REGULATION OF ACCESSIBILITY OF THE VARIABLE GENE SEGMENTS OF THE MOUSE IMMUNOGLOBULIN KAPPA LIGHT CHAIN GENE LOCUS

APPROVED BY SUPERVISORY COMMITTEE

_____

William T. Garrard, Ph.D.

_____

Margaret A. Phillips, Ph.D.

_____

Kristen W. Lynch, Ph.D.

_____

Richard H. Scheuermann, Ph.D.

REGULATION OF ACCESSIBILITY OF THE VARIABLE GENE SEGMENTS OF THE MOUSE IMMUNOGLOBULIN KAPPA LIGHT CHAIN GENE LOCUS

by

KATHERINE MEYERS BREKKE

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

August 2005

REGULATION OF ACCESSIBILITY OF THE VARIABLE GENE SEGMENTS OF THE MOUSE IMMUNOGLOBULIN KAPPA LIGHT CHAIN GENE LOCUS

Publication No. _____

Katherine Meyers Brekke, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2005

Supervising Professor: William T. Garrard, Ph.D.

Understanding the development of the expression of antibody heavy and light chain molecules is required to fully explain the adaptive immune response. This involves understanding the mechanisms underlying the selection of individual Vκ genes for V-J joining. It has been assumed that this selection involves alterations in the chromatin structure surrounding Vκ genes. Therefore, the focus of my work has centered on understanding the factors influencing the selection of a particular Vκ gene as well as the accompanying alterations in chromatin structure. I have assembled the complete nucleotide sequence of the 3.2 Mb *Mus musculus* Igκ gene locus and mapped all of the functional and pseudo- Vκ genes onto the sequence. I have also determined the patterns of potential transcription factor

binding sites surrounding these Vκ genes. These analyses revealed statistically significant patterns of transcription factor motifs that are clustered either upstream of the transcription start site, in the intronic region, or downstream of the recombination signal sequence (RSS). The conservation of these sites has been tied to the presence of DNase I hypersensitive sites at the pre-B cell stage of development. These hypersensitive sites appear just prior to the V-J joining step and disappear after the Igκ locus has completed the rearrangement process. These sites are present at or near the promoter of the particular Vκ gene and are sometimes present near the RSS as well. I also looked for the presence of transcripts from the promoters of the Vκ genes that had HS_P and HS_{RSS}. I found that the presence of transcripts and the presence of HS was uncoupled. There were instances where both HS and transcripts were present, neither was present, or only one phenomenon was observed. This led me to conclude that these events were separable. The presence of HS_{RSS} as well as the corresponding binding sites for transcription factors in the downstream region, suggests a model for generating accessibility of Igκ V genes for V-J joining.

TABLE OF CONTENTS

Prior Publications

Li, S., R. E. Hammer, J. B. George-Raizen, K. C. Meyers and W. T. Garrard. 2000. High-level rearrangement and transcription of yeast artificial chromosome-based mouse Igκ transgenes containing distal regions of the contig. *J. Immunol. 164:812.*

Liu, Z-.M., J. B. George-Raizen, S. Li, K. C. Meyers, M. Y. Chang, and W. T. Garrard. 2002. Chromatin structural analyses of the mouse Igκ gene locus reveal new hypersensitive sites specifying a transcriptional silencer and enhancer. *J. Biol. Chem. 277:32640.*

Brekke, K.M., and W.T. Garrard. 2004. Assembly and analysis of the mouse immunoglobulin kappa gene sequence. *Immunogenetics. 56:490*

List of Figures

List of Tables

CHAPTER ONE

General Introduction

The immunoglobulin (Ig) and T cell receptor (TCR) genes provide a favorable system for teasing out the mechanisms of regulation of multi-gene loci and understanding tissue-specific gene expression. The Igκ locus in mouse is generates 95% of the light chain molecules (McIntire and Rouse, 1970). More that 10^{10} antibody molecules with different specificities are generated by the B cell compartment (Tonegawa, 1983). The tightly regulated process that generates antibody molecules is referred to as V(D)J recombination. V(D)J recombination is both tissue-specific and developmental stage specific. Numerous *cis*-acting elements have been described that regulate this process, but the methods by which the chromatin becomes accessible for the recombination machinery are relatively unknown. This suggests that there is a significant level of regulation of the accessibility of the Vκ genes in the locus.

Immunoglobulin Gene Structure and V(D)J Recombination

In most instances, an individual immunoglobulin molecule is a tetramer of two identical heavy chains and two identical light chains. In the mouse, the heavy chain can be one of eight isotypes, μ, δ, γ1, γ2a, γ2b, γ3, α, or ε. Each is generated from the same locus. The light chain can be either of the κ or λ isotype, each of which is generated by a separate locus (reviewed by (Spiegelberg, 1974). The nomenclature in the field names the highly divergent

amino terminal domains, variable (V) regions and the conserved, carboxyl terminal end the constant (C) region. Heavy chain molecules also have an intermediate, diversity (D) segment between the variable and constant regions. Each IgH and IgL molecule has a single joining (J) region adjacent to the C region (reviewed by (Blackwell and Alt, 1989; Lieber, 1992; Shapiro and Weigert, 1987)).

B Cell Development. B cell development is generally defined by the status of Ig gene rearrangement as well as the accompanying expression of cell surface markers, size and growth rate (reviewed by (Burrows and Cooper, 1997; Ghia et al., 1998; Loffert et al., 1994; Papvasiliou et al., 1997; Rolink et al., 1995)). In adult mammals, B lymphocytes are derived from precursor stem cells in the bone marrow. The pro-B cell is the earliest identifiable stage of B cell development. It is during this stage when IgH rearrangement occurs, first by DJ joining, then V to DJ (Alt et al., 1981; Alt et al., 1984; Kamps and Cooper, 1982; Levitt and Cooper, 1980; Perry et al., 1981; Siden et al., 1981). During this time the cells are expressing terminal deoxynucleotidyl transferase (TdT), and the recombination activating genes (RAG-1 and –2) (Komori et al., 1993; Oettinger et al., 1990; Rolink et al., 1993; Rolink et al., 1991; Schatz et al., 1989; ten Boekel et al., 1995). Successful rearrangement of the IgH locus marks the transition to the pre-B cell stage. During this time, the IgH molecule is expressed on the cell surface in conjunction with surrogate light chain molecules (Karasuyama et al., 1990; Kudo and Melchers, 1987; Papvasiliou et al., 1996; Sakaguchi and Melchers, 1986) and reviewed in (Melchers et al., 1993). This expression of μ protein initiates a feedback control mechanism, allelic exclusion, that prevents further rearrangement

or expression of the other allele (Alt et al., 1981; Coleclough et al., 1981). During the pre-B cell stage, RAG-1 and –2 expression is decreased and increased again and the IgL loci rearrange (Bauer and Scheuermann, 1993; Constantinescu and Schlissel, 1997). After a functional IgL chain is created, antigen-independent B cell development is complete, immature B cells leave the bone marrow and undergo further maturation in the periphery (reviewed by (Henderson and Calame, 1998; Meffre et al., 2000)).

