of cyclin D3 in soluble and nuclear matrix-bound fraction. In an earlier study from the lab, we speculated that different regions of cyclin D3 protein maybe attributable to these two cellular functions. Therefore, chimeric constructs containing either the N- or C-terminal 1/3 of cyclin D3 paired with cyclin D2 for the remaining protein were made

and transduced in *Ccnd3*$^{-/-}$ pro-B cells, and Vκ repression was assessed by qPCR. We found that C-terminus, bur not N-terminus (contains cdk4/6 binding motifs) of cyclin D3 was involved in transcriptional regulation. However, the results were not drastic, suggesting that there might be additional feature(s) of cyclin D3 or of the two fractions themselves that can explain the dual compartmentalization. It is likely that alternative partner in the soluble fraction and matrix-bound fraction help maintain this dual localization. As hypothesized, soluble cyclin D3, and not the nuclear matrix bound cyclin D3, binds cdk4/6(Powers et al. 2012). As an extension of this idea, only those cells with appropriate matrix bound partner would show cyclin D3 localization in the matrix. Performing e-Pol II IP using nuclear matrix extraction protocol on fibroblast and pro-B cells may allow to see if composition of nuclear matrix is fundamentally different between the two cell types.

In addition to *Igκ*, *Ccnd3*$^{-/-}$ pro-B cells showed dysregulation of genes of neuronal origin, suggesting that in pro-B cells, nm-cyclin D3 could have a role in repression of non-lineage genes. Interestingly, the neuronal genes dysregulated in *Ccnd3*$^{-/-}$ pro-B cells show ectopic expression, and do not show continued dysregulation in *WT* small pre-B (Heng, Painter, and Consortium 2008), where cyclin D3 levels are down-regulated. A probable hypothesis is that nm-cyclin D3 in *WT* pro-B cells is a part of a protein cascade that represses non-lineage genes, and therefore upon entry into small pre-B cells, *Igκ* is expressed due to down-regulation of cyclin D3 and presence of small pre-B specific factors, whereas other non-lineage genes are repressed back by downstream members of the cascade due to lack of neuron-specific factors to sustain their expression. Consistent with this hypothesis, cyclin E expressed later in cell cycle

also has nuclear matrix bound fraction (Munkley et al. 2011). If nm-cyclin D3 was initial member of the cascade, its absence would result in global relief of repressed genes, including both *Igκ* and neuronal genes, as observed in *Ccnd3*$^{-/-}$ pro-B cells. Nuclear matrix bound proteins that specifically associate with nm-cyclin D3 in *WT* pro-B cells can be transduced in *Ccnd3*$^{-/-}$ pro-B cells, to test if they recapitulate nm-cyclin D3 mediated repression of neuronal genes.

Other V genes, including *IgH* and *Tcrγ* V genes were also de-repressed in *Ccnd3*$^{-/-}$ pro-B cells. Like *Igκ*, both *IgH* and *Tcrγ* have diverse V gene families lying either upstream of a (D)J cluster (*IgH*) or interspersed with Js (*Tcrγ*) (Lefranc et al. 2005, de Almeida, Hendriks, and Stadhouders 2015, Ebert, Hill, and Busslinger 2015). These similarities suggest that monoallelic choice at these loci may also be determined by V accessibility. As with pro-B cells, cyclin D3 is responsible for proliferation of pro-T cells (Sicinska et al. 2003). However, instead of IL-7R, cyclin D3 is expressed downstream of pre-TCR. It is likely that *Ccnd3*$^{-/-}$ pro-T cells also prematurely express *Tcrγ* variable genes, which can be assessed by performing single cell RNA-seq on these cells. If *Igκ* and *Tcrγ* were both regulated by nm-cyclin D3, this would highlight a novel mechanism of transcriptional regulation and allelic choice of AgR by matrix-bound proteins.

Apart from OR and protocadherin genes that undergo monoallelic rexpression, as high as 15-20% autosomal genes show random monoallelic expression (RME) (Chess 2013). To date, two categories of RME genes have been identified, one termed dynamic RME genes that show biallelic expression early during development, but show skewing towards monoallelic

expression after entry into more differentiated stage, and another termed fixed RME genes that show consistent random monoallelic expression and clonal maintenance of this signature throughout development(Deng et al. 2014, Eckersley-Maslin et al. 2014). Due to "randomly monoallelic" nature of this category of genes, some cells in a given tissue will express these genes from maternal or paternal allele and show clonal maintenance, amongst other cells where the genes will be expressed biallelically. Furthermore, RME expression can occur in the absence of single nucleotide polymorphisms(Gendrel et al. 2014) at the two alleles, suggesting that RME expression is not a result of differences in transcription efficiency that can result from SNPs located in important regulatory regions.

Majority of genes dysregulated in *Ccnd3*$^{-/-}$ pro-B cells were RME genes, however, whether they show monoallelic expression throughout earlier developmental stages such as in case of fixed RME genes or acquire monoallelic expression during differentiation such as in case of dynamic RME genes is not known. Understanding fixed or dynamic RME nature of genes dysregulated in *Ccnd3*$^{-/-}$ pro-B cells will help understand if there is an obvious skewing towards one category versus another. Since dynamic RME has been specifically noted for genes expressed in differentiated stages (Eckersley-Maslin et al. 2014), we speculate genes dysregulated in *Ccnd3*$^{-/-}$ pro-B cells, to be dynamic RME genes. One way to test RME nature of these genes is to perform single-cell RNA seq on CLPs, pre-pro-B and pro-B cells of divergent cross between B6x CAST mice deficient in cyclin D3, and perform SNP analysis to identify allele-specific expression of these genes in these stages. Since monoallelic or biallelic nature of genes is clonally maintained, multiple clones can be expanded from each stage and tested for allele-specific expression of candidate genes (Eckersley-Maslin et al. 2014) by RNA-FISH. If

these genes were dynamic RME genes, we would see pre-dominant biallelic expression in CLPs and pre-pro-B stage but increasingly monoallelic expression in *Ccnd3*$^{-/-}$ pro-B stage. Since most of these genes are neuronal genes, testing neuronal lineage for their mono- versus biallelic expression would further solidify their RME nature.

Random autosomal monoallelic expression is a recent area of study, and therefore it is not known if monoallelic expression is always downstream of nm-cyclin D3 or other D-type cyclins. It is unlikely this is the case since D-type cyclins are expressed in specific stages of development and in specific tissues, but RME expression has been noted in multiple tissues. It would be interesting to see how fixed and dynamic RME is associated with expression pattern of D-type cyclins in various tissues.

Cyclin D3 repressed ~250 genes in pro-B cells (Powers et al. 2012), which is only a subset of RME genes reported (Deng et al. 2014, Savova et al. 2016). This suggests that there are additional proteins involved in transcriptional regulation of these genes. It would be interesting to see if regulation of RME genes by nm-bound proteins is a general mechanism of transcriptional regulation. All proteins identified to bind nuclear matrix currently are reported as part of a database https://rostlab.org/services/nmpdb/index.html, created by Sven Mika at Rost Group in Columbia University, New York. The database includes information about nm-bound protein names, their organism and the cell-type in which nm association was observed, and therefore may provide a valuable tool to delve into potential role of these proteins in transcriptional regulation of RME genes.

Based on our findings of asymmetric e-Pol II niche around Vκ gene segments, it is possible that asymmetric niche surrounding monoallelic genes is a general rule, established early on during development in the case of fixed RME genes or set-up during differentiation for dynamic RME genes. DNA FISH of genes that hold up to their RME potential by single-cell RNA seq and clonal analyses, together with e-Pol II staining can be used to test this hypothesis. If true, we expect any gene that shows fixed RME from early developmental stage to later stages, will maintain asymmetric e-Pol II niche all throughout development. However, any gene with dynamic RME will only show asymmetric niche in later differentiation stages.

V segments of *Tcrγ* chain but not *Tcrα* or *β* was dysregulated in *Ccnd3*$^{-/-}$ pro-B cells. Interestingly, Vβ and Vα show biallelic transcription (Jia, Kondo, and Zhuang 2007, Ranganath et al. 2008) and ~30% biallelic in-frame rearrangement, respectively (Melchers 2015). This observation is consistent with our hypothesis that cyclin D3 primarily represses monoallelic genes. It would be interesting to test if e-Pol II niche for Vβ and Vα genes is symmetric and therefore contributes to their biallelic nature. If true, this would mean that asymmetric niche is characteristic of monoallelic genes.

Apart from AgR genes, protocadherin and OR genes are also prototype for monoallelic expression (Chess 2005, Monahan and Lomvardas 2015). Olfactory receptors are encoded by a diverse family of receptor genes (approximately 1400 in mouse)(Monahan and Lomvardas 2015) that are expressed monoallelically and monogenically in terminally differentiated neurons. Protocadherin genes are expressed by alternative promoter choice of 5' variable gene segments

that is then spliced to one of the constant segments in the 3' region (Monahan et al. 2012). Likewise, similar to *Igκ,* OR cluster is expressed as a read-through transcript that then associate with constant exon by splicing. A striking similarity of these genes and AgR genes is that all three categories of genes are clustered and are confined within topological domains defined by CTCF sites (Chen and Maniatis 2013). These similarities between protocadherin, OR and *Igκ* genes suggest that these gene clusters may have evolved similarly, and in a way that allowed monoallelic and monogenic choice from multi-genic loci. Our comparison of AgR with OR and protocadherin genes for monogenic choice is valid, however, AgR undergo recombination while the other two category of genes do not. To enable comparison among these categories of genes, single cell RNA seq should be performed on pro-B cells from B6 X CAST F1 hybrids that are doubly deficient for cyclin D3, which would express V genes from *Igκ* without recombination.

We noticed that both Vκ and Jκ colocalized to the same transcription factory in *WT* small pre-B cells, which suggests that their colocalization marks potential ongoing recombination. However, co-occupancy or juxtaposition of Vκ and Jκ within a factory was not necessary for initial accessibility, since Vκ transcription did occur in the absence of juxtaposition with Jκ in *Ccnd3*$^{-/-}$ pro-B cells. This data suggests that initial Vκ and Jκ transcription may occur in different transcription factories, independent of each other. Our single cell RNA seq data and Vκ-Jκ 2-color DNA FISH did not provide information on transcription and contraction simultaneously. For example, in our single cell RNA seq data, we saw transcription of multiple V loops, however, we lacked information about their proximity to Jκ. Similarly, our Vκ-Jκ 2-color DNA FISH captured transcription and contraction of single Vκ loop in small pre-B cells, however,

lacked information about whether other V loops were active and juxtaposed to Jκ in the same cell. By using BAC probe specific to multiple V loops and Jκ may allow us to precisely understand how multiple V loop activation ties with juxtaposition with Jκ and co-occupancy with e-Pol II factories.

Cyclin D3 repressed both Vκ transcription and Vκ-Jκ contraction. It is likely that the two are related as transcription has been implicated in locus contraction (Corcoran 2010, Verma-Gaur et al. 2012). Interestingly, Vκ-Jκ contraction in *Ccnd3*$^{-/-}$ pro-B cells was only partial on the transcribing Vκ allele, suggesting that additional mechanisms are required for complete contraction. These may include Jκ transcription, which is absent in these cells and stable capture of Vκ loops by RAG or other components of the recombination machinery, such as BRWD1 (Mandal et al. 2015, Schatz and Ji 2011). One way to test if initial contraction is dependent upon Vκ transcription is to culture CD19^{+}B220^{+}IgM^{-}CD43^{low} small pre-B cells, where Vκ undergoes initial transcription, in presence of elongation inhibitor and perform Vκ-Jκ 2-color DNA FISH and distance measurements on the acquired images. The alternative hypothesis that contraction leads to transcription could also be true, as Vκ-Jκ in *Ccnd3*$^{-/-}$ pro-B cells may have reached a critical contraction distance necessary for transcription. This hypothesis can be tested by performing Vκ-Jκ 2-color DNA FISH on BRWD1 mutant CD19^{+}B220^{+}IgM^{-}CD43^{low} small pre-B cells, which show drastically reduced Vκ transcription (unpublished). Reduced Vκ-Jκ contraction frequency in BrWD1 small pre-B cells will mean that contraction has a potential role in transcription.