The Mechanism of V(D)J Recombination. V(D)J recombination occurs only in B and T lymphocytes. Ig genes are only rearranged and expressed in B cells. V(D)J recombination is a tissue-, lineage-, and developmental stage-specific process that uses a common set of protein machinery in the cells in which it occurs (reviewed by (Sleckman et al., 1996)). Each gene segment (V, D, and J) is flanked by a recombination signal sequence (RSS). An RSS is made up of a conserved heptamer and an AT-rich conserved nonamer separated by either a 12 or a 23 bp nonconserved "spacer" (Early et al., 1980; Sakano et al., 1981). The RSS associated with each coding region is both necessary and sufficient to direct the recombinase to the corresponding segment in a non-chromatin context (Akira et al., 1987; Hesse et al., 1989; Wei and Lieber, 1993). The 12 and 23 bp spacers also restrict the process to the correct segments. Recombination only occurs between a 12 bp spacer segment and one with a 23 bp spacer. This is known as the "12/23 rule" (Early et al., 1980; vanGent et al., 1996). The protein Artemis is responsible for opening the hairpin formed by cleavage of the RSS. This step occurs with the assistance of Ku and DNA-PK (Karanjawala et al., 2002; Ma et al., 2002). Once the recombinase recognizes the RSS, there is the introduction of a double strand

break (DSB) between the coding regions from each of the segments and then joining of the coding ends. V(D)J recombination is evolutionarily conserved to the point that immunoglobulin genes from chicken and rabbit, as well as human can be recombined in mice (Bucchini et al., 1987; Goodhardt et al., 1987).

This process uses a common set of protein machinery in all of the cell types in which it occurs. The proteins involved in this process are RAG-1, RAG-2, TdT, DNA-protein kinase (DNA-PK), Artemis, Ku70, Ku-80, XRCC4 and DNA ligase IV, (Bassing et al., 2002; Fugmann et al., 2000; Ma et al., 2002; Ramsden et al., 1997). The products of the recombination activating genes, RAG-1 and –2 initiate recombination by binding to and cleaving the RSSs flanking the recombining gene segments. They also initiate the formation of hairpins at the coding joint precursor fragment ends (Agrawal and Schatz, 1997; Bailin et al., 1999; Besmer et al., 1998; Gallo et al., 1994; Hiom and Gellert, 1997; Kirch et al., 1998; McBlane et al., 1995; Oettinger, 1992; Oettinger et al., 1990; Schatz et al., 1989; Swanson and Desiderio, 1998). HMG1 and HMG2 interact with the RAG proteins to facilitate binding and the recombination process and help to enforce the 12/23 rule (Aidinis et al., 1999; Kwon et al., 1998; Shirakata et al., 1991; vanGent et al., 1997; West and Lieber, 1998). Artemis then cleaves the hairpins caused by RAG-1/-2 and the Ku protein, a heterodimer of Ku70 and Ku80, is involved in binding the ends of the cleaved DNA (Gao et al., 1998; Gu et al., 1997; Ma et al., 2002; Pergola et al., 1993; Taccioli et al., 1993) and DNA ligase IV and XRCC4 are involved in rejoining the ends to form rearranged Ig genes (Leu et al., 1997; Li et al., 1995; McElhinny et al., 2000; Ramsden and Gellert, 1998). XRCC4 functions by forming a complex with DNA ligase IV and enhancing its activity (Critchlow et al., 1997; Grawunder

et al., 1997; Li et al., 1995; Modesti et al., 1999). The biological targets for DNA-PK in V(D)J recombination are not known, but it is required to reconstitute the reaction *in vitro* and the knockout mice display a block in the hairpin cutting stage of the reaction (Bogue et al., 1998; Hartley et al., 1995; Rukumara et al., 2000; Taccioli et al., 1998). Two other proteins have been shown to be involved in cellular signaling of DNA damage and are thought to have a role in the joining reaction, but their roles are not well understood. These proteins are Nbs1 and the histone H2A variant H2A.X (Chen et al., 2000; Huang et al., 1986; Moshous et al., 2001; Paull and Gellert, 2000; Paull et al., 2000; Schlissel, 2002)

Structure and organization. Ig loci are all organized in a similar manner. In each immunoglobulin locus, the C exons are the most 3' segment(s) (Fig. 1.1). There are several known enhancers in these regions. Upstream of the C region are the J segments, next the D segments (if present) followed by widely dispersed V genes in the most 5' regions of the loci. There are no known conserved global regulatory elements in the 5' regions of these genes (reviewed by (Lansford et al., 1996)). There are regulatory elements surrounding each of the V genes, but not any known enhancers or locus control regions (LCRs). I will discuss the regulatory regions surrounding the V genes in the next two sections.

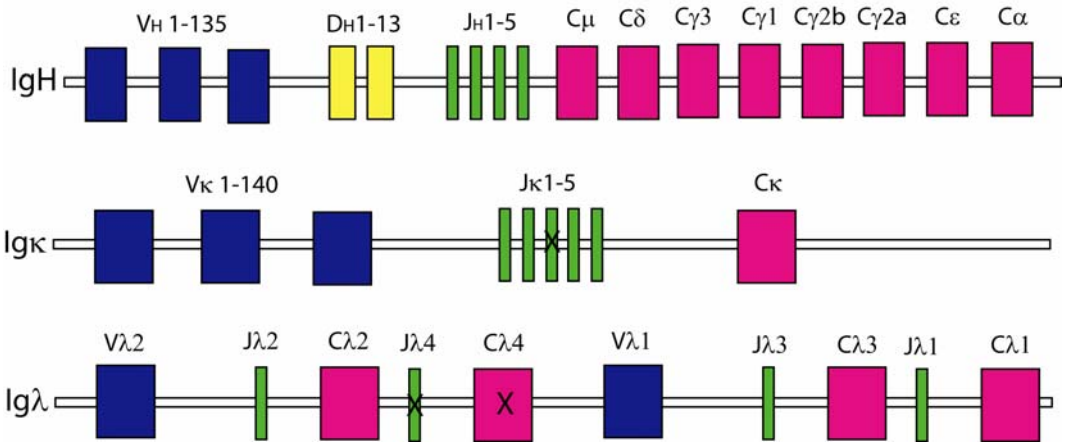

Figure 1.1 – Cartoon representations of the three immunoglobulin loci of mouse. Drawings are not to scale. The black X's indicate non-functional joining or constant regions.

The mouse Igκ locus is not only the largest antigen receptor locus, it is also the largest known single gene (George et al., 1995; Schupp et al., 1997). It is located on chromosome 6. Igκ has five Jκ regions, one of which (Jκ3) is non-functional due to a cryptic RSS (Tonegawa, 1983). Most recent estimates have suggested that there are 140 V genes in the estimated 3.3 Mb of DNA in the 5' portion of the locus (George et al., 1995; Kirschbaum et al., 1996; Schable et al., 1999). These Vκ gene segments are both clustered in families and interspersed amongst each other (Kofler et al., 1992). Vκ genes exist in both transcriptional orientations relative to Cκ. This means that rearrangement can occur either by deleting a portion of the gene, or by inversion (Zachau, 1989).

cis-acting Regulatory Elements in Igκ

The mouse Igκ locus contains a number of well-defined *cis*-acting regulatory elements that have been characterized by previous studies. These include the promoters and RSSs that are located at each individual V gene throughout the locus, three enhancer elements, a matrix association region (MAR), two germline promoters, the KI/KII elements, the intervening sequence silencer (Sis), and the downstream recombining sequence (RS). With the exception of V genes and their cognate promoters and RSSs, all of these elements are located in the most 3' 30 kb region of this 3.2 Mb locus. In the next section, I will review these elements and their functions in the regulation of Igκ.