We detected a subset of Vκ accessible per cell. It is likely that every subset of Vκ relies on a dominant TF for its expression. There is a strong correlation between E2A binding and transcriptional activation of Vκs (Stadhouders et al. 2014). Since E2A binding to Eκi is important for Jκ accessibility, it is possible that it also has important role in Vκ activation. Targeted mutation of E2A binding sites leads to highly reduced germline transcription at Jκ and recombination efficiency ((Inlay et al. 2004), and ectopic E2A expression in non-lymphoid cell together with RAG expression is enough to induce *Igκ* recombination (Romanow et al. 2000). Similar to E2A, other TFs such as Ikaros, PU.1, PAX5, EBF1 also bind subset of Vκ(de Almeida, Hendriks, and Stadhouders 2015), and may have a role in their transcription and usage. It is likely that germline transcription of particular V segments is dependent upon whether they land the transcription factory with their designated TF(s), since there is some evidence that transcription factories are loaded with transcription factors that can coregulate multiple genes (Sutherland and Bickmore 2009). To address this query, DNA probes spanning multiple Vκ loops can be used in combination with e-Pol II and TF staining. Although immunoFISH is not high throughput, analysis can still be done at single cell level, and therefore maybe informative. Furthermore, ectopic expression of specific TF alone or in combination with other TFs and RAG in non-lymphoid cells can allow us to test if they are sufficient for recombination. It is likely that Vκ repertoire due to ectopic expression of one TF is different than expression of another TF.

We saw multiple Vκ segments expressed per cell, however, the number of Vs expressed was variable between cells, with numbers ranging from 1 to 12 per cell. We have shown that nm-cyclin D3 down-regulation is important for Vκ expression, however, whether Vκ expression co-

relates with the level of cyclin D3 decline has not been tested. It is likely that Vκ expression starts with few Vs and progressively increases in number as the level of cyclin D3 decline. This correlation can be tested by plotting cyclin D3 expression levels against number of Vκ expressed within single cells. Alternatively, probing for expression of multiple V genes using DNA-RNA FISH and assessing intensity of nm-cyclin D3 localized in the nuclear body per cell, would also allow us to test this hypothesis.

The finding that multiple V segments are accessible per cell refutes notions about single accessible V segment leading to single recombination event. We noticed 1-3 highly used Vs accessible per cell, which suggests a competitive model for recombination. We propose that only V loops with highly used Vs have high probability of synapse formation with Jκ and recombination. However, from our single cell RNA-seq data, we could not assess this competition visually. For example, we could not say if all accessible V loops formed synapse with Jκ. We saw that even though both distal and proximal Vs were accessible per cell, only proximal Vs underwent recombination, suggesting that synapse preferentially forms with proximal V loops in $CD43^{low}$ cells. It is likely that in these cells, distal Vs were accessible but were not juxtaposed with Jκ. Alternatively, it is also likely that both distal and proximal Vs were both accessible and juxtaposed with Jκ, however, complete engagement with Jκ only happened with proximal Vs. The former scenario is more probable because we already know from *Ccnd3*$^{-/-}$ pro-B contraction data that distal V loop accessibility does not equate to juxtaposition with Jκ.

Understanding how transcriptional accessibility correlates to assembly over Jκ is important to understand how multiple accessible loops compete for recombination. One way to address these questions is to use BAC probes spanning each of the V loops and test if there is simultaneous engagement with Jκ in fixed small pre-B cells. Furthermore, cells can be stained for γ-H2AX, a H2A histone variant phosphorylated by ataxia telangiectasia mutated (ATM) kinase during RAG-mediated double stranded breaks (Kuo and Yang 2008), to mark loop (s) that succeed in the competition. It is likely that more than one accessible V loops are cleaved by RAGs, but only one is joined with Jκ.

Positioning of Vκ loops with respect to Jκ appears to help determine their transcription and usage. For example, both by DNA FISH and 3D simulation, intermediate loop extended outward and away from Jκ, while distal and proximal Vκ loops were juxtaposed to Jκ(Roldan et al. 2005). Remarkably, while distal and proximal Vs are highly transcribed and used, intermediate loop is both less transcribed and less used for recombination, despite having similar promoter elements and other features comparable to distal and proximal V genes (Martinez-Jean, Folch, and Lefranc 2001, Aoki-Ota et al. 2012). This finding suggests that their poor transcription and usage maybe associated with how they are folded and positioned with respect to transcription and recombination factories. One way to test whether position of intermediate loop renders it inaccessible would be to switch the location of distal or proximal V loops with intermediate loop using CRISPR/Cas9 mediated genome editing in pre-B cell line, and test by V loop specific DNA FISH and e-Pol II staining, if intermediate loop now becomes transcriptionally and recombinationally active. The alternative should also be true, where distal

or proximal loops now become less transcribed and less used.

There is some level of confusion associated with the nature of Vκ expression per cell. From our single cell RNA seq data, we saw that germline transcription at single Vκ segment was always monoallelic. However, we did notice that other non-homologous Vκ segments were expressed from the alternate allele in a single cell. This creates misperception, as to whether Vκ gene segments should be treated as a single locus or separately for each Vκ segments. If *Igκ* locus is treated as one, the fact that we notice sporadic transcription at another allele makes Vκ accessibility biallelic in a fraction of cells (14%). However, since transcription never happened biallelically from a homologous region, transcription should still be monoallelic at single V segment.

Read-through Vκ-Vκ transcription may have a role beyond Vκ accessibility. Single-cell RNA seq data revealed a lot of read-through transcripts that do not participate in making rearranged *Igκ* transcript. Since these transcripts span multiple V segments, with some >200kb in size, they may belong to the category of long non-coding RNA, which have been associated with many functions including, imprinting and epigenetic regulation (Cao 2014). Of note, one of the imprinted genes, *Kcnq1ot1*, dysregulated in *Ccnd3*$^{-/-}$ pro-B cells is a well characterized long non-coding RNA that is involved in epigenetic silencing of adjacent imprinted gene cluster (Redrup et al. 2009). Although silencing by *Kcnq1ot1* occurs in *cis*, mechanisms in *trans* also exist. In particular, anti-sense transcripts have been shown to be able to silence genes in *trans* (Werner 2013). In the context of AgR, *trans* silencing, if true, is more likely, since *IgH* chain has been

shown to undergo anti-sense transcription in pro-B cells (Bolland et al. 2007). Since active *Igκ* allele is associated with recombining allele and not excluded allele, suggests that germline transcription maybe involved in *trans* silencing. However, we have not evaluated whether Vκ read-through transcription is anti-sense in orientation in our data, and therefore needs to be tested. If *trans* silencing by non-coding Vκ transcripts were true it would provide an additional mechanism of reinforcing monoallelic expression besides asymmetric e-Pol II niche.

One way to test if expression of these non-coding Vκ transcripts has a role in monoallelic expression of Vκ and repression in *trans*, is to generate multiple clones from $CD19^{+}B220^{+}IgM^{-}CD43^{low}$ single cells from B6X CAST F1 mice. Since every clone is expected to express Vκ transcripts from single allele, knocking down B6-specific Vκ transcription (say) from a clone using B6-specific siRNA, possible expression from CAST allele can be tested. Although uncommon, variant specific siRNA has been successfully designed and used in *in vitro* system (Birmingham et al. 2007). If no transcript were detected from the CAST allele, it would suggest that these germline Vκ transcripts are not directly involved in *trans* silencing and that e-Pol II depleted niche posses a barrier for transcription of CAST allele, as we proposed in our model. If however, transcripts were detected from the CAST allele, it would mean that that e-Pol II depleted niche alone is not sufficient to pose accessibility barrier for the alternate allele, and that these non-coding transcripts maybe involved in reinforcing repression of the alternate or excluded allele.

For monoallelically recombined locus such as *Igκ,* it is important to distinguish differential marking of the active allele from the silenced allele in an allele-specific manner. Therefore, analyses done in bulk cells that are heterogeneous, may lead to erroneous interpretations. Recently, DNA methylation levels at transcriptionally active Vκ gene segments were shown to increase from pro-B to small pre-B cells by bulk bisulfite sequencing (Orlanski et al. 2016). However, due to lack of discriminatory power, whether DNA-methylation marked the silenced allele of transcriptionally active V genes or the active allele itself could not be confirmed. Based on known role of DNA methylation in marking repressed genes, we propose that in this case, DNA methylation marks the silenced *Igκ* allele. Our data indirectly supports this hypothesis, as we saw active Vκ gene promoters bound by CTCF sites, which only bind unmethylated gene promoters (Davalos-Salas et al. 2011). One way to test if DNA methylation is specific to silenced allele, is to make multiple pre-B clones from divergent F1 cross of B6 x CAST mice, identify active Vκ allele using allele-specific SNPs and perform restriction digest with methyl sensitive enzyme and allele specific sequencing to identify the methylated allele, as has been published for Jκ (Farago et al. 2012).

Our lab and others have shown that Vκ genes are not marked by canonical histone marks (Xu and Feeney 2009, Mandal et al. 2011). However, we do not know if this is a result of bulk analysis of small pre-B cells, which can underestimate gene-specific modifications that maybe visible only at single cell level. Therefore, one way to address this question is to perform single cell ChIP on $CD43^{low}$ small pre-B cells for activating histone marks, using published single cell ChIP protocol(Rotem et al. 2015) and test if histone marks now become apparent. Another way

to test Vκ gene specific histone marks would be to expand $CD43^{low}$ small pre-B cell clones, which are likely to be more homogenous in Vκ expression, to then see if now the histone marks become more evident on the active V genes.

It is likely that histone modifications of AgR and other monoallelic gene clusters such as OR and protocadherin genes follow similar rules. Interestingly, activating histone mark H3K4me3 is only present on few chosen promoters of protocadherin genes in the cluster, with this mark lacking in majority of other protocadherin genes in the same cluster (Guo et al. 2012). It is possible that this histone mark seen on few protocadherin genes is due to relatively small size of protocadherin cluster compared to AgR, and therefore is reflective of those few genes that are frequently and homogenously active between cells. This is unlike AgR locus which has >100V genes and therefore could be marked more heterogeneously. Apart from single cell ChIP for H3K4me3 proposed above, another way to test if the observed lack of or low histone modification in AgR and protocadherin gene cluster is due to large gene number or locus size, is to truncate these loci in *Ccnd3*$^{-/-}$ pro-B cells *in vitro*, which is expected to decreases the number of available Vs, and make histone modifications more apparent.

Igκ undergoes asynchronous replication established early during development (Mostoslavsky et al. 2001). We observed that in *Ccnd3*$^{-/-}$ pro-B cells and small pre-B cells the early replicating allele was significantly associated with transcribing allele located in tighter e-Pol II niche, which in turn interacted with Jκ and was poised for recombination. This observation is consistent with earlier reports from Bergmann group, showing that early replicating *Igκ* allele

is fated for recombination (Farago et al. 2012). This data suggest that there is a linear relationship between early replicating allele, allele in tighter e-Pol II niche and recombining allele.

If this linear relationship is true, similar to asynchronous replication, asymmetric e-Pol II niche surrounding Vκ alleles may also be established early during development, and maintained until Vκ expression and recombination in small pre-B. In order for early replicating allele to be fated for transcription and recombination, the allele and associated niche should be clonally maintained throughout development. However, early replicating allele does not show clonal maintenance in initial stages of development as either allele is found to be early replicating in the progeny (Farago et al. 2012), and only shows clonal maintenance past pro-B stage. This suggests that asymmetric niche may only be relevant past pro-B stage, where both asynchronous replication and asymmetric niche are clonally maintained, which is also consistent with clonality of recombined cells. One way to test the dynamics and clonality of asymmetric niche establishment in relation to replication and transcription would be to perform Vκ and Jκ DNA FISH and e-Pol II staining in cells of earlier developmental stages and measure e-Pol II niche dimensions around Vκ alleles.

We saw many instances where Vκ replication was directly coupled to transcription, suggesting that Vκ can be transcriptionally accessible during replication in S phase. This is possible since downregulation of cyclin D3, necessary for Vκ expression, is already in place in S and later phases, suggesting that Vκ transcription may not require complete

exit from cell cycle. Unlike replication and transcription, no coupling occurs between replication and recombination, as evident by restriction of RAG mediated breaks in G_0-G_1 phase and proteosomal degradation of RAG2 to G_1-S phase. Since transcription does not require coupling with replication for sustained expression, it suggests that something about this coupling marks the chosen allele to continue expression as the cells enter G_0 phase and encounter RAGs. Together these data suggests that replication, transcription and recombination are tightly coordinated between different stages of cell cycle.

Monoallelic genes are interspersed among biallelically transcribed genes that replicate synchronously, suggesting that replication asynchrony is coordinated throughout chromosome. For example, asynchronous replication of Vκ extends throughout *Igκ* locus (Mostoslavsky et al. 2001) and likely extends throughout entire chromosome 6 (Singh et al. 2003, Ensminger and Chess 2004). Similarly, X-chromosome also undergoes asynchronous replication, which is established chromosome-wide early in development (Simon et al. 1999) and is inherited in a stable manner by daughter cells. Interestingly, establishment of chromosome-wide asynchronous replication of monoallelic genes is not necessarily associated with their expression in that lineage. This finding allows us to test our own set of monoallelic genes from *Ccnd3*$^{-/-}$ pro-B cells for chromosome-wide asynchronous replication. One way to do so would be to use gene-specific DNA FISH on *Ccnd3*$^{-/-}$ pro-B cells for monoallelic genes in a specific chromosome. Alternatively, freshly synthesized DNA (early replicating) in *Ccnd3*$^{-/-}$ pro-B cells can be pulse-labeled by BrDU and cells FACs sorted by DNA content. BrdU DNA can then be isolated from each fraction and assayed using primers specific for target monoallelic genes in a particular chromosome, as published before(Farago et al. 2012).

In our single cell RNA-seq data, we saw transcription within CTCF defined loops, as reported before (Oti et al. 2016). Interestingly, transcription of Vκ genes occurred as read-through transcription spanning multiple Vs within these loops. One way to test if read –through transcription is a result of confinement defined by CTCF sites at *Igκ* is to alter the CTCF binding element flanking selected loops using CRISPR/Cas9 mediated editing, and test if the pattern of Vκ expression changes.

Interestingly, we also saw read-through transcription between two adjacent CTCF defined loops in our single cell RNA seq data. One interpretation of this finding is that loops can be formed by alternative CTCF pairing (Fig S2.8.5)(Merkenschlager and Nora 2016), and that regions of read-through transcription marks CTCF sites that have paired in a single cell. For example, read-through transcription between L2 and L3 in a cell suggests that the particular cell has incorporated 5' CTCF from L2 and 3' CTCF from L3, and is being expressed as a single loop (Fig S2.8.5). One way to test if L2 and L3 can be induced to make a single loop, would be to mutate 3' CTCF element from L2 and 5' CTCF element of L3 using CRISPR/guide RNA method, and see if transcription occurs through the newly made loop. If this is indeed a result of alternative pairing of CTCF sites on *Igκ* it means that alternative pairing of CTCF will not only have consequences on which loops are formed but also on which *Igκ* allele is looped appropriately to favor transcription and recombination.

Our model of Vκ transcription assumes static CTCF defined loops. However, this is not necessarily the case. Recently, "loop extrusion model" has been proposed to explain genome-

wide contact domains observed by Hi-C (Sanborn et al. 2015). The model proposes that DNA is pre-loaded with "extrusion complexes" made of doughnut shaped proteins that extrude DNA loops until convergent CTCF motifs are encountered. Based on the extrusion model, loop formation is dynamic and dependent upon the location of extrusion and CTCF sites encountered during the process, as opposed to static model, which assumes CTCF sites randomly interacting in space (Fudenberg et al. 2016). Random interaction of CTCF sites poses the risk of making intra and inter-chromosomal knots, whereas loop extrusion would prevent this outcome. In the context of *Igκ*, it is possible that random interaction of V segments to fixed RNA Pol II centers nucleates extrusion from the particular region until extrusion finds flanking CTCF sites to make a loop. Potential for alternative pairing of CTCF sites and read-through transcription between loops hints for a possibility of dynamic looping of *Igκ*.

We noticed co-relation between CTCF peak strength and regions of read-through Vκ transcription, with lower strength CTCF sites particularly clustering in the distal region, where most read-through transcription was found. For example, L1, L2 and L3, loops that showed several instances of read-through transcription were flanked by lower strength CTCF peaks compared to high strength CTCF peaks located at 5', intermediate and 3' end of *Igκ*, where read-through transcription was not common. As an example, no read-through transcription was noted between L3 and L4, which are separated by high CTCF peaks, suggesting that the region between L3 and L4 is a true boundary that prevents flanking regions from communicating with each other. This is consistent with loop extrusion model, where extrusion complexes extrude progressively larger loops, but stall at domain boundaries

(Fudenberg et al. 2016). This suggests that even by extrusion model, extrusion complexes cannot extrude between L3 and L4, which is marked by a high CTCF binding site. To precisely say that region between L3 and L4 is a boundary, which prevents read-through transcription, the intervening CTCF sites can be mutated, and tested to see if now the barrier is breached and transcription now spans between L3 and L4. An alternative way to test this possibility would be to shift the particular CTCF site 3' of its original location, and test if moving this insulator allows transcription of those additional Vs from L4 together with L3. Performing single cell Hi-C (Nagano et al. 2015) on a divergent F1 cross, would allow us to confirm this finding with much precision compared to bulk Hi-C methods. Furthermore, by mutating subset of low strength CTCF peaks the general impact on Vκ transcription can be assessed.

We noticed that in multiple cells, read-through transcription was always initiated from the same Vκ region, which suggests that there are specific "hot spots" for germline Vκ transcription. We also noticed that these sites were predominantly close to CTCF sites, which we postulate to anchor e-Pol II(Chernukhin et al. 2007). It is likely that transcription factors binding to these sites in particular have most effect on transcription. One consequence of read-through transcription is that transcription factors may only be necessary at the initiating Vκ segments. To confirm that these specific hot-spots are important for read-through transcription and therefore accessibility of Vκ, CTCF sites proximate to these hot-spots can be mutated or deleted in an *in vitro* assay and efficiency of read through transcription can then be tested thereafter.

These hotspots for transcription may also explain why there are more transcriptionally active Vs than Vs used in recombination(Aoki-Ota et al. 2012). In other words, if transcription initiates repeatedly from one Vκ, other Vs in the particular loop would be captured as read-through transcript even though they themselves may not have all the right components to be involved in recombination. Only those Vs that are expressed as part of read-through transcription and can initiate transcription of rearranged Vs can be used for recombination.

In addition to cohesin, CTCF has been reported to partner with multiple other proteins (Zlatanova and Caiafa 2009). Contrary to popular belief that CTCF is primarily found at TAD borders, and is involved in barrier function, studies in mammals have shown that only 15% of genomic CTCF-binding sites is attributable to this function (Dixon et al. 2012). This percentage is comparable to percentage of CTCF sites that are located near promoters (15%) (Ong and Corces 2014), which is suggestive of a role opposing their barrier function. CTCF peaks align with peaks for RNA polymerase II, cohesin, PARP1 and YY1, further highlighting its diverse functionality (Zlatanova and Caiafa 2009). Based on distribution of CTCF peaks along *Igκ* and transcriptional activity, we think binding partners of low strength CTCF peaks are involved in gene activation, whereas those binding high strength CTCF peaks that mark boundaries or TADs are involved in insulation. These hypotheses can be addressed by categorizing CTCF sites by their peak strength and then by binding site for their partners such as RNA Pol II, YY1, and PARP1.

CTCF sites also have a role in *Igκ* contraction and looping(Xiang, Park, and Garrard 2013). In particular, the CTCF site between L3 and L4 appears to be a pivot point for looping Vκ onto Jκ, as shown by its preferential pairing with CTCF sites in the 3' end of the locus. Removing this intermediate CTCF site, could have both implication in distal loop formation and juxtaposition with Jκ that can be tested with 2-color DNA FISH for distal Vκ and Jκ.

From polymer chain simulations we saw that constraining *Igκ* in a matrix defined cylindrical space compressed and ordered Vκ containing loops such that Vκ genes close to CTCF sites, were both exposed towards the surface of the cylinder and were accessible to nm e-Pol II. In contrast, same confinement folded intermediate loop L4 towards the interior of the cylinder and made it relatively unavailable for transcription. This gene topology predicts that nm e-Pol II tends to engage Vκ genes near CTCF sites and then stochastically read in either direction. As CTCF binding can limit processivity of e-Pol II elongation, only reading away from CTCF sites would be productive. This is evident by the pattern of Vκ expression observed in single small pre-B cells. One way to test CTCF imposed directionality of transcription would be flip transcription orientation of contiguous Vκ genes, such that they face CTCF sites and assess if transcription is now inhibited.

In our polymer chain simulation, we only used CTCF sites with high probability of contact. The findings from the simulation are consistent with DNA FISH data, showing that folding of 5' end brings distal Vκ loops closer to Jκ unlike intermediate loop, which is less structured and faces away from Jκ. The Vκ loops when constrained within a cylinder were more

organized compared to when unconstrained. However, the model did not show drastic difference in how highly transcribed and poorly transcribed Vκs were laid out with respect to e-Pol II cylinder. It is likely that we are missing additional parameters. One caveat to our simulation is that we assumed CTCF interactions as static, which may not be true. From our single cell RNA-seq data, we observed that relatively lower strength CTCF sites in the distal region appeared to participate in alternative CTCF pairing. Therefore, by running multiple independent simulations where each likely CTCF pairs are weighed differently in each simulation, may allow us to capture variability in loop formation. Doing so would likely place highly transcribed and poorly transcribed loops at different distances and orientations within the cylinder, and may explain their usage.

For our simulation, we fixed 3' end of *Igκ*, which is expected to anchor the matrix by MAR located near Eκi at the 3' end of the locus. *In vivo* consequence of MAR deletion is that *Igκ* undergoes recombination prematurely in *WT* pro-B cells(Yi et al. 1999), which suggests that MAR has a role in stage-specific expression and rearrangement of *Igκ*. Simulation in absence of fixed 3' end may allow us to see how this affects loop formation and orientation of distal, intermediate and proximal loops with respect to each other.

Most literature surrounding *Igκ* regulation focuses on Jκ, which is marked by various activating histone modifications, germline transcription and RAG binding prior to recombination(Xu and Feeney 2009, Ji et al. 2010). Therefore, all the allelic exclusion models pertaining to *Igκ* have always used Jκ as reference (Vettermann and Schlissel 2010). Studies

have shown monoallelic demethylation and monoallelic deposition of activating histone marks as well as RAG deposition over Jκ to argue monoallelic recombination. This is in contrast to germline transcription of Jκ, which has been noted in both Jκ alleles. The question that remains is why biallelic transcription does not translate to biallelic deposition of histone marks and RAG peaks over Jκ. It is likely that the studies that observed monoallelic demethylation, histone modification and RAG binding (Farago et al. 2012) were analyzing late pre-B stage, where allelic choice has been already made and recombination is already ongoing whereas those that observed biallelic Jκ transcription(Amin et al. 2009) were looking at an earlier stage, prior to allelic choice. This kind of variability is likely because even within 71 $CD19^{+}B220^{+}IgM^{-}CD43^{low}$ small pre-B cells analyzed, presumed to be pure $CD43^{low}$ population, we observed contamination of $CD43^{+}$ and $CD43^{-}$ -like cells, which were drastically different in terms of Vκ and Jκ expression. For example, in our single cell RNA-seq data, we observed biallelic Jκ transcription in $CD43^{low}$ unarearranged cells but monoallelic Jκ germline transcription in $CD43^{-}$ small pre-B cells that had undergone rearrangement. The inconsistency between biallelic germline Jκ transcription and monoallelic RAG deposition is based on our observation that we never observed $CD43^{low}$ cells that both underwent biallelic transcription and monoallelic recombination. Furthermore, the ones that had undergone recombination in one allele were also monoallelic for germline Jκ expression. Therefore, our data suggest that biallelic germline transcription of Jκ is an initial step before monoallelic choice is made, whereas monoallelic RAG binding occurs after the choice has been already made.

To come up with a unifying theory on *Igκ* allelic exclusion, it is important to look at germline transcription and demethylation/RAG deposition in the same or exactly matched population. One way to test this theory would be to make multiple clones from $CD43^{low}$ and $CD43^{-}$ small pre-B cells, and test these clones for biallelic *Igκ* transcription by RNA seq and RAG deposition/demethylation by ChIP-seq. One advantage of working with clones is abundance of DNA and RNA available for experiments and homogeneity of accessible Vs. Our hypothesis is that in $CD43^{low}$ small pre-B cells, we would see biallelic Jκ germline transcription with either no or biallelic deposition of RAG whereas in $CD43^{-}$ small pre-B cells, we would see monoallelic Jκ germline transcription with monoallelic deposition of RAGs on the transcribing allele. An alternative scenario for why RAG deposition, demethylation happens in one allele and heterochromatin recruitment occurs on another allele, despite biallelic germline transcription could be that when initial monoallelic choice is being made, heterochromatic recruitment of excluded allele does not necessarily shut off germline transcription. This is consistent with the observation that genes can be transcribed in heterochromatin (Saksouk, Simboeck, and Dejardin 2015).

There is apparent temporal regulation of Vκ and Jκ accessibility as well as usage of proximal versus distal Vs for recombination. For example, we noticed that in few cells, Jκ was accessible prior to Vκ. This finding is consistent with bimodular regulation of Vκ and Jκ, as seen by differences in extent of histone marks and RAG binding observed in the two regions. This outcome is also possible owing to the difference in size of Vκ and Jκ segments. The Jκ cluster extends as 10kb, from distal Jκ promoter to Cκ, whereas Vκ extends 3 Mb in the 5' end of the

locus, and therefore, unlike Vκ, activation of Jκ does not need large-scale changes. Increasing the number of $CD43^{low}$ single cell RNA seq sample size will allow us to test if there is a defined stage where Jκ transcription precedes Vκ transcription.