The enhancers

The intronic enhancer (iEκ). iEκ, the first enhancer identified in Igκ, is located between the Jκ region and the Cκ exon and is sufficient to direct lymphoid specific expression of a heterologous sequence (Bergman et al., 1984; Emorine et al., 1983; Picard and Schaffner, 1984; Queen and Baltimore, 1983). This enhancer contains a κB element, a binding site for NF-κB (Lenardo et al., 1987; Sen and Baltimore, 1986a; Sen and Baltimore, 1986b) and three E box sequences (Ephrussi et al., 1985; Lenardo et al., 1987), all of which are required for full enhancer activity. This enhancer is initially activated at the pre-B cell stage, when Igκ is undergoing V-J joining, and stays active throughout the remaining stages of B cell development. iEκ has been implicated in the demethylation of the Igκ locus.

Methylated constructs containing a rearranged Vκ gene sequence were transfected into cultured B cells. It was found that the constructs containing iEκ underwent stage-specific demethylation, while those missing the enhancer sequence did not (Lichtenstein et al., 1994). To further delineate the function of iEκ, this enhancer was deleted from mice by a Cre-loxP strategy (Xu et al., 1996). This deletion demonstrated that iEκ is not required for V-J joining, but required to maintain the normal levels of κ+ B cells; the κ:λ ratio in the knockout mice was decreased from the normal 10:1, to a 1:1 ratio (Xu et al., 1996). An earlier study deleted iEκ by inserting a *neo*^{r} gene in place of the enhancer (Takeda et al., 1993). In these mice, V-J joining was completely ablated. However, placement of the *neo*^{r} gene 3' to iEκ also resulted in the same phenotype. Given that the later work with a Cre-loxP knockout demonstrated a less severe phenotype (Xu et al., 1996), it suggests that the first study did not directly address the role of iEκ in the recombination process.

The 3' enhancer (E3'). A second enhancer (E3'), was identified 9 kb downstream of the Cκ exon (Meyer and Neuberger, 1989). Several lines of evidence suggested the existence of another enhancer. These included: deletion studies of rearranged Igκ transgenes in plasmacytoma cells and transgenic mice (Blasquez et al., 1989; Xu et al., 1989), high level transcription of endogenous, but not transfected Igκ genes in cells lacking NF-κB (Atchison and Perry, 1987; Atchison and Perry, 1988).

The downstream enhancer (Ed). A third enhancer has recently been identified in the Igκ locus (Liu et al., 2002). This enhancer was uncovered by DNase I hypersensitive site

analysis of the downstream region of Igκ and was found to be conserved in the downstream region of the human Igκ locus. This DNase I mapping was carried out to identify factors that could act as boundary elements or further regulate the locus in different ways. It was suggested that the function of this enhancer is still later in development than iEκ or E3', because only terminally differentiated plasmacytoma cells exhibit the hypersensitive site (Liu et al., 2002). The factor binding sites crucial for the activity of this enhancer are NF-κB and E2A preferred E box sequences.

The matrix association region (MAR). A nuclear matrix association region (MAR) is located 5' of iEκ (Cockerill and Garrard, 1986). This MAR/enhancer pattern is conserved in human and rabbit Igκ genes suggesting an important role in Igκ regulation (Sperry et al., 1989; Whitehurst et al., 1992). These sequences are greater that 70% A-T and bind chromosomal proteins involved in forming higher order nuclear structures (Sperry et al., 1989). The Igκ MAR increases the expression of Igκ and can protect integrated constructs against position effects (Blasquez et al., 1989; Xu et al., 1989). A cell line with the MAR deleted on one allele exhibited hyper-recombination (Hale and Garrard, 1998); in mice the MAR deletion led to premature V-J joining in pro-B cells and reduced somatic hypermutation (Yi et al., 1999). These results suggest a negative role for the MAR in V-J joining and a positive role in hypermutation.

The intervening sequence silencer (Sis). A silencing element is present in the intervening sequence between Vκ21-1 and the Jκ genes (Liu et al., 2002). This element was

also identified in a screen for novel regulatory elements in the Igκ locus. Its silencing activity occurs specifically in pre-B cells in a segment of DNA that is 3.6 kb in length. The sequence is A/T rich, and contains three potential Ikaros binding sites.

Germline promoters. Two promoters reside upstream of the Jκ region. These germline promoters are B cell specific in their activities and produce transcripts that originate either 3.5 kb or 50 bp upstream of the Jκ region (Leclerq et al., 1989; Martin and VanNess, 1990; VanNess et al., 1981). These transcripts are spliced and expressed at high levels prior to the rearrangement event, but are not translated (Martin and VanNess, 1990). It has been hypothesized that germline transcription plays a role in determining the accessibility of the chromatin prior to the recombination event. These two events are concomitant (Schlissel and Baltimore, 1989; Schlissel et al., 1991b), but have not been shown to be coupled or causal. Germline transcription has been shown to be biallelic on both the heavy chain and κ locus. It was thought that germline transcription did not play a role in determining which allele will be selected first for rearrangement (Delpy et al., 2003; Singh et al., 2003). However, more recent evidence with a GFP-Igκ knock-in, has shown that a single allele is transcribed from these promoters and that the transcribed allele is more likely to undergo V-J joining (Liang et al., 2004).

V region promoters. Each of the 95 upstream variable gene segments contains its own cognate promoter (Bergman et al., 1984; Falkner and Zachau, 1984; Parslow et al., 1984). These promoters each contain their own TATA box (Bergman et al., 1984) and an

octamer motif (Falkner and Zachau, 1984; Parslow et al., 1984) located within 70 bp of the transcription start site. Transcription from these promoters occurs in pre-B cells and initiates from the same place as in rearranged genes (Schlissel et al., 1991b).

The KI/KII elements. A *cis*-acting element that is located immediately upstream of Jκ1 and is composed of two palindromic motifs, termed KI and KII appears to regulate V-J joining (Weaver and Baltimore, 1987). The targeted mutation of this site in mice resulted in an 80% reduction of rearrangement on the mutated allele, while not affecting transcription (Ferradini et al., 1996). Similar results were observed in the 38B9 cell line; no reduction in transcription from either of the germline promoters and a 50% reduction in rearrangement on the targeted allele (Liu and VanNess, 1999). The product of the Pax-5 gene has been shown to bind to these sequences and likely plays a role in regulating Igκ rearrangement (Tian et al., 2004).

The recombining sequence (RS). A cryptic recombination sequence resides downstream of Ed. Igλ-producing B cells frequently exhibit several types of Igκ rearrangements to this sequence (Durdik et al., 1984; Moore et al., 1985; Nadel et al., 1990; Shimizu et al., 1991), suggesting it plays a role in regulating the κ/λ ratio and selection. Recombination to the RS has shown to occur at high frequencies by single cell PCR experiments; 33% of early bone marrow B cells and 15% of splenic B cells exhibit RS recombination (Yamagami et al., 1999). There is a homolog in the human locus, the kappa deleting element (Kde), that when rearranged to leads to rearrangement of the Igλ locus

(Feddersen et al., 1990; Graninger et al., 1988; Siminovitch et al., 1985). In mice, there is a unique transcript originating from the RS element in pre-B cells. There are also changes in chromatin structure and transcriptional activation status surrounding the RS element. These are similar to the changes that occur in other areas of the locus prior to V(D)J recombination (Daitch et al., 1992).