Similarly, in our single cell RNA seq data, we found exclusive use of proximal Vκ in recombination. It is possible that proximal and distal Vκ recombination is also temporally regulated, with initial rearrangement events favoring proximal usage. This can happen when proximal Vκs and Jκ are accessible and are in close proximity to each other but distal Vκs are accessible but not yet proximate to Jκ. This idea is applicable to *IgH* where DJ rearrangement in the proximal end can occur in the absence of distal looping as seen with $PAX5^{-/-}$, $YY1^{-/-}$ mice (Fuxa et al. 2004, Liu et al. 2007). Segregating proximal and distal rearrangement of *Igκ* temporally maybe important to increase the efficiency of proximal usage, as distal rearrangement in general appears to have high efficiency (Aoki-Ota et al. 2012). Although we saw recombination products exclusively in the proximal loops, we only had few cells to analyze. Furthermore, we do not know if this proximal usage is favored because distal Vs have not yet encountered Jκ in these cells or whether this skewing in proximal usage happens even when both distal and proximal Vκs are accessible and proximate to Jκ. One way to test these two scenarios is to take clones of $CD43^{low}$ and $CD43^{-}$ cells and separate them into two fractions to perform rearrangement analysis in one and Vκ-Jκ contraction in another. Alternatively, these two populations can be single cell sorted to confirm their distal or proximal usage by RNA-seq. Based on our single cell RNA seq data, it likely that we will find initial proximal usage in $CD43^{low}$ and distal usage in $CD43^{-}$ small pre-B clones, which is expected to show minimal

contraction of distal Vs in CD43^{low} small pre-B clones and increased distal contraction to Jκ in CD43^{-} small pre-B clones. If true, this result would verify that proximal and distal usage is temporally regulated.

Igκ recombination appears to be an inefficient process. Only few single cells captured for single cell RNA seq had recombined and those that had, were all out-of frame. Upon close analysis, this out-of frame rearrangement was due to random cleavage by RAG at nearby CA sites that were not part of the recombination signal sequence (CACAGTG). Furthermore, majority of cells which appeared to have undergone recombination (due to presence of single Jκ promoter expression), did not have a recombination product, suggesting that perhaps majority of the recombined out-of-frame transcripts are degraded by non-sense mediated decay (NMD). One way to test inefficient RAG cleavage and degradation of unproductively rearranged transcript is to analyze transcripts from CD43^{-} small pre-B cells deficient in NMD associated proteins such as upf1, upf2, or upf3. This analysis will also allow us to confirm inefficiency of RAG cleavage and to test if this is specific to proximal Vκs or is applicable to the entire locus. Because distal Vs are more favored for recombination, we may see differences on efficiency of RAG cleavage within *Igκ*.

During our analysis of Vκ expression from small pre-B single cells, there were multiple instances where we saw high expression of RNA that did not match annotated V region. Furthermore, these transcripts were noted in small pre-B, but not in pro-B RNA-seq samples. One possibility is that similar to Jκ, regions of Vκ also have enhancer elements and that the transcripts so observed are enhancer associated RNA, also called enhancer RNA (eRNA)(Li,

Notani, and Rosenfeld 2016). The importance of these presumed enhancer elements for Vκ transcription can be assessed by mutating or deleting this region at the DNA level, and testing for eRNA transcription as well as activation of nearby Vs. It is likely that these enhancer-like region interact with their target promoters (Kim et al. 2010) within or outside Vκs. Small pre-B enhancer landscape obtained by H3K4me1/2 marks would help solidify existence of these enhancer regions, and small pre-B Hi-C may help identify potential target promoters that these supposed Vκ eRNAs interact with.

Throughout the study, there were few instances where additional evidence was necessary to solidify the findings. For example, we saw specific colocalization of Vκ with e-Pol II factories, assembled on the nuclear matrix, however, apart from co-staining with SATB1, we did not demonstrate that these factories were in fact immobile. One way to confirm that *Igκ* is transcribed in e-Pol II transcription factories assembled on the nuclear matrix is to disintegrate *WT* small pre-B nuclear matrix with caspases (Kivinen, Kallajoki, and Taimen 2005) *in vitro*, and test by DNA FISH analysis if e-Pol II factories now dissociate from the nuclear matrix and no longer associate with *Igκ* in small pre-B cells. Similarly, we used *Tcrβ* as negative control to show that colocalization of *Igκ* with e-Pol II was specific to *Igκ* and B cells. However, to specifically say that colocalization of *Igκ* to e-Poll II only happens in small pre-B cells, *Igκ* localization should be tested in stages of T cell development or in other non-lymphoid tissues. Lastly, we investigated complete cyclin D3 knockout mice, and therefore whether the phenotype and transcriptional profile of *Ccnd3*$^{-/-}$ pro-B cells is specific to B-cells or a cumulative effect of absence of cyclin D3 throughout development is not known. In particular, expression of neuronal

genes in *Ccnd3*$^{-/-}$ pro-B cells could be a result of this collective defect. Therefore, a B-cell specific or HSC-specific deletion of cyclin D3 may provide a more precise understanding of transcription-related role of cyclin D3 in B cells and hematopoietic cells in general.

There are also some technical caveats in our study. We used Leica's GSD to visualize transcription factories and nm-cyclin D3 at super-resolution. GSD is based on capture of photons released from fluorophores that stochastically return from non-fluorescent "dark state" to ground state (also called "blinking", see Fig 1.17.1), and reconstruction of the original image. However, not all fluorophores reaching the dark state will return to ground state with equal efficiency, and therefore reconstruction of super-resolution images maybe incomplete. This suggests that the number of e-Pol II, cyclin D3 and SATB1 foci observed in our GSD images is an underestimate. Every fluorophore differs in its extent of stability in dark state, and efficiency of return (Dempsey et al. 2011). One way to make sure that the reconstruction of e-Pol II, D3 and SATB1 foci in pro-B cells is near complete is to test multiple fluorophores, for their efficiency to return to ground state (less stable in "dark state"). Furthermore, the concentration of primary antibody can be increased so that the initial labeling is high to help maximize the return.

Similarly, in some single cells, B6 and CAST alleles had to be assigned on low Vκ reads. In this case, a V segment may appear monoallelic simply due to low available reads to analyze. We did, however, detect at least one cell for each Vκ segment that was medium expressed (threshold of 100) and was assignable as monoallelic. To address if low read counts led to erroneous allelic assignment, multiple small pre-B clones can be analyzed for monoallelic expression of specific Vs. Since monoallelic genes are clonally maintained (Krueger and

Morison 2008), this approach will provide enough reads that will in turn allow us to assign alleles with more precision.

Using DNA FISH assay, we saw that 35% of Vκ genes in *Ccnd3*$^{-/-}$ pro-B cells and *WT* small pre-B cells were monoallelically colocalized with e-Pol II factories. However, combination with RNA FISH showed transcription in absence of Vκ colocalization with e-Pol II in a fraction of cells. This result goes along earlier finding that transcription occurs in burst (Finan and Cook 2011), which suggests that there is a possibility that monoallelic Vκ expression observed by confocal imaging could be a result of asynchronized transcription between the two alleles. Therefore, tracking few Vκ gene segments dynamically using TetO-R-EYFP and CDK9-RFP system would confirm that monoallelic capture of Vκ transcripts is not a result of asynchronized transcriptional burst at Vκ, but rather due to true monoallelic expression.

From our single cell RNA-seq analysis, we found that CD43^{low} cells did not comprise a pure population, as we collected both pro-B (CD43^{+}) and small pre-B (CD43^{-}) cells by using this gating strategy. Furthermore, we captured only 35 cells out of 144 cells that were expressing Vκ and Jκ transcripts in germline configuration. One way to extract a more homogenous population of cells is to perform genome wide transcription analysis of each of those 35 single cells, and identify potential surface markers that can more accurately identify this category of cells.

Similarly, *Igκ* polymer simulations were performed based on usual features of gene structure and those *Igκ* TADs predicted to be most common. It is possible that further knowledge

of specific gene structure, and consideration of alternative loops likely to form in some cells, would provide a more precise and predictive model.

Conclusion of this study is that monoallelic choice is determined by asymmetric localization of single allele in e-Pol II rich niche. In the context of *Igκ*, chosen Vκ allele is surrounded by e-Pol II arrayed in a tight e-Pol II matrix where Vκ loops are able to stochastically engage one or more nm e-Pol II complexes after cell cycle exit. Often contiguous Vκ genes in the same orientation are transcribed from multiple loops, with proximal loops with higher preference for recombination in early stages. We propose that this mechanism of loop capture transcription defines a pre-selection repertoire of accessible Vκs from which one V segment is productively captured by the recombination machinery, and forms the basis of monogenic choice from a multi-genic locus. It is likely that this loop capture mechanism is sequential with first capture mediated by nm e-Pol II and second capture mediated by RAG1/2 (Hu et al. 2015).

Overall, our data reveal that in addition to gene intrinsic mechanisms, the nuclear niche in which a gene resides can play critical roles in determining allelic choice and in restricting expression of single gene from diverse gene families. The shape and dimensionality of the niche can dictate which allele is chosen and which genes are stochastically selected for transcription. Furthermore, productive transcription within these nuclear niches is intimately tied to cell cycle exit, and thus provides a mechanism for monogenic choice in differentiated cells. Similarities with clustered gene families suggest that loop capture transcription within CTCF defined domains, and regulation by nm-cyclin D3, is a general mechanism of monogenic choice among clustered gene families.

References

Abarrategui, I., and M. S. Krangel. 2006. "Regulation of T cell receptor-alpha gene recombination by transcription." *Nat Immunol* 7 (10):1109-15.

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

Akhtar, A., and S. M. Gasser. 2007. "The nuclear envelope and transcriptional control." *Nat Rev Genet* 8 (7):507-17.

Albrethsen, J., J. C. Knol, and C. R. Jimenez. 2009. "Unravelling the nuclear matrix proteome." *J Proteomics* 72 (1):71-81.

Amin, R. H., D. Cado, H. Nolla, D. Huang, S.A. Shinton, Y. Zhou, R.R. Hardy, and M.S. Schlissel. 2009. "Biallelic, ubiquitous transcription from the distal germline Ig-kappa locus promoter during B cell development." *Proc Nat Acad Sci, USA* 106:522-527.

Amin, R. H., and M. S. Schlissel. 2008. "Foxo1 directly regulates the transcription of recombination-activating genes during B cell development." *Nat Immunol* 9 (6):613-22.

Aoki-Ota, Miyo, Ali Torkamani, Takayuki Ota, Nicholas Schork, and David Nemazee. 2012. "Skewed primary Igκ repertoire and V-J joining in C57BL/6 mice: implications for recombination accessibility and receptor editing." *J Immunol* 188 (5):2305-2315.

Atchison, M. L. 2014. "Function of YY1 in long-distance DNA interactions." *Front Immunol* 5:45.

Bain, G., E. C. Maandag, D. J. Izon, D. Amsen, A. M. Kruisbeek, B. C. Weintraub, I. Krop, M. S. Schlissel, A. J. Feeney, M. van Roon, and et al. 1994. "E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements.[comment]." *Cell.* 79 (5):885-92.

Barlow, D. P., and M. S. Bartolomei. 2014. "Genomic imprinting in mammals." *Cold Spring Harb Perspect Biol* 6 (2).

Bemark, M., A. Martensson, D. Liberg, and T. Leanderson. 1999. "Spi-C, a novel Ets protein that is temporally regulated during B lymphocyte development." *J Biol Chem* 274 (15):10259-67.

Berezney, R., and D. S. Coffey. 1974. "Identification of a nuclear protein matrix." *Biochem Biophys Res Commun* 60 (4):1410-7.

Bickmore, W. A., and B. van Steensel. 2013. "Genome architecture: domain organization of interphase chromosomes." *Cell* 152 (6):1270-84.

Birmingham, A., E. Anderson, K. Sullivan, A. Reynolds, Q. Boese, D. Leake, J. Karpilow, and A. Khvorova. 2007. "A protocol for designing siRNAs with high functionality and specificity." *Nat Protoc* 2 (9):2068-78.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. "Trimmomatic: a flexible trimmer for Illumina sequence data." *Bioinformatics* 30 (15):2114-20.

Bolland, D. J., A. L. Wood, R. Afshar, K. Featherstone, E. M. Oltz, and A. E. Corcoran. 2007. "Antisense intergenic transcription precedes Igh D-to-J recombination and is controlled by the intronic enhancer Emu." *Mol Cell Biol* 27 (15):5523-33.

Cao, J. 2014. "The functional role of long non-coding RNAs and epigenetics." *Biol Proced Online* 16:11.

Chen, W. V., and T. Maniatis. 2013. "Clustered protocadherins." *Development* 140 (16):3297-302.