Transcription Factors in B Cell Development

A number of transcription factors play crucial roles in B cell development. Some of these factors are ubiquitous, others specific to lymphoid cells, or B cells themselves. Many of these factors are also active throughout B cell development, while some are on only during a limited time frame. In the following pages, I will review known factors and the time frame they are present in B cells.

Oct-1, -2. The conserved octamer element in the promoters of all Ig V genes is the first sequence that was identified as important and almost completely conserved in Ig gene transcription (Falkner and Zachau, 1984; Parslow et al., 1984). The proteins that bind to these sequences are the first identified members of the POU-domain family of proteins, Oct-1 and Oct-2 (Clerc et al., 1988; Muller et al., 1988; Scheidereit et al., 1988; Sturm et al., 1988). Oct-1 is a ubiquitous transcription factor and Oct-2 is restricted mainly to the lymphoid lineage with some expression also seen in the nervous system, kidney and testis. Mice

generated without Oct-2 die at birth for unknown reasons, but B cells develop normally in these mice through the IgM^+ stage (Corcoran et al., 1993).

The roles that Oct-1 and –2 play in transcriptional activation in B cells are not well understood. Although Oct-1 is ubiquitous and Oct-2 is lymphoid specific, both can activate reporter plasmids equally well in B cells (Annweiler et al., 1992). The Annweiler et al. paper also demonstrated that Oct-2 is unable to activate transcription in non-B cells and shows that Ig expression is not correlated completely with expression levels of Oct-2. An accessory molecule was identified that could bind to either Oct-1 or –2 and activate it in B cells. This protein was isolated by a number of groups simultaneously and is referred to as either OCA-B, OBF-1, or BOB-1 (Gstaiger et al., 1995; Luo and Roeder, 1995; Strubin et al., 1995). It was shown that this protein was able to activate octamer containing promoters in a B cell specific fashion (Luo and Roeder, 1995) and is expressed specifically in B cells. Oca-B appears at least to regulate a subset of Igκ genes.

Single and double knockouts of these genes lend some insight to the roles each of these proteins plays in murine development. The absence of Oca-B in mice leads to normal B cell development through the mature B cell stage, but the mice display defects in isotype switching and do not form germinal centers (Kim et al., 1996; Nielsen et al., 1996; Schubart et al., 1996). The overall level of transcription of Igκ or Igμ genes appears unaffected in the Oca-B -/- mice (Schubart et al., 2001). Analysis of mice missing both Oca-B and Oct-2 in the B cell compartment showed that B cell development could proceed normally through the mature cell stage, but there was a significant reduction of numbers of cells overall (Schubart et al., 2001). Closer, more detailed analysis of Oca-B -/- mice showed that Oca-B was

required for normal levels of transcription from some Vκ genes (Casellas et al., 2002). This study also showed that the repertoire of the Oca-B mice was significantly different from that from wild-type mice.

E2A. The E2A factor is a member of the basic helix-loop-helix (bHLH) family of transcription factors. E2A was originally identified by its ability to bind to a sequence (CANNTG) in the IgH intronic enhancer through DNA footprinting experiments (Ephrussi et al., 1985). The E2A transcript gives rise to two alternatively spliced products, E12 and E47, which are both ubiquitously expressed and generally bind DNA as heterodimers with cell type specific family members (Davis et al., 1987; Lassar et al., 1989; Lassar et al., 1991; Murre et al., 1989a; Murre et al., 1989b). Generally, these proteins are involved in regulating the expression of genes that direct patterns of development of a particular cell lineage.

E12 and E47 have well-documented roles in lymphoid and B cell development. In mice the absence of the E2A genes leads to a complete block in B cell development past the commitment stage (Bain et al., 1994; Zhuang et al., 1994). A further study demonstrates individual roles of E47 and E12 in B cell development (Bain et al., 1997). Here it is shown that mice expressing only E12 can undergo some reduced level of D_H-J_H joining, but then cease differentiation. However, expression of E47 alone can lead to a small fraction of cells continuing through to the mature B cell stage (Bain et al., 1997).

A bHLH factor specific to B cells has not been identified. The AP-4 factor has been shown to bind to Igκ promoter E-boxes in gel-shift assays (Aranburu et al., 2001). E47 forms heterodimers specifically in the B cell lineage (Shen and Kadesch, 1995) and

formation of these heterodimers is mediated by hyperphosphorylation of the E47 protein (Sloan et al., 1996). Interestingly, this specific dimer may be a B cell specific differentiation mechanism. It has been observed that the E2A gene products are not required for differentiation of ES cells into skeletal or cardiac muscle cells, erythrocytes, neurons or cartilage (Zhuang et al., 1992). But as the E2A knockout mouse has demonstrated, E2A is essential for populating the B cell compartment (Bain et al., 1994; Zhuang et al., 1994). It has also been shown that overexpression of E47 in a pre-T cell line is sufficient to induce D_H-J_H rearrangement (Schlissel et al., 1991a); its expression in either a pre-T cell line or NIH 3T3 fibroblasts leads to germline transcription of the IgH locus (Choi et al., 1996; Schlissel et al., 1991a).

Not only has E2A been identified as a factor that can regulate transcription of Ig loci, but it has been shown to regulate the recombination process as well. In a human embryonic kidney cell line, expression of either E12 or E47 in conjunction with RAG1 and RAG2, led to a detectable level of cleaved signal ends, as well as Igκ rearrangement products (Romanow et al., 2000). This activity appears to be a result of direct interaction with the chromatin of the Igκ locus, as there was no detectable expression of other factors involved in B cell development.

Ikaros. The Ikaros gene encodes a series of proteins that are generated by alternative splicing. These proteins bind DNA as either homo- or heterodimers through a kruppel-like zinc finger domain (Hahm et al., 1994; Molnar and Georgopoulos, 1994; Sun et al., 1996). Aiolos, another family member, appears to be involved in regulating T cell development

(Hahm et al., 1998; Morgan et al., 1997). Both Ikaros and Aiolos are expressed throughout B cell development (Georgopoulos et al., 1992; Morgan et al., 1997). Mice with a homozygous knockout of Ikaros display a very early block in B cell development, prior to the pro-B cell stage (Wang et al., 1996). The dominant-negative knock-in of Ikaros results in an even more severe phenotype, with a complete block in all lymphopoiesis, defects in erythroid and myeloid cell development and death within three weeks (Georgopoulos et al., 1994), suggesting important roles for Ikaros and related factors in development of a variety of cell types. It has also been demonstrated that Ikaros proteins can both activate and repress transcription depending on the context (Koipally et al., 1999; Molnar and Georgopoulos, 1994; Sun et al., 1996). This mediation appears to be through regulating the chromatin structure of the gene of interest (Hahm et al., 1998; Kim et al., 1999). Ikaros has been identified bound to transcriptionally silent genes (including Ig loci) in pericentric heterochromatin and is proposed to play a role in maintaining this silent state (Cobb et al., 2000; Ernst et al., 1996; Klug et al., 1998).

Lef-1. The lymphocyte enhancer factor 1 (Lef-1) is a member of the T cell-specific transcription factor (TCF) family of transcription factors, a subset of the high mobility group (HMG) family (Korinek et al., 1998; van de Wetering et al., 1991). Lef-1 is expressed early in B and T cell development (Oosterwegel et al., 1991; Travis et al., 1991; van Genderen et al., 1994). Lef-1 does not have its own transcriptional activation potential, but acts as a scaffold protein organizing members of the Wnt signaling pathway (Behrens et al., 1996; Bruhn et al., 1997; Giese et al., 1995; Huber et al., 1996). Lef-1 appears to play a role in

regulating mitosis of early B cells, as pro-B cell proliferation is impaired in Lef-1 knockout mice (Reya et al., 2000).