Chernukhin, I., S. Shamsuddin, S. Y. Kang, R. Bergstrom, Y. W. Kwon, W. Yu, J. Whitehead, R. Mukhopadhyay, F. Docquier, D. Farrar, I. Morrison, M. Vigneron, S. Y. Wu, C. M. Chiang, D. Loukinov, V. Lobanenkov, R. Ohlsson, and E. Klenova. 2007. "CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide." *Mol Cell Biol* 27 (5):1631-48.

Chess, A. 2005. "Monoallelic expression of protocadherin genes." *Nat Genet* 37 (2):120-1.

Chess, A., I. Simon, H. Cedar, and R. Axel. 1994. "Allelic inactivation regulates olfactory receptor gene expression." *Cell* 78 (5):823-834.

Chess, Andrew. 2013. "Random and Non-Random Monoallelic Expression." *Neuropsychopharmacology* 38 (1):55-61.

Choi, N. M., S. Loguercio, J. Verma-Gaur, S. C. Degner, A. Torkamani, A. I. Su, E. M. Oltz, M. Artyomov, and A. J. Feeney. 2013. "Deep sequencing of the murine IgH repertoire reveals complex regulation of nonrandom V gene rearrangement frequencies." *J Immunol* 191 (5):2393-402.

Ciemerych, M. A., A. M. Kenney, E. Sicinska, I. Kalaszczynska, R. T. Bronson, D. H. Rowitch, H. Gardner, and P. Sicinski. 2002. "Development of mice expressing a single D-type cyclin." *Genes Dev* 16 (24):3277-89.

Clark, M. R., M. Mandal, K. Ochiai, and H. Singh. 2014. "Orchestrating B cell lymphopoiesis through interplay of IL-7 receptor and pre-B cell receptor signalling." *Nat Rev Immunol* 14 (2):69-80.

Cocea, L., A. De Smet, M. Saghatchian, S. Fillatreau, L. Ferradini, S. Schurmans, J. C. Weill, and C. A. Reynaud. 1999. "A targeted deletion of a region upstream from the Jkappa cluster impairs kappa chain rearrangement in cis in mice and in the 103/bcl2 cell line." *J Exp Med* 189 (9):1443-50.

Coleman, M. L., C. J. Marshall, and M. F. Olson. 2004. "RAS and RHO GTPases in G1-phase cell-cycle regulation." *Nat Rev Mol Cell Biol* 5 (5):355-66.

Cooper, A. Byron, Catherine M. Sawai, Ewa Sicinska, Sarah E. Powers, Piotr Sicinski, Marcus R. Clark, and Iannis Aifantis. 2006. "A unique function for cyclin D3 in early B cell development." *Nature immunology* 7 (5):489-497.

Corcoran, Anne E. 2010. "The epigenetic role of non-coding RNA transcription and nuclear organization in immunoglobulin repertoire generation." *Seminars in immunology* 22 (6):353-361.

Davalos-Salas, M., M. Furlan-Magaril, E. Gonzalez-Buendia, C. Valdes-Quezada, E. Ayala-Ortega, and F. Recillas-Targa. 2011. "Gain of DNA methylation is enhanced in the absence of CTCF at the human retinoblastoma gene promoter." *BMC Cancer* 11:232.

de Almeida, Claudia Ribeiro, Rudi W. Hendriks, and Ralph Stadhouders. 2015. "Chapter five - Dynamic control of long-range genomic interactions at the immunoglobulin κ light-chain locus." In *Adv in Immunol*, edited by Murre Cornelis, 183-271. Academic Press.

Degner, S. C., J. Verma-Gaur, T. P. Wong, C. Bossen, G. M. Iverson, A. Torkamani, C. Vettermann, Y. C. Lin, Z. Ju, D. Schulz, C. S. Murre, B. K. Birshtein, N. J. Schork, M. S. Schlissel, R. Riblet, C. Murre, and A. J. Feeney. 2011. "CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells." *Proc Natl Acad Sci U S A* 108 (23):9566-71.

Dempsey, G. T., J. C. Vaughan, K. H. Chen, M. Bates, and X. Zhuang. 2011. "Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging." *Nat Methods* 8 (12):1027-36.

Deng, Qiaolin, Daniel Ramsköld, Björn Reinius, and Rickard Sandberg. 2014. "Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells." *Science (New York, N.Y.)* 343 (6167):193-196.

Dengler, H.S., G.V. Baracho, S. A. Omori, S. Bruckner, K. C. Arden, D. H. Castrillon, R. A. DePinho, and R. C. Rickert. 2008. "Distinct functions for the transcription factor Foxo1 at various stages of B cell differentiation." *Nat Immunol* 12:1388-1398.

Difilippantonio, M. J., J. Zhu, H. T. Chen, E. Meffre, M. C. Nussenzweig, E. E. Max, T. Ried, and A. Nussenzweig. 2000. "DNA repair protein Ku80 suppresses chromosomal aberrations and malignant transformation." *Nature* 404 (6777):510-4.

Dixon, J. R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren. 2012. "Topological domains in mammalian genomes identified by analysis of chromatin interactions." *Nature* 485 (7398):376-80.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. 2013. "STAR: ultrafast universal RNA-seq aligner." *Bioinformatics* 29 (1):15-21.

Duber, S., H. Engel, A. Rolink, K. Kretschmer, and S. Weiss. 2003. "Germline transcripts of immunoglobulin light chain variable regions are structurally diverse and differentially expressed." *Mol Immunol* 40 (8):509-16.

Durrin, L. K., and T. G. Krontiris. 2002. "The thymocyte-specific MAR binding protein, SATB1, interacts in vitro with a novel variant of DNA-directed RNA polymerase II, subunit 11." *Genomics* 79 (6):809-17.

Ebert, A., S. McManus, H. Tagoh, J. Medvedovic, G. Salvagiotto, M. Novatchkova, I. Tamir, A. Sommer, M. Jaritz, and M. Busslinger. 2011. "The distal V(H) gene cluster of the Igh locus contains distinct regulatory elements with Pax5 transcription factor-dependent activity in pro-B cells." *Immunity* 34 (2):175-87.

Ebert, Anja, Louisa Hill, and Meinrad Busslinger. 2015. "Chapter three - Spatial regulation of V–(D)J recombination at antigen receptor loci." In *Advances in Immunology*, edited by Murre Cornelis, 93-121. Academic Press.

Eckersley-Maslin, Mélanie A., David Thybert, Jan H. Bergmann, John C. Marioni, Paul Flicek, and David L. Spector. 2014. "Random Monoallelic Gene Expression Increases upon Embryonic Stem Cell Differentiation." *Developmental Cell* 28 (4):351-365.

Ensminger, A. W., and A. Chess. 2004. "Coordinated replication timing of monoallelically expressed genes along human autosomes." *Hum Mol Genet* 13 (6):651-8.

Farago, M., C. Rosenbluh, M. Tevlin, S. Fraenkel, S. Schlesinger, H. Masika, M. Gouzman, G. Teng, D. Schatz, Y. Rais, J. H. Hanna, A. Mildner, S. Jung, G. Mostoslavsky, H. Cedar, and Y. Bergman. 2012. "Clonal allelic predetermination of immunoglobulin-kappa rearrangement." *Nature* 490 (7421):561-5.

Ferradini, L., H. Gu, A. De Smet, K. Rajewsky, C. A. Reynaud, and J. C. Weill. 1996. "Rearrangement-enhancing element upstream of the mouse immunoglobulin kappa chain J cluster." *Science* 271 (5254):1416-20.

Ferrai, Carmelo, Sheila Q. Xie, Paolo Luraghi, Davide Munari, Francisco Ramirez, Miguel R. Branco, Ana Pombo, and Massimo P. Crippa. 2010. "Poised transcription factories prime silent uPA gene prior to activation." *PLoS biology* 8 (1):e1000270.

Ferreiros-Vidal, I., T. Carroll, B. Taylor, A. Terry, Z. Liang, L. Bruno, G. Dharmalingam, S. Khadayate, B. S. Cobb, S. T. Smale, M. Spivakov, P. Srivastava, E. Petretto, A. G. Fisher, and M. Merkenschlager. 2013. "Genome-wide identification of Ikaros targets elucidates its contribution to mouse B-cell lineage specification and pre-B-cell differentiation." *Blood* 121 (10):1769-82.

Fey, E. G., P. Bangs, C. Sparks, and P. Odgren. 1991. "The nuclear matrix: defining structural and functional roles." *Crit Rev Eukaryot Gene Expr* 1 (2):127-43.

Finan, Kieran, and Peter R. Cook. 2011. "Transcriptional Initiation: Frequency, Bursting, and Transcription Factories." In *Genome Organization and Function in the Cell Nucleus*, 235-254. Wiley-VCH Verlag GmbH & Co. KGaA.

Fitzsimmons, Sean P., Ralph M. Bernstein, Edward E. Max, Jane A. Skok, and Marjorie A. Shapiro. 2007. "Dynamic Changes in Accessibility, Nuclear Positioning, Recombination, and Transcription at the Igκ Locus." *The Journal of Immunology* 179 (8):5264-5273.

Flemming, A., T. Brummer, M. Reth, and H. Jumaa. 2003. "The adaptor protein SLP-65 acts as a tumor suppressor that limits pre-B cell expansion." *Nat Immunol* 4 (1):38-43.

Fudenberg, Geoffrey, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A. Mirny. 2016. "Formation of chromosomal domains by loop extrusion." *Cell Rep* 15 (9):2038-2049.

Fudenberg, Geoffrey, and Leonid A. Mirny. 2012. "Higher-order chromatin structure: bridging physics and biology." *Current Opinion in Genetics & Development* 22 (2):115-124.

Fuxa, M., J. Skok, A. Souabni, G. Salvagiotto, E. Roldan, and M. Busslinger. 2004. "Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene." *Genes Dev* 18 (4):411-22.

Gao, Y., D. O. Ferguson, W. Xie, J. P. Manis, J. Sekiguchi, K. M. Frank, J. Chaudhuri, J. Horner, R. A. DePinho, and F. W. Alt. 2000. "Interplay of p53 and DNA-repair protein XRCC4 in tumorigenesis, genomic stability and development." *Nature* 404 (6780):897-900.

Gendrel, A. V., M. Attia, C. J. Chen, P. Diabangouaya, N. Servant, E. Barillot, and E. Heard. 2014. "Developmental dynamics and disease potential of random monoallelic gene expression." *Dev Cell* 28 (4):366-80.

Gerdes, T., and M. Wabl. 2004. "Autoreactivity and allelic inclusion in a B cell nuclear transfer mouse." *Nat Immunol* 5 (12):1282-7.

Getzenberg, R. H. 1994. "Nuclear matrix and the regulation of gene expression: tissue specificity." *J Cell Biochem* 55 (1):22-31.

Ghamari, Alireza, Mariëtte P. C. van de Corput, Supat Thongjuea, Wiggert A. van Cappellen, Wilfred van Ijcken, Jeffrey van Haren, Eric Soler, Dirk Eick, Boris Lenhard, and Frank G. Grosveld. 2013. "In vivo live imaging of RNA polymerase II transcription factories in primary cells." *Genes & Development* 27 (7):767-777.

Glaser, R. L., J. P. Ramsay, and I. M. Morison. 2006. "The imprinted gene and parent-of-origin effect database now includes parental origin of de novo mutations." *Nucleic Acids Res* 34 (Database issue):D29-31.

Goldmit, Maya, Yanhong Ji, Jane Skok, Esther Roldan, Steffen Jung, Howard Cedar, and Yehudit Bergman. 2005. "Epigenetic ontogeny of the Igk locus during B cell development." *Nature Immunology* 6 (2):198-203.

Guo, Y., K. Monahan, H. Wu, J. Gertz, K. E. Varley, W. Li, R. M. Myers, T. Maniatis, and Q. Wu. 2012. "CTCF/cohesin-mediated DNA looping is required for protocadherin alpha promoter choice." *Proc Natl Acad Sci U S A* 109 (51):21081-6.

Guo, Y., Q. Xu, D. Canzio, J. Shou, J. Li, D. U. Gorkin, I. Jung, H. Wu, Y. Zhai, Y. Tang, Y. Lu, Y. Wu, Z. Jia, W. Li, M. Q. Zhang, B. Ren, A. R. Krainer, T. Maniatis, and Q. Wu. 2015. "CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function." *Cell* 162 (4):900-10.

Hartley, S. B., J. Crosbie, R. Brink, A. B. Kantor, A. Basten, and C. C. Goodnow. 1991. "Elimination from peripheral lymphoid tissues of self-reactive B lymphocytes recognizing membrane-bound antigens." *Nature.* 353 (6346):765-9.

Heng, T.S., M.W. Painter, and Immunological Genome Project Consortium. 2008. "The Immunological Genome Project: networks of gene expression in immune cells." *Nat Immunol* 9:1091-1094.