Pax-5. The product of the Pax-5 genes is the B cell specific activator protein (BSAP). This paired box transcription factor is expressed at the earliest stages of B cell development and continues through until the mature B cell stage. Pax-5 is not detected in nuclear extracts from antibody secreting plasma cells, T cells or erythroid cells (Barberis et al., 1990). The Pax-5 null mice display a block in the pro-B to pre-B cell transition, as well as neurological defects (Nutt et al., 1997; Urbanek et al., 1994). The IgH loci in these mice undergo D_H-J_H rearrangement, but arrest in development at the V_H-DJ_H stage. These mice express normal levels of E2A and Ebf, both crucial for B cell development (Nutt et al., 1997). One of BSAP's targets in B cell development is CD19, as it is the only gene detected with a severely altered expression pattern in these mice. This suggests a role for this gene being involved in activating genes responsible for the signaling pathways in B cell development. Further support for this idea comes from the fact that BSAP also regulates expression of the *mb*-1 gene (Nutt et al., 1998; Nutt et al., 1997). Another target is the KI /KII site in the V-J intervening sequence (Tian et al., 1997). BSAP binds this sequence and may regulate the activation of part of the Igκ gene. Although there are binding sites for BSAP in the promoter region of the surrogate light chain genes (Okabe et al., 1992), BSAP does not appear to have an effect on their regulation (Nutt et al., 1998; Nutt et al., 1997). Mice expressing ectopic Pax-5 under the control of the Ikaros promoter progress through B

cell development and prevent T lymphopoiesis by blocking the Notch activation pathway (Souabni et al., 2002).

NF-κB. NF-κB was first identified as a factor binding to the Igκ Ei and activating transcription of the Igκ gene (Sen and Baltimore, 1986b). The name NF-κB now refers to a family transcription factors that bind DNA as dimers in a variety of cell types on divergent genes. Each member of this family possesses a rel homology domain (RHD) that is responsible for DNA-binding and dimerization, reviewed in (Grilli et al., 1993; Kopp and Ghosh, 1995; Verma et al., 1995). This domain also contains the nuclear localization sequence (NLS) and is responsible for binding to IκB, the inhibitory subunits of NF-κB. The IκB inhibitors function by binding to the NF-κB subunit and masking its NLS, thereby sequestering it in the cytoplasm (Baeuerle and Baltimore, 1988). This mechanism of control allows for rapid activation of NF-κB at the appropriate time. All that is required is the degradation of the inhibitor and a rapid shuttling of the factor into the nucleus to activate transcription.

Because of the numerous roles of the NF-κB complexes and their inhibitors in a variety of systems from apoptosis to oncogenesis, I will limit my review to NF-κB activities and effects on Igκ gene transcription and early B cell development. It was originally observed that there was little detectable NF-κB activity in pre-B cell lines unless they were activated with lipopolysaccharide (LPS) or certain cytokines (Grilli et al., 1993). However, all of these cell lines were derived by immortalizing them with the Abelson virus. Cultures later derived from mice and maintained on stromal feeder layers did express detectable levels

of NF-κB and undergo germline Igκ gene transcription in the absence of LPS (Chen et al., 1994; Klug et al., 1994). There are NF-κB sites in all of the known enhancers, iEκ, E3' and Ed (Liu and Garrard, 2005; Liu et al., 2002; Sen and Baltimore, 1986b). The NF-κB site in iEκ is occupied in both pro-B and pre-B cells (Shaffer et al., 1997). This suggests that while NF-κB binding may be required for activation of this enhancer, other factors are important for the pro-to-pre-B cell transition.

Targeted disruptions have been made of each of the individual subunits of NF-κB. Although some of the phenotypes displayed are quite severe, even lethal in the case of p65 (RelA), the roles these factors play in B cell development appears to be redundant (Beg et al., 1995; Burkly et al., 1995; Doi et al., 1997; Köntgen et al., 1995; Sha et al., 1995; Weih et al., 1995). Pre-B cells express either hetero- or homodimers of p50, p52, p65 and c-Rel (Baldwin, 1996). p65 and c-Rel are the activating members of complexes that are induced during pre-B cell differentiation (Liou et al., 1994; Miyamoto et al., 1994). Because p65 was the originally identified factor with a role in Igκ expression and the knockout was embryonically lethal, fetal liver cells from these mice were used to reconstitute either lethally irradiated or SCID donor mice (Doi et al., 1997; Horwitz et al., 1997). In these mice, all of the hematopoetic lineages were reconstituted, however fetal liver adoptive transfer experiments from p65-/-, p50-/- mice resulted in a complete block in lymphopoesis (Horwitz et al., 1997). When the roles of p65 and c-Rel were examined more closely in B cells, it was found that Igκ germline transcription was not affected (Köntgen et al., 1995; Sha et al., 1995; Xu et al., 1996). However, a cell line expressing a *trans*-dominant form of the inhibitor IκBα (IκBΔN) is unable to import either c-Rel or p65 into the nucleus. In these cells with a dual

block in the activation pathway there is also a block in the activation of Igκ germline transcription and V-J joining (Scherer et al., 1996). A further phenotype of the IκBΔN cell line is a loss of inducibility by LPS (O'Brien et al., 1997). These cells are still responsive to interferon-γ activation, suggesting that there are at least two independent pathways for activation of germline transcription and V-J joining in pre-B cells (O'Brien et al., 1997).

Early B Cell Factor. Early B cell factor (Ebf) is novel factor containing a zinc-binding motif required for DNA binding and a bHLH-like domain required for homodimerization (Hagman et al., 1995). This protein is expressed in all stages of B cell development, but is not found in plasma cells (Hagman et al., 1993). Mice with a homozygous knockout of Ebf are viable, but the B cells arrest in development prior to any recombination events (Lin and Grosschedl, 1995).

PU.1 and Pip. PU.1 is a member of the large Ets family of transcription factors. These proteins all contain a conserved 85 amino acid motif responsible for binding to purine rich sequences. In general these factors require interactions with other factors to activate and repress their target genes, reviewed in (Crepieux et al., 1994; Leiden, 1993). PU.1, specifically is expressed in monocytes and granulocytes in addition to lymphoid cells (Klemsz et al., 1990). It is critical to the development of B cells. Homozygous knockout mice for PU.1 are embryonically lethal and display a lack of lymphocyte and myeloid lineages in their fetal livers (McKercher et al., 1996; Scott et al., 1994). Ets binding sites are located in a number of genes involved in B cell development. It is known that there are Ets

sites present in the E3' of Igκ (Pongubala et al., 1992). For PU.1 to activate transcription from this site, it requires the association of a second factor, Pip/NF-EM5, a member of the interferon regulatory factor family (Eisenbeis et al., 1995; Pongubala et al., 1992).