Herzog, Sebastian, Eva Hug, Sonja Meixlsperger, Ji-Hye Paik, Ronald A. DePinho, Michael Reth, and Hassan Jumaa. 2008. "SLP-65 regulates immunoglobulin light chain gene recombination through the PI(3)K-PKB-Foxo pathway." *Nature Immunology* 9 (6):623-631.

Hewitt, S. L., B. Yin, Y. Ji, J. Chaumeil, K. Marszalek, J. Tenthorey, G. Salvagiotto, N. Steinel, L. B. Ramsey, J. Ghysdael, M. A. Farrar, B. P. Sleckman, D. G. Schatz, M. Busslinger, C. H. Bassing, and J. A. Skok. 2009. "RAG-1 and ATM coordinate monoallelic recombination and nuclear positioning of immunoglobulin loci." *Nat Immunol* 10 (6):655-64.

Hewitt, Susannah L., Deborah Farmer, Katarzyna Marszalek, Emily Cadera, Hong-Erh Liang, Yang Xu, Mark S. Schlissel, and Jane A. Skok. 2008. "Association between the Igk and Igh immunoglobulin loci mediated by the 3' Igk enhancer induces 'decontraction' of the Igh locus in pre-B cells." *Nature Immunology* 9 (4):396-404.

Hodawadekar, S., K. Park, M. A. Farrar, and M. L. Atchison. 2012. "A developmentally controlled competitive STAT5-PU.1 DNA binding mechanism regulates activity of the Ig kappa E3' enhancer." *J Immunol* 188 (5):2276-84.

Hu, J., Y. Zhang, L. Zhao, R. L. Frock, Z. Du, R. M. Meyers, F. L. Meng, D. G. Schatz, and F. W. Alt. 2015. "Chromosomal loop domains direct the recombination of antigen receptor genes." *Cell* 163 (4):947-59.

Iborra, F. J., A. Pombo, D. A. Jackson, and P. R. Cook. 1996. "Active RNA polymerases are localized within discrete transcription "factories' in human nuclei." *Journal of Cell Science* 109 (Pt 6):1427-1436.

Inlay, M. A., H. H. Gao, V. H. Odegard, T. Lin, D. G. Schatz, and Y. Xu. 2006. "Roles of the Ig kappa light chain intronic and 3' enhancers in Igk somatic hypermutation." *J Immunol* 177 (2):1146-51.

Inlay, M. A., H. Tian, T. Lin, and Y. Xu. 2004. "Important roles for E protein binding sites within the immunoglobulin kappa chain intronic enhancer in activating Vkappa Jkappa rearrangement." *J Exp Med* 200 (9):1205-11.

Inlay, Matthew, Frederick W. Alt, David Baltimore, and Yang Xu. 2002. "Essential roles of the kappa light chain intronic enhancer and 3' enhancer in kappa rearrangement and demethylation." *Nature Immunology* 3 (5):463-468.

Jackson, D. A., and P. R. Cook. 1985. "Transcription occurs at a nucleoskeleton." *The EMBO journal* 4 (4):919-925.

Jackson, D. A., and P. R. Cook. 1986. "Replication occurs at a nucleoskeleton." *The EMBO journal* 5 (6):1403-1410.

Jackson, D. A., A. B. Hassan, R. J. Errington, and P. R. Cook. 1993. "Visualization of focal sites of transcription within human nuclei." *The EMBO journal* 12 (3):1059-1065.

Jackson, D. A., F. J. Iborra, E. M. Manders, and P. R. Cook. 1998. "Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei." *Molecular Biology of the Cell* 9 (6):1523-1536.

Ji, Yanhong, Wolfgang Resch, Elizabeth Corbett, Arito Yamane, Rafael Casellas, and David G. Schatz. 2010. "The In Vivo Pattern of Binding of RAG1 and RAG2 to Antigen Receptor Loci." *Cell* 141 (3):419-431.

Jia, J., M. Kondo, and Y. Zhuang. 2007. "Germline transcription from T-cell receptor Vbeta gene is uncoupled from allelic exclusion." *Embo j* 26 (9):2387-99.

Jiang, H., F. C. Chang, A. E. Ross, J. Lee, K. Nakayama, and S. Desiderio. 2005. "Ubiquitylation of RAG-2 by Skp2-SCF links destruction of the V(D)J recombinase to the cell cycle." *Mol Cell* 18 (6):699-709.

Johnson, Kristen, Tamar Hashimshony, Catherine M. Sawai, Jagan M. R. Pongubala, Jane A. Skok, Iannis Aifantis, and Harinder Singh. 2008. "Regulation of Immunoglobulin Light-Chain Recombination by the Transcription Factor IRF-4 and the Attenuation of Interleukin-7 Signaling." *Immunity* 28 (3):335-345.

Jumaa, H., M. Mitterer, M. Reth, and P. J. Nielsen. 2001. "The absence of SLP65 and Btk blocks B cell development at the preB cell receptor-positive stage." *Eur J Immunol* 31 (7):2164-9.

Keane, T. M., L. Goodstadt, P. Danecek, M. A. White, K. Wong, B. Yalcin, A. Heger, A. Agam, G. Slater, M. Goodson, N. A. Furlotte, E. Eskin, C. Nellaker, H. Whitley, J. Cleak, D. Janowitz, P. Hernandez-Pliego, A. Edwards, T. G. Belgard, P. L. Oliver, R. E. McIntyre, A. Bhomra, J. Nicod, X. Gan, W. Yuan, L. van der Weyden, C. A. Steward, S. Bala, J. Stalker, R. Mott, R. Durbin, I. J. Jackson, A. Czechanski, J. A. Guerra-Assuncao, L. R. Donahue, L. G. Reinholdt, B. A. Payseur, C. P. Ponting, E. Birney, J. Flint, and D. J. Adams. 2011. "Mouse genomic variation and its effect on phenotypes and gene regulation." *Nature* 477 (7364):289-94.

Kee, B. L., M. W. Quong, and C. Murre. 2000. "E2A proteins: essential regulators at multiple stages of B-cell development." *Immunological reviews* 175:138-149.

Kim, J., S. Sif, B. Jones, A. Jackson, J. Koipally, E. Heller, S. Winandy, A. Viel, A. Sawyer, T. Ikeda, R. Kingston, and K. Georgopoulos. 1999. "Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes." *Immunity* 10 (3):345-55.

Kim, T. K., M. Hemberg, J. M. Gray, A. M. Costa, D. M. Bear, J. Wu, D. A. Harmin, M. Laptewicz, K. Barbara-Haley, S. Kuersten, E. Markenscoff-Papadimitriou, D. Kuhl, H. Bito, P. F. Worley, G. Kreiman, and M. E. Greenberg. 2010. "Widespread transcription at neuronal activity-regulated enhancers." *Nature* 465 (7295):182-7.

Kimura, H., Y. Tao, R. G. Roeder, and P. R. Cook. 1999. "Quantitation of RNA polymerase II and its transcription factors in an HeLa cell: little soluble holoenzyme but significant amounts of polymerases attached to the nuclear substructure." *Mol Cell Biol* 19 (8):5383-92.

Kitsberg, D., S. Selig, M. Brandeis, I. Simon, I. Keshet, D. J. Driscoll, R. D. Nicholls, and H. Cedar. 1993. "Allele-specific replication timing of imprinted gene regions." *Nature* 364 (6436):459-463.

Kivinen, K., M. Kallajoki, and P. Taimen. 2005. "Caspase-3 is required in the apoptotic disintegration of the nuclear matrix." *Exp Cell Res* 311 (1):62-73.

Kosak, S. T., J. A. Skok, K. L. Medina, R. Riblet, M. M. Le Beau, A. G. Fisher, and H. Singh. 2002. "Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development." *Science* 296 (5565):158-62.

Kozar, Katarzyna, Maria A. Ciemerych, Vivienne I. Rebel, Hirokazu Shigematsu, Agnieszka Zagozdzon, Ewa Sicinska, Yan Geng, Qunyan Yu, Shoumo Bhattacharya, Roderick T. Bronson, Koichi Akashi, and Piotr Sicinski. 2004. "Mouse development and cell proliferation in the absence of D-cyclins." *Cell* 118 (4):477-491.

Krueger, C., and I. M. Morison. 2008. "Random monoallelic expression: making a choice." *Trends Genet* 24 (6):257-9.

Kuo, L. J., and L. X. Yang. 2008. "Gamma-H2AX - a novel biomarker for DNA double-strand breaks." *In Vivo* 22 (3):305-9.

Langmead, B., and S. L. Salzberg. 2012. "Fast gapped-read alignment with Bowtie 2." *Nat Methods* 9 (4):357-9.

Lazorchak, Adam S., Mark S. Schlissel, and Yuan Zhuang. 2006. "E2A and IRF-4/Pip promote chromatin modification and transcription of the immunoglobulin kappa locus in pre-B cells." *Molecular and Cellular Biology* 26 (3):810-821.

Lefranc, M. P., V. Giudicelli, Q. Kaas, E. Duprat, J. Jabado-Michaloud, D. Scaviner, C. Ginestoux, O. Clement, D. Chaume, and G. Lefranc. 2005. "IMGT, the international ImMunoGeneTics information system." *Nucleic Acids Res* 33 (Database issue):D593-7.

Li, Wenbo, Dimple Notani, and Michael G. Rosenfeld. 2016. "Enhancers as non-coding RNA transcription units: recent insights and future perspectives." *Nat Rev Genet* 17 (4):207-223.

Li, Z., D. I. Dordai, J. Lee, and S. Desiderio. 1996. "A conserved degradation signal regulates RAG-2 accumulation during cell division and links V(D)J recombination to the cell cycle." *Immunity* 5 (6):575-589.

Lieber, M. R., K. Yu, and S. C. Raghavan. 2006. "Roles of nonhomologous DNA end joining, V(D)J recombination, and class switch recombination in chromosomal translocations." *DNA Repair (Amst)* 5 (9-10):1234-45.

Lin, H., and R. Grosschedl. 1995. "Failure of B-cell differentiation in mice lacking the transcription factor EBF." *Nature.* 376 (6537):263-7.

Lin, W.C., and S. Desiderio. 1993. "Regulation of V(D)J Recombination Activator Protein RAG-2 by Phosphorylation." *Science* 260 (5110):953-959.

Lin, Y.C., C. Benner, R. Mansson, S. Heinz, K. Miyazaki, M. Miyazaki, V. Chandra, C. Bossen, C.K. Glass, and C. Murre. 2012. "Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate." *Nat Immunol* 13:1196-1205.

Lin, Y.C., S. Jhunjhunwala, C. Benner, S. Heinz, E. Welinder, R. Mansson, M. Sigvardsson, J. Hagman, C.A. Espinoza, J. Duthowski, T. Ideker, C.K. Glass, and C. Murre. 2010. "A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate." *Nat Immunol* 11:635-643.

Liu, H., M. Schmidt-Supprian, Y. Shi, E. Hobeika, N. Barteneva, H. Jumaa, R. Pelanda, M. Reth, J. Skok, K. Rajewsky, and Y. Shi. 2007. "Yin Yang 1 is a critical regulator of B-cell development." *Genes Dev* 21 (10):1179-89.

Lu, Runqing, Kay L. Medina, David W. Lancki, and Harinder Singh. 2003. "IRF-4,8 orchestrate the pre-B-to-B transition in lymphocyte development." *Genes & Development* 17 (14):1703-1708.

Lucas, J. S., Y. Zhang, O. K. Dudko, and C. Murre. 2014. "3D trajectories adopted by coding and regulatory DNA elements: first-passage times for genomic interactions." *Cell* 158 (2):339-52.

Ma, S., A. Turetsky, L. Trinh, and R. Lu. 2006. "IFN regulatory factor 4 and 8 promote Ig light chain kappa locus activation in pre-B cell development." *J Immunol* 177 (11):7898-904.

Ma, Shibin, Simanta Pathak, Malay Mandal, Long Trinh, Marcus R. Clark, and Runqing Lu. 2010. "Ikaros and Aiolos inhibit pre-B-cell proliferation by directly suppressing c-Myc expression." *Molecular and Cellular Biology* 30 (17):4149-4158.

Ma, Shibin, Simanta Pathak, Long Trinh, and Runqing Lu. 2008. "Interferon regulatory factors 4 and 8 induce the expression of Ikaros and Aiolos to down-regulate pre-B-cell receptor and promote cell-cycle withdrawal in pre-B-cell development." *Blood* 111 (3):1396-1403.

Malin, S., S. McManus, C. Cobaleda, M. Novatchkova, A. Delogu, P. Bouillet, A. Strasser, and M. Busslinger. 2010. "Role of STAT5 in controlling cell survival and immunoglobulin gene recombination during pro-B cell development." *Nat Immunol* 11:171-179.

Mandal, M., 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. 2015. "Histone reader BRWD1 targets and restricts recombination to the Igk locus." *Nat Immunol* 16 (10):1094-103.