Regulation of Igκ by chromatin structure changes

Accessibility Hypothesis. The accessibility hypothesis posits that each V(D)J recombination event is regulated at the level of the chromatin accessibility of each coding gene segment. If this hypothesis is true, then mechanisms must exist to regulate the chromatin structure of these coding gene segments at the appropriate time in development in the correct cell type. This idea was first suggested by Frederick Alt and colleagues in 1986 (Yancopoulos and Alt, 1985; Yancopoulos and Alt, 1986) with experimental support first evident in 1996 (Stanhope-Baker et al., 1996). Recent evidence has also shown that DNA sequences with chromatin assembled *in vitro* are more resistant to RAG-1 and –2 cleavage than naked DNA (Golding et al., 1999; Kwon et al., 1998). Therefore, chromatin remodeling complexes and post-translational modifications of histones have been the focus of much recent work. These events are known to generate accessiblity in chromatin and have been areas of active research for many genes, including the Ig and TCR loci.

Other support for the accessiblity hypothesis is provided by the observation that transiently transfected recombination substrates can be induced to undergo rearrangement after expression of RAG-1 and –2 in non-B cells (Oettinger et al., 1990; Schatz et al., 1989), but the endogenous Ig genes remain intact (Romanow et al., 2000). This accessibilty can be

modulated by addition of other factors. Co-transfection of either E2A or EBF proteins in addition to RAG-1 and –2 leads to rearrangement of the endogenous loci (Goebel et al., 2001; Romanow et al., 2000). It has also been observed that mouse Vκ and Jκ regions are more generally DNase I sensitive at the pro/pre-B cell transition (Maes et al., 2001). The RSS sequences themselves are also targets for the recombinase cleavage both *in vivo* and *in vitro* in isolated nuclei only at the pre-B cell stage of development (Baumann et al., 2003a; Constantinescu and Schlissel, 1997; Goldmit et al., 2002; Stanhope-Baker et al., 1996).

Histone Acetylation. The acetylation of the histones surrounding Ig and TCR genes has been of much interest. Differential acetylation patterns provide a framework in which to analyze patterns of chromatin "opening" and accessibility. It has been demonstrated that regions of these genes active for recombination or transcription are hyperacetylated surrounding the promoters and enhancers (Agata et al., 2001; Huang et al., 2001; Johnson et al., 2003; Mathieu et al., 2000; McMurry et al., 1997; McMurry and Krangel, 2000). In the IgH locus, both the MARs which flank the enhancer and the LCR have been shown to have increased levels of acetylated histones associated with them when active (Fernandez et al., 2001; Madisen et al., 1998). In the case of the enhancer, occupancy of sites is the same on wild-type or MAR-deleted alleles. However, chromatin immunoprecipitation (ChIP) experiments showed that the MAR is required for acetylation of nucleosomes throughout the locus (Fernandez et al., 2001). In the case of the IgH LCR, the affected nucleosomes were also located at a distance from the element (Madisen et al., 1998). Treatment of cultured cells with histone deacetylase (HDAC) inhibitors leads to activation of rearrangement of the

appropriate antigen receptor locus (Cherry and Baltimore, 1999; Huang et al., 2001; Mathieu et al., 2000; McBlane and Boyes, 2000). In addition, model *in vitro* studies have shown that histone acetylation contributes to RSS accessibility (Kwon et al., 2000). These results could be attributable to indirect effects, but are highly suggestive that hyperacetylation contributes to the formation of accessible chromatin.

DNA Methylation. The methylation status of CpG dinucleotides surrounding recombining DNA segments has also been observed to affect the accessibility of the DNA to the recombinase. *In vitro* studies have shown that methylated substrates are resistant to recombination in transfected cells (Cherry and Baltimore, 1999; Hsieh and Lieber, 1992). For the Igκ locus, it has been observed that V-J joining only occurs on alleles whose Jκ region is demethylated (Mostoslavsky et al., 1998). However, both demethylation and histone hyperacetylation appear to be required for V-J joining to occur (Ji et al., 2003).

Scope of Present Study

Immunoglobulin genes present a unique system in which to study tissue-specific gene expression and site-specific DNA recombination as well as the relationship between chromatin structure and gene expression. However, although much has been learned concerning the regulatory elements in the downstream region of the locus, the remaining >3 Mb remains largely unstudied. This includes the DNA sequences, patterns of transcription factor binding sites and transcriptional activation. To determine fully whether the

accessibility hypothesis is indeed correct, a systematic study of this region must be undertaken. In this context, the goal of my research was to understand the regulation of the accessibility of the V regions of the Igκ locus. To carry out this aim, I conducted the following experimental approaches.

1. I assembled the complete nucleotide sequence of the Igκ locus. This included defining the boundaries on either end and mapping all potential Vκ genes, pseudogenes and relics onto it.
2. I determined the patterns of transcription factor binding sites surrounding Vκ genes in the locus. These were analyzed by both family and position in the locus.
3. I analyzed the chromatin structure surrounding the Vκ genes throughout B cell development to determine what specific changes occur prior to rearrangement. This included tests for germline transcripts from particular Vκ gene promoters.

CHAPTER TWO

Nucleotide Sequence of Igκ Locus and Transcription Factor Site Analysis of the Locus

Summary

The mechanisms regulating V gene usage leading to the immunoglobulin (Ig) repertoire have been of interest for many years but are only partially defined. To gain insight into these processes, I have assembled the nucleotide sequence of the *Mus musculus* Igκ locus using data recently made available from genome-wide sequencing efforts. I found the locus to be 3.21 Mb in length and mapped all known functional, pseudo- and relic V gene segments onto the sequence, along with known regulatory elements. I corrected errors in former gene assignments, positions and orientations and identified a novel Vκ4 gene segment. The 5' boundary of the locus is defined by the presence of the tumor-associated calcium-signal transducer-2 gene located 19 kb upstream of Vκ24-140, the most distal V gene. No non-Vκ genes were found in the sequence of the locus. Detailed analysis of the sequences 0.5 kb upstream, within, and 0.5 kb downstream of each potentially functional V gene revealed interesting patterns of statistically significant clustering of transcription factor consensus binding sites, generally specific to a particular family. I found E boxes were clustered not only in promoter regions, but also nearby recombination signal sequences. Family members of Vκ4/5 genes exhibit a conserved pattern of octamer sites in their downstream regions, as well as Ebf sites in their introns, and Lef-1 sites in their upstream regions.

Introduction

A diverse antibody repertoire is crucial to an organism‘s adaptive immune response. Primary combinatorial and junctional diversity in the repertoire develops through the creation of immunoglobulin (Ig) molecules by V(D)J gene rearrangements of the IgH and IgL gene loci (reviewed in (Sleckman et al., 1996)). Positive and negative selection of B lymphocytes, receptor editing, somatic mutation and the level of transcription of individually rearranged genes provide further variation to the distribution of different Ig molecules in the periphery (Casellas et al., 2002; Casellas et al., 2001).