Mandal, Malay, Sarah E. Powers, Mark Maienschein-Cline, Elizabeth T. Bartom, Keith M. Hamel, Barbara L. Kee, Aaron R. Dinner, and Marcus R. Clark. 2011. "Epigenetic repression of the Igk locus by STAT5-mediated recruitment of the histone methyltransferase Ezh2." *Nature immunology* 12 (12):1212-1220.

Mandal, Malay, Sarah E. Powers, Kyoko Ochiai, Katia Georgopoulos, Barbara L. Kee, Harinder Singh, and Marcus R. Clark. 2009. "Ras orchestrates exit from the cell cycle and light-chain recombination during early B cell development." *Nature Immunology* 10 (10):1110-1117.

Martin, D. J., and B. G. van Ness. 1990. "Initiation and processing of two kappa immunoglobulin germ line transcripts in mouse B cells." *Mol Cell Biol* 10 (5):1950-8.

Martinez-Jean, C., G. Folch, and M. P. Lefranc. 2001. "Nomenclature and overview of the mouse (Mus musculus and Mus sp.) immunoglobulin kappa (IGK) genes." *Exp Clin Immunogenet* 18 (4):255-279.

Maston, Glenn A., Sara K. Evans, and Michael R. Green. 2006. "Transcriptional regulatory elements in the human genome." *Annual Review of Genomics and Human Genetics* 7:29-59.

Matthews, A. G., and M. A. Oettinger. 2009. "RAG: a recombinase diversified." *Nat Immunol* 10 (8):817-21.

Max, E. E., J. G. Seidman, and P. Leder. 1979. "Sequences of five potential recombination sites encoded close to an immunoglobulin kappa constant region gene." *Proceedings of the National Academy of Sciences of the United States of America* 76 (7):3450-3454.

Medvedovic, J., A. Ebert, H. Tagoh, I. M. Tamir, T. A. Schwickert, M. Novatchkova, Q. Sun, P. J. Huis In 't Veld, C. Guo, H. S. Yoon, Y. Denizot, S. J. Holwerda, W. de Laat, M. Cogne, Y. Shi, F. W. Alt, and M. Busslinger. 2013. "Flexible long-range loops in the VH gene region of the Igh locus facilitate the generation of a diverse antibody repertoire." *Immunity* 39 (2):229-44.

Melchers, F. 2005. "The pre-B-cell receptor: selector of fitting immunoglobulin heavy chains for the B-cell repertoire." *Nat Rev Immunol* 5 (7):578-84.

Melchers, F. 2015. "Checkpoints that control B cell development." *J Clin Invest* 125 (6):2203-10.

Melnik, Svitlana, Binwei Deng, Argyris Papantonis, Sabyasachi Baboo, Ian M. Carr, and Peter R. Cook. 2011. "The proteomes of transcription factories containing RNA polymerases I, II or III." *Nature Methods* 8 (11):963-968.

Merkenschlager, M., and E. P. Nora. 2016. "CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation." *Annu Rev Genomics Hum Genet*.

Mitchell, Jennifer A., and Peter Fraser. 2008. "Transcription factories are nuclear subcompartments that remain in the absence of transcription." *Genes & Development* 22 (1):20-25.

Monahan, K., and S. Lomvardas. 2015. "Monoallelic expression of olfactory receptors." *Annu Rev Cell Dev Biol* 31:721-40.

Monahan, K., N. D. Rudnick, P. D. Kehayova, F. Pauli, K. M. Newberry, R. M. Myers, and T. Maniatis. 2012. "Role of CCCTC binding factor (CTCF) and cohesin in the generation of single-cell diversity of protocadherin-alpha gene expression." *Proc Natl Acad Sci U S A* 109 (23):9125-30.

Morse, Philip M. 1929. "Diatomic molecules according to the wave mechanics. II. vibrational levels." *Phys Rev* 34 (1):57-64.

Mostoslavsky, R., N. Singh, T. Tenzen, M. Goldmit, C. Gabay, S. Elizur, P. Qi, B. E. Reubinoff, A. Chess, H. Cedar, and Y. Bergman. 2001. "Asynchronous replication and allelic exclusion in the immune system." *Nature* 414 (6860):221-225.

Mostoslavsky, Raul, Nandita Singh, Andrei Kirillov, Roberta Pelanda, Howard Cedar, Andrew Chess, and Yehudit Bergman. 1998. "κ chain monoallelic demethylation and the establishment of allelic exclusion." *Genes & Development* 12 (12):1801-1811.

Munkley, J., N. A. Copeland, V. Moignard, J. R. Knight, E. Greaves, S. A. Ramsbottom, M. E. Pownall, J. Southgate, J. F. Ainscough, and D. Coverley. 2011. "Cyclin E is recruited to the nuclear matrix during differentiation, but is not recruited in cancer cells." *Nucleic Acids Res* 39 (7):2671-7.

Murre, C., P. S. McCaw, H. Vaessin, M. Caudy, L. Y. Jan, Y. N. Jan, C. V. Cabrera, J. N. Buskin, S. D. Hauschka, A. B. Lassar, and et al. 1989. "Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence." *Cell* 58 (3):537-44.

Nag, A., S. Vigneau, V. Savova, L. M. Zwemer, and A. A. Gimelbrant. 2015. "Chromatin signature identifies monoallelic gene expression across mammalian cell types." *G3* 5 (8):1713-20.

Nagano, T., Y. Lubling, E. Yaffe, S. W. Wingett, W. Dean, A. Tanay, and P. Fraser. 2015. "Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell." *Nat Protoc* 10 (12):1986-2003.

Nemazee, D. A., and K. Buerki. 1989. "Clonal deletion of B lymphocytes in a transgenic mouse bearing anti-MHC class I antibody genes." *Nature* 337:562-566.

Nutt, S. L., P. Urbanek, A. Rolink, and M. Busslinger. 1997. "Essential functions of Pax5 (BSAP) in pro-B cell development: difference between fetal and adult B lymphopoiesis and reduced V-to-DJ recombination at the IgH locus." *Genes & Development.* 11 (4):476-91.

Ong, Chin-Tong, and Victor G. Corces. 2014. "CTCF: an architectural protein bridging genome topology and function." *Nature Reviews. Genetics* 15 (4):234-246.

Orlanski, S., V. Labi, Y. Reizel, A. Spiro, M. Lichtenstein, R. Levin-Klein, S. B. Koralov, Y. Skversky, K. Rajewsky, H. Cedar, and Y. Bergman. 2016. "Tissue-specific DNA demethylation is required for proper B-cell differentiation and function." *Proc Natl Acad Sci U S A* 113 (18):5018-23.

Oti, M., J. Falck, M. A. Huynen, and H. Zhou. 2016. "CTCF-mediated chromatin loops enclose inducible gene regulatory domains." *BMC Genomics* 17:252.

Papantonis, Argyris, and Peter R. Cook. 2011. "Fixing the model for transcription: the DNA moves, not the polymerase." *Transcription* 2 (1):41-44.

Passegue, E., A. J. Wagers, S. Giuriato, W. C. Anderson, and I. L. Weissman. 2005. "Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates." *J Exp Med* 202 (11):1599-611.

Pena-Hernandez, R., M. Marques, K. Hilmi, T. Zhao, A. Saad, M. A. Alaoui-Jamali, S. V. del Rincon, T. Ashworth, A. L. Roy, B. M. Emerson, and M. Witcher. 2015. "Genome-wide targeting of the epigenetic regulatory protein CTCF to gene promoters by the transcription factor TFII-I." *Proc Natl Acad Sci U S A* 112 (7):E677-86.

Peschon, J. J., P. J. Morrissey, K. H. Grabstein, F. J. Ramsdell, E. Maraskovsky, B. C. Gliniak, L. S. Park, S. F. Ziegler, D. E. Williams, C. B. Ware, and et al. 1994. "Early lymphocyte expansion is severely impaired in interleukin 7 receptor-deficient mice." *J Exp Med* 180 (5):1955-60.

Plimpton, Steve. 1995. "Fast parallel algorithms for short-range molecular dynamics." *J Comput Phys* 117 (1):1-19.

Pongubala, J. M., D.L. Northrup, D. W. Lancki, K.L. Medina, T. Treiber, E. Bertolino, M. K. Thomas, R. Grosschedl, D. Allman, and H. Singh. 2008. "Transcription factor EBF restricts alternative lineage options and promotes B cell fate commitment independently of Pax5." *Nat Immunol* 9:203-215.

Powers, S. E., M. Mandal, S. Matsuda, A. V. Miletic, M. H. Cato, A. Tanaka, R. C. Rickert, S. Koyasu, and M. R. Clark. 2012. "Subnuclear cyclin D3 compartments and the coordinated regulation of proliferation and immunoglobulin variable gene repression." *J Exp Med* 209 (12):2199-213.

Quinlan, A. R., and I. M. Hall. 2010. "BEDTools: a flexible suite of utilities for comparing genomic features." *Bioinformatics* 26 (6):841-2.

Quong, M. W., A. Martensson, A. W. Langerak, R. R. Rivera, D. Nemazee, and C. Murre. 2004. "Receptor editing and marginal zone B cell development are regulated by the helix-loop-helix protein, E2A." *J Exp Med* 199 (8):1101-12.

Ragoczy, T., and M. Groudine. 2010. "Getting connected in the globin interactome." *Nat Genet* 42 (1):16-7.

Ranganath, S., A. C. Carpenter, M. Gleason, A. C. Shaw, C. H. Bassing, and F. W. Alt. 2008. "Productive coupling of accessible Vbeta14 segments and DJbeta complexes determines the frequency of Vbeta14 rearrangement." *J Immunol* 180 (4):2339-46.

Rao, S. S., M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. L. Aiden. 2014. "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping." *Cell* 159 (7):1665-80.

Redrup, L., M. R. Branco, E. R. Perdeaux, C. Krueger, A. Lewis, F. Santos, T. Nagano, B. S. Cobb, P. Fraser, and W. Reik. 2009. "The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing." *Development* 136 (4):525-30.

Reinius, Björn, and Rickard Sandberg. 2015. "Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation." *Nat Rev Genet* 16 (11):653-664.

Reyes, J. C., C. Muchardt, and M. Yaniv. 1997. "Components of the human SWI/SNF complex are enriched in active chromatin and are associated with the nuclear matrix." *The Journal of Cell Biology* 137 (2):263-274.

Ribeiro de Almeida, C., R. Stadhouders, M. J. de Bruijn, I. M. Bergen, S. Thongjuea, B. Lenhard, W. van Ijcken, F. Grosveld, N. Galjart, E. Soler, and R. W. Hendriks. 2011. "The DNA-binding protein CTCF limits proximal Vkappa recombination and restricts kappa enhancer interactions to the immunoglobulin kappa light chain locus." *Immunity* 35 (4):501-13.

Rieger, M. A., and T. Schroeder. 2012. "Hematopoiesis." *Cold Spring Harb Perspect Biol* 4 (12).

Roldan, E., M. Fuxa, W. Chong, D. Martinez, M. Novatchkova, M. Busslinger, and J. A. Skok. 2005. "Locus 'decontraction' and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene." *Nat Immunol* 6 (1):31-41.

Rolink, A., U. Grawunder, D. Haasner, A. Strasser, and F. Melchers. 1993. "Immature surface Ig+ B cells can continue to rearrange kappa and lambda L chain gene loci." *J Exp Med* 178 (4):1263-70.

Romanow, W. J., A. W. Langerak, P. Goebel, I. L. Wolvers-Tettero, J. J. van Dongen, A. J. Feeney, and C. Murre. 2000. "E2A and EBF act in synergy with the V(D)J recombinase to generate a diverse immunoglobulin repertoire in nonlymphoid cells." *Mol Cell* 5 (2):343-53.

Rotem, A., O. Ram, N. Shoresh, R. A. Sperling, A. Goren, D. A. Weitz, and B. E. Bernstein. 2015. "Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state." *Nat Biotechnol* 33 (11):1165-72.

Saksouk, N., E. Simboeck, and J. Dejardin. 2015. "Constitutive heterochromatin formation and transcription in mammals." *Epigenetics Chromatin* 8:3.

Salles, Patricy de Andrade, Juliana Chagas Fortes, Maria Izabel Florindo Guedes, and Mathias Weller. 2007. "Expression of D-type cyclins in differentiating cells of the mouse spinal cord." *Genetics and Molecular Biology* 30:702-708.

Sanborn, A. L., S. S. Rao, S. C. Huang, N. C. Durand, M. H. Huntley, A. I. Jewett, I. D. Bochkov, D. Chinnappan, A. Cutkosky, J. Li, K. P. Geeting, A. Gnirke, A. Melnikov, D. McKenna, E. K. Stamenova, E. S. Lander, and E. L. Aiden. 2015. "Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes." *Proc Natl Acad Sci U S A* 112 (47):E6456-65.

Sato, H., F. Saito-Ohara, J. Inazawa, and A. Kudo. 2004. "Pax-5 is essential for kappa sterile transcription during Ig kappa chain gene rearrangement." *J Immunol* 172 (8):4858-65.

Savova, V., J. Patsenker, S. Vigneau, and A. A. Gimelbrant. 2016. "dbMAE: the database of autosomal monoallelic expression." *Nucleic Acids Res* 44 (D1):D753-6.

Sawasdichai, A., H. T. Chen, N. Abdul Hamid, P. S. Jayaraman, and K. Gaston. 2010. "In situ subcellular fractionation of adherent and non-adherent mammalian cells." *J Vis Exp* (41).

Schaniel, C., M. Gottar, E. Roosnek, F. Melchers, and A. G. Rolink. 2002. "Extensive in vivo self-renewal, long-term reconstitution capacity, and hematopoietic multipotency of Pax5-deficient precursor B-cell clones." *Blood.* 99 (8):2760-6.

Schatz, D.G., and Y. Ji. 2011. "Recombination centers and the orchestration of V(D)J recombination." *Nat Rev Immunol* 11:251-263.

Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. "NIH Image to ImageJ: 25 years of image analysis." *Nat Methods* 9 (7):671-5.

Schoenfelder, Stefan, Tom Sexton, Lyubomira Chakalova, Nathan F. Cope, Alice Horton, Simon Andrews, Sreenivasulu Kurukuti, Jennifer A. Mitchell, David Umlauf, Daniela S. Dimitrova, Christopher H. Eskiw, Yanquan Luo, Chia-Lin Wei, Yijun Ruan, James J. Bieker, and Peter Fraser. 2010. "Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells." *Nature Genetics* 42 (1):53-61.

Schwickert, T. A., H. Tagoh, S. Gultekin, A. Dakic, E. Axelsson, M. Minnich, A. Ebert, B. Werner, M. Roth, L. Cimmino, R. A. Dickins, J. Zuber, M. Jaritz, and M. Busslinger. 2014. "Stage-specific control of early B cell development by the transcription factor Ikaros." *Nat Immunol* 15 (3):283-93.

Seet, C.S., R.L. Brumbaugh, and B. L. Kee. 2004. "Early B cell factor promotes B lymphopoiesis with reduced interleukin 7 responsiveness in the absence of E2A." *J Exp Med* 199:1689-1700.

Sellars, M., P. Kastner, and S. Chan. 2011. "Ikaros in B cell development and function." *World J Biol Chem* 2 (6):132-9.

Shaw, A.C., W. Swat, R. Ferrini, L. Davidson, and F.W. Alt. 1999. "Activated Ras signals developmental progression of recombinase-activating gene (RAG)-deficient pro-B lymphocytes." *J Exp Med* 189 (1):123-129.

Shih, H. Y., and M. S. Krangel. 2013. "Chromatin architecture, CCCTC-binding factor, and V(D)J recombination: managing long-distance relationships at antigen receptor loci." *J Immunol* 190 (10):4915-21.

Sicinska, E., I. Aifantis, L. Le Cam, W. Swat, C. Borowski, Q. Yu, A. A. Ferrando, S. D. Levin, Y. Geng, H. von Boehmer, and P. Sicinski. 2003. "Requirement for cyclin D3 in lymphocyte development and T cell leukemias." *Cancer Cell* 4 (6):451-61.

Simon, I., T. Tenzen, B. E. Reubinoff, D. Hillman, J. R. McCarrey, and H. Cedar. 1999. "Asynchronous replication of imprinted genes is established in the gametes and maintained during development." *Nature* 401 (6756):929-932.

Singh, N., F. A. Ebrahimi, A. A. Gimelbrant, A. W. Ensminger, M. R. Tackett, P. Qi, J. Gribnau, and A. Chess. 2003. "Coordination of the random asynchronous replication of autosomal loci." *Nat Genet* 33 (3):339-41.

Skok, J.A., K. Brown, V. Azura, M.-L. Caparros, J. Baxter, K. Takacs, N. Dilon, D. Gray, R. Perry, M. Merkenschlager, and A.G. Fisher. 2001. "Nonequivalent nuclear location of immunoglobulin alleles in B lymphocytes." *Nat Immunol* 2:8484-854.

Srinivasan, L., Y. Sasaki, D.P. Calado, B. Zhang, J.H. Paik, R.A. DePinho, J.L. Kutok, J. F. Kearney, K.L. Otipoby, and K. Rajewsky. 2009. "PI3 kinase signals BCR-dependent mature B cell survival." *Cell* 139:573-586.

Stadhouders, Ralph, Marjolein J. W. de Bruijn, Magdalena B. Rother, Saravanan Yuvaraj, Claudia Ribeiro de Almeida, Petros Kolovos, Menno C. Van Zelm, Wilfred van Ijcken, Frank Grosveld, Eric Soler, and Rudi W. Hendriks. 2014. "Pre-B cell receptor signaling induces immunoglobulin κ locus accessibility by functional redistribution of enhancer-mediated chromatin interactions." *PLoS biology* 12 (2):e1001791.

Stanhope-Baker, P., K. Hudson, A. L. Shaffer, A. Constantinescu, and M. Schlissel. 1996. "Cell type-specific chromatin structure determines the targeting of V(D)J recombinase activitiy in vitro." *Cell* 85:887-897.

Staufenbiel, M., and W. Deppert. 1984. "Preparation of nuclear matrices from cultured cells: subfractionation of nuclei in situ." *J Cell Biol* 98 (5):1886-94.

Sutherland, H., and W. A. Bickmore. 2009. "Transcription factories: gene expression in unions?" *Nat Rev Genet* 10 (7):457-66.

Takagi, N. 1974. "Differentiation of X chromosomes in early female mouse embryos." *Exp Cell Res* 86 (1):127-35.

Thompson, E.C., 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. 2007. "Ikaros DNA-binding proteins as integral components of B cell developmental-stage-specific regulatory circuits." *Immunity* 26:335-344.

Tian, J., T. Okabe, T. Miyazaki, S. Takeshita, and A. Kudo. 1997. "Pax-5 is identical to EBB-1/KLP and binds to the VpreB and lambda5 promoters as well as the KI and KII sites upstream of the Jkappa genes." *Eur J Immunol* 27 (3):750-5.

Tiegs, S. L., D. M. Russell, and D. Nemazee. 1993. "Receptor editing in self-reactive bone marrow B cells." *J Exp Med* 177:1009-1020.

Tokoyoda, Koji, Takeshi Egawa, Tatsuki Sugiyama, Byung-Il Choi, and Takashi Nagasawa. 2004. "Cellular niches controlling B lymphocyte behavior within bone marrow during development." *Immunity* 20 (6):707-718.

van Loo, Pieter Fokko, Gemma M. Dingjan, Alex Maas, and Rudi W. Hendriks. 2007. "Surrogate-light-chain silencing is not critical for the limitation of pre-B cell expansion but is for the termination of constitutive signaling." *Immunity* 27 (3):468-480.

Verma-Gaur, J., A. Torkamani, L. Schaffer, S. R. Head, N. J. Schork, and A. J. Feeney. 2012. "Noncoding transcription within the Igh distal V(H) region at PAIR elements affects the 3D structure of the Igh locus in pro-B cells." *Proc Natl Acad Sci U S A* 109 (42):17004-9.

Vettermann, C., and M.S. Schlissel. 2010. "Allelic exclusion of immunoglobulin genes: models and mechanisms." *Immunol Rev* 237:22-42.

Vettermann, Christian, Greg A. Timblin, Vivian Lim, Ernest C. Lai, and Mark S. Schlissel. 2015. "The proximal J kappa germline-transcript promoter facilitates receptor editing through control of ordered recombination." *PloS One* 10 (1):e0113824.

Wang, Tian-Yun, Zhong-Min Han, Yu-Rong Chai, and Jun-He Zhang. 2010. "A mini review of MAR-binding proteins." *Molecular Biology Reports* 37 (7):3553-3560.

Weeks, John D, David Chandler, and Hans C Andersen. 1971. "Role of repulsive forces in determining the equilibrium structure of simple liquids." *J Chem Phys* 54 (12):5237-5247.

Werner, A. 2013. "Biological functions of natural antisense transcripts." *BMC Biol* 11:31.

Whyte, W. A., D. A. Orlando, D. Hnisz, B. J. Abraham, C. Y. Lin, M. H. Kagey, P. B. Rahl, T. I. Lee, and R. A. Young. 2013. "Master transcription factors and mediator establish super-enhancers at key cell identity genes." *Cell* 153 (2):307-19.

Wilson, Rosemary H. C., and Dawn Coverley. 2013. "Relationship between DNA replication and the nuclear matrix." *Genes to Cells: Devoted to Molecular & Cellular Mechanisms* 18 (1):17-31.

Xiang, Y., S. K. Park, and W. T. Garrard. 2013. "Vkappa gene repertoire and locus contraction are specified by critical DNase I hypersensitive sites within the Vkappa-Jkappa intervening region." *J Immunol* 190 (4):1819-26.

Xiang, Y., S. K. Park, and W. T. Garrard. 2014. "A major deletion in the Vkappa-Jkappa intervening region results in hyperelevated transcription of proximal Vkappa genes and a severely restricted repertoire." *J Immunol* 193 (7):3746-54.

Xu, C. R., and A. J. Feeney. 2009. "The epigenetic profile of Ig genes is dynamically regulated during B cell differentiation and is modulated by pre-B cell receptor signaling." *J Immunol* 182 (3):1362-9.

Xu, S., K. G. Lee, J. Huo, T. Kurosaki, and K. P. Lam. 2007. "Combined deficiencies in Bruton tyrosine kinase and phospholipase Cgamma2 arrest B-cell development at a pre-BCR+ stage." *Blood* 109 (8):3377-84.

Yalcin, B., K. Wong, A. Agam, M. Goodson, T. M. Keane, X. Gan, C. Nellaker, L. Goodstadt, J. Nicod, A. Bhomra, P. Hernandez-Pliego, H. Whitley, J. Cleak, R. Dutton, D. Janowitz, R. Mott, D. J. Adams, and J. Flint. 2011. "Sequence-based characterization of structural variation in the mouse genome." *Nature* 477 (7364):326-9.

Yancopoulos, G. D., and F. W. Alt. 1985. "Developmentally controlled and tissue-specific expression of unrearranged VH gene segments." *Cell* 40 (2):271-281.

Yao, Z., 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. 2006. "Stat5a/b are essential for normal lymphoid development and differentiation." *Proc Nat Acad Sci* 103 (4):1000-1005.

Yasui, D., M. Miyano, S. Cai, P. Varga-Weisz, and T. Kohwi-Shigematsu. 2002. "SATB1 targets chromatin remodelling to regulate genes over long distances." *Nature* 419 (6907):641-5.

Yi, M., P. Wu, K. W. Trevorrow, L. Claflin, and W. T. Garrard. 1999. "Evidence that the Igkappa gene MAR regulates the probability of premature V-J joining and somatic hypermutation." *Journal of Immunology (Baltimore, Md.: 1950)* 162 (10):6029-6039.

Young, F., B. Ardman, W. Shinkai, R. Lansford, T.K. Blackwell, M. Mendelshn, A. Rolink, F. Melchers, and F.W. Alt. 1994. "Influnence of immunoglobulin heavy- and light-chain expression on B-cell differentation." *Genes Dev.* 8:1043-1057.

Zandi, S., R. Mansson, P. Tsapogas, J. Zetterblad, D. Bryder, and M. Sigvardsson. 2008. "EBF1 is essential for B-lineage priming and establishment of a transcription factor network in common lymphoid progenitors." *J Immunol* 181 (5):3364-72.

Zeitlin, S., A. Parent, S. Silverstein, and A. Efstratiadis. 1987. "Pre-mRNA splicing and the nuclear matrix." *Molecular and Cellular Biology* 7 (1):111-120.

Zhang, X., N. Tang, T. J. Hadden, and A. K. Rishi. 2011. "Akt, FoxO and regulation of apoptosis." *Biochim Biophys Acta* 1813 (11):1978-86.

Zhu, C., K. D. Mills, D. O. Ferguson, C. Lee, J. Manis, J. Fleming, Y. Gao, C. C. Morton, and F. W. Alt. 2002. "Unrepaired DNA breaks in p53-deficient cells lead to oncogenic gene amplification subsequent to translocations." *Cell* 109 (7):811-21.

Zink, D., A. H. Fischer, and J. A. Nickerson. 2004. "Nuclear structure in cancer cells." *Nat Rev Cancer* 4 (9):677-87.

Zlatanova, J., and P. Caiafa. 2009. "CTCF and its protein partners: divide and rule?" *J Cell Sci* 122 (Pt 9):1275-84.