To gain insight into the mechanisms of how the Ig repertoire might be established, it is essential to consider the genomic organizations of the IgH and IgL loci. In this work, I focus on Igκ, the IgL locus that encodes ~95% of the light chain molecules of the mouse. Previous studies of the mouse Igκ locus have established a nearly complete physical map based on yeast and bacterial artificial chromosome (YAC and BAC) contigs, revealing it to be the largest multi-gene-family locus thus far identified with respect to genomic length; Igκ spans an estimated 3.5 megabases (Mb) (George et al., 1995; Kirschbaum et al., 1996; Kirschbaum et al., 1998; Kirschbaum et al., 1999; Röschenthaler et al., 2000; Schable et al., 1999; Thiebe et al., 1999; Zachau, 2004). Early sequencing efforts of the Jκ-Cκ region established the presence of four functional and one non-functional Jκ regions, and a single Cκ exon (Max et al., 1981). More recently, sequencing of segments within upstream genomic clones has identified 96 potentially functional Vκ genes that have been grouped into 18 families based on sequence homologies, as well as numerous Vκ pseudogenes and relics, all of

which have been physically mapped onto the contig (Thiebe et al., 1999). Members of a given Vκ gene family were found to be semi-clustered but partially interspersed with members of other Vκ gene families (George et al., 1995; Thiebe et al., 1999). The most 5' Vκ gene, hf24 (renamed Vκ24-140 in this study), is a member of the Vκ24 family and located some 3.3 Mb away from the Jκ-Cκ region (Thiebe et al., 1999). The most 3' Vκ gene is Vκ21G (George et al., 1997), here termed Vκ21-1, located 18 kb away from the Jκ1 gene segment (Kirschbaum et al., 1998; Liu et al., 2002). The 3' boundary of the Igκ locus is adjacent to the gene encoding ribose-5-phosphate isomerase-A (Rpia) (Apel et al., 1995), but the gene that defines its 5' boundary remained to be identified. In addition, due to the previous lack of a more detailed systematic analysis, it is unclear whether any non-Vκ genes reside within this enormous locus. Data recently made available from genome wide sequencing efforts have allowed me to address these issues, as well as to map all Vκ genes onto this sequence.

It is unclear how a particular Vκ gene is selected for recombination to a particular Jκ region. This choice does not appear to be derived from chromosomal position (Li and Garrard, 2003), but instead several V genes throughout the locus are preferred in early rearrangements, both in mice and cell lines (Kalled and Brodeur, 1990; Kaushik et al., 1989; Ramsden et al., 1994b). In fact, the Vκ21 gene family, located most proximally to the Jκ-Cκ region of the locus (Kirschbaum et al., 1998), is rarely used in the earliest detectable rearrangement events in fetal liver (Kaushik et al., 1989; Ramsden et al., 1994b).

It has been proposed that rearrangement is regulated by appropriately modulating chromatin accessibility to the recombination machinery (Yancopoulos and Alt, 1986). Because transcription factors are known to recruit proteins that

mediate chromatin accessibility (for reviews, see (Felsenfeld et al., 1996; Felsenfeld and Groudine, 2003; Hampsey and Reinberg, 2003)), it seems likely that *cis*-acting DNA elements responsible for generating chromatin accessibility to Vκ genes reside within each Vκ gene's local vicinity, rather than in the well-studied Jκ-Cκ region at the 3' end of the locus. Indeed, V gene germline transcription, histone acetylation, and histone methylation have been associated with V(D)J joining (Casellas et al., 2002; McMurry and Krangel, 2000; Ng et al., 2003; Schlissel and Baltimore, 1989; Schlissel et al., 1991b). Unfortunately, previous studies characterizing potential *cis*-acting elements in Vκ genes have been limited to the available short upstream sequences in a subset of Vκ genes' promoters and to very short downstream sequences terminating near recombination signal sequences (RSS). Yet, despite this limit, these studies have identified octamer motifs in essentially all Vκ genes' promoters (Bemark et al., 1998; Thiebe et al., 1999), as well as E boxes in a subset of these promoters (Bemark et al., 1998). In view of the importance of several specific transcription factors in specifying B cell development and Ig gene expression (Reya and Grosschedl, 1998; Schebesta et al., 2002), I have taken advantage of my assembly of the Igκ gene locus sequence to map consensus binding sites for these factors onto each V gene sequence and its corresponding upstream and downstream flanks. This approach has uncovered several regions in which potential binding sites for transcription factors cluster within and around Vκ genes in a family-specific fashion. The clusters were found either in the upstream region, the intron, or downstream of the RSS. The most obvious novel finding is an enrichment of E-boxes or octamer motifs close to the 3' ends of many Vκ gene RSSs. These

findings provide new insight on potential molecular mechanisms for targeting Vκ genes for recombination.

Materials and Methods

Assembly of the Igκ gene sequence. The sequence was determined first by extracting the 500 kb segments of DNA (scaffolds) spanning the region from 66 Mb to 72 Mb from the Celera Discovery System (IDs: GA_x54KRFPKN04:9000001..9500000 through GA_x54KRFPKN04:12500001..13000000) and then aligning them as a contiguous sequence in Vector NTI (Informax, Bethesda, MD). The Celera database is a composite sequence from the C57Bl/6, A/J, 129 and DBA/2 strains of mice. To fill in the gaps, I utilized Build 30 of the C57Bl/6 mouse genome available from NCBI (http://www.ncbi.nlm.nih.gov/genome/seq/NCBIContigInfo.html). Regions of approximately 1 kb on either side of a gap were used to BLAST Build 30 (http://www.ncbi.nlm.nih.gov/genome/seq/MmBlast.html). Hits that were within the correct chromosomal region and at least 99% identical were considered to close gaps. These limits were chosen because they exceed Celera's criteria for matching shotgun fragments. Each hit had to match on both ends and also possess 99% or better identity for up to 3 kb on either side of the gap to fill it in. 47 of the gaps were filled completely in this manner (Table 2.1 and results). The flanks of the remaining gaps were then used to BLAST the high quality HTGS sequences that had not been mapped onto a chromosome. After I assembled this sequence, I mapped all of the known variable gene segments, the joining and constant regions, pseudogenes, relics, boundary genes and enhancers onto it (See results for details). For a known feature, a >99% identity with the sequence previously determined (http://biochemie.web.med.unimuenchen.de/zachau/kappa.htm) was

required to determine a match. Table 2.1 contains details on gap regions and what was used to fill them in.

Searching sequence for putative transcription factor binding sites. I wrote scripts in Perl, with the assistance of Monty Brekke, to search the Igκ locus and individual V genes for the DNA sequences of interest. Exact sequences searched for are listed in Table 2.3, along with references that provided consensus sequences. These scripts counted the motifs in each V gene as well as generated scale diagrams of their placement along the sequence.

Analysis for additional genes. Sequence was analyzed using **GENSCAN** (http://genes.mit.edu/GENSCAN.html) developed by Chris Burge (Burge and Karlin, 1997) using the parameter matrix, HumanIso.smat. The returned results included predicted peptides, promoter regions, exons, and polyA$^+$ sites, with the probability of each exon listed. I discarded potential gene sequences with no exon whose probability was greater than 50%. Predicted peptides from the algorithm were blasted against the SWISSPROT database at NCBI (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi).

Analysis of sequence for repeated sequences. Sequence was submitted to the RepeatMasker server (**repeatmasker@ftp.genome.washington.edu**), developed by A. Smit and P. Green (unpublished data). Analysis was against a rodent specific repeat database at the most sensitive setting. Analysis was against cross_match version 0.990329 and RepBase Update 6.3, vs 05152001. Sequence was considered repetitive if it matched the repeat sequence with >75% identity.

Statistical Analysis. Transcription factor binding motifs were analyzed using a cumulative binomial probability distribution: $\sum_{i=r}^{n} \frac{n!}{(n-r)!r!} p^{i}(1-p)^{n-i}$

where: n represents the number of V regions examined, r represents the number of V regions with motif, and p represents the probability of observing ≥1 of the motif in question based on the nucleotide composition of the locus, i.e. $p(Ebox) = 1 - (\{1 - [(pC)(pA)(pC + pG)^2(pT)(pG)]\}^{295})$, for a sequence length of 300 bp (pC=pG=0.2, pT=pA=0.3 for Igκ). This gives the probability that at least one of the motifs in question is present in the sequence of interest. This method of analysis was chosen as the best method after consultation with Dr. William Frawley in our Academic Computing Services Department and Dr. Nick Grishin in the Biochemistry Department. Specifics of the motifs analyzed and lengths for each region analyzed are listed in Table 2.3. This analysis is a simplified view of the significance. It does not take into account the fact that the locus most likely arose from gene duplication. It also does not take into account the conservation of position of elements. The significance is likely understated by the fact that if multiple instances of the motif exist, this is not taken into account and the positioning of the elements is not considered. It also would overstate the significance because of absence of consideration of the gene duplication of V genes.

Results

Assembly of the mouse Igκ gene locus sequence. To piece together the sequence of the Igκ locus, I first imported the scaffolds assembled by the Celera Discovery System comprising the sequence of the locus into VectorNTI. Celera had assembled the scaffolds from their sequence made up primarily of BACs. They linked the fragments using their own "mate-pair" method which allows determination of order and orientation of the sequenced DNA fragments as well as the length of the gaps remaining in the sequence with a high degree of confidence (Myers et al., 2000; Venter et al., 2001). After importing over 4 Mb of scaffold sequence, there were 148 gaps of unknown sequence in the locus. The gaps ranged in size from just a few bp to an estimated 30 kb. The first step to fill these gaps was to use the sequence from the analogous region in the NCBI database. 47 of these gaps were filled in completely using this information, while two gaps were only partially filled in this manner. The sequence flanking the 101 remaining gaps lacked enough identity overlap or was apparently located in regions of unknown sequence in the NCBI contig to be filled in this way. The flanking regions of these gaps were BLASTed as described in the Methods. One of the gaps only partially filled with the contig was completed with HTGS sequence, and an additional 47 gaps were filled in this way. One gap was partially filled-in with the HTGS sequence, leaving a much smaller region of undetermined sequence. Approximately 2% of the total sequence of the locus, distributed in 53 regions remains undetermined at this point (Table 2.1).

Position of 5' end of gap	Gap filled by NCBI contig?	Gap filled or corrected with? (N/A if still a gap)	Approximate length of gap (in bp), if still remaining	Type of repeat sequence gap is in
5749	Y			
8116	N	N/A	19	LTR
43128	N	RP23-123I10 replaced flanks; RP24-348E11 for center		
57260	N	N/A	60	LINE
70597	N	RP23-123I10		
80268	N	RP23-123I10		
91145	N	RP23-337A17		
140666	N	NT_050069.1 (unmapped contig)		
148908	N	N/A	2000	N/A
160845	Y			
277289	N	NT_039363.1 (mapped to chromosome 6, not positioned)		
325327	N	NT_039363.1 (same as above)		
326368	N	NT_039363.1 (same as above)		
375164	N	N/A	20	LTR
453984	Y			
456404	Y			
467595	Y			
547186	Y			
582624	Y			
591800	Y			
596282	Y			
601413	Y			
604164	Y			
605459	N	N/A	400	LTR
607459	N	N/A	620	LINE
614095	Y			
632354	Y			
635291	Y			
642970	Y			
658136	Y			
687449	Y			
696634	Y			
705845	Y (648 bp duplicate)			
708692	Y			
745864	Y			
748035	Y			
777999	Y			
791474	Y			
812849	N	RP23-29F15		
819097	Y			
820484	N	N/A	1300	LINE
828769	Y			
863557	N	RP23-29F15		
865702	N	RP23-29F15		
883680	Y (CT repeat, unclear how long)	RP23-29F15 (determined length of repeat)		
888557	N	N/A	140	LINE
891769	N	N/A	20	N/A

897797	Y			
950757	Y			
1004805	N	RP23-29F15		
1027055	Y			
1072085	Y (5' end)	RP24-405O23 (remainder)		
1073865	N	N/A	3500	LTR
1081226	N	N/A	20	N/A
1098991	N	N/A	20	LINE
1100666	N	RP24-405O23		
1101830	N	N/A	4000	LINE
1117957	N	RP23-405O23		
1126938	N	RP23-405O23		
1135310	N	RP23-405O23		
1141228	N	RP23-405O23		
1185653	Y			
1247641	Y			
1297054	Y			
1307441	Y			
1308551	N	RP23-211L5		
1395629	Y			
1397602	Y			
1410768	Y			
1435423	N	RP23-344N8		
1443618	N	RP23-344N8		
1476950	N	RP23-344N8		
1488714	N	N/A	3500	LINE
1539407	N	N/A	1600	LTR and LINE
1541734	N	N/A	1000	LTR and LINE
1623490	N	NT_052183.1 (unmapped contig)		
1660937	N	NT_052282.1 (unmapped contig)		
1663839	N	N/A	10	LINE
1674540	N	N/A	30	N/A
1700968	N	NT_062389.1 (unmapped contig)		
1704272	N	N/A	20	LINE
1719260	N	N/A	270	LINE and SINE
1721715	N	N/A	1500	LTR
1737387	N	N/A	2300	N/A
1752905	N	N/A	20	LINE/CT repeat
1798095	Y			
1874835	N	RP24-568F14		
1878614	N	RP24-568F14		
1897602	N	3' end fixed with RP24-568F14	20	LINE
1910776	N	N/A	20	LINE
1916191	N	RP24-568F14		
1940866	N	RP24-568F14		
1944504	N	RP24-568F14		
1982903	N	RP24-405H19		
1985095	N	N/A	20	LTR
1988200	N	N/A	2200	LTR
2042644	N	N/A	1900	LINE
2045847	N	N/A	2900	LINE
2075924	N	RP23-24I14		
2112782	N	RP23-24I14		
2138313	N	NT_042284.1 (unmapped		

		contig, fixed 5' end) RP23-24I14 (fixed 3' end)		
2156302	Y (21 bp overlap)			
2239874	N	N/A	1800	LINE
2242852	N	N/A	20	LINE
2245480	N	N/A	130	N/A
2248140	N	N/A	2700	LINE
2251693	N	N/A	1950	N/A
2288699	N	RP24-475M8		
2322643	Y			
2376076	N	RP24-475M8		
2378508	N	N/A	20	LINE
2411648	Y			
2435147	Y (28 bp overlap)			
2454915	Y			
2470080	N	N/A	350	LINE
2473695	Y			
2483416	N	RP24-314D8		
2499087	Y – 1.2 kb on 3' end	N/A	3400	LINE
2564256	N	N/A	20	N/A
2591679	N	N/A	500	LTR
2639490	N	N/A	5700	LINE
2660506	Y			
2683578	N	N/A	23000	LINE/SINE
2716989	N	N/A	300	LINE
2719113	N	N/A	2900	LINE
2728102	N	N/A	3500	LINE
2734999	N	N/A	1400	LTR
2738769	Y (4 bp overlap)			
2748026	N	N/A	500	LINE
2755758	N	NT_066791.1 (unmapped contig)		
2765548	N	NT_070146.1 (unmapped contig)		
2781362	N	N/A	200	LTR
2811546	N	NT_075060.1		
2821255	N	NT_075060.1		
2831032	N	RP24-392D13		
2865290	N	RP24-392D13		
2923682	N	N/A	20	LINE
2927153	N	RP24-392D13		
2927562	N	N/A	5700	LINE
2971240	N	RP24-392D13		
2987182	N	NT_058603.1 (unmapped contig)		
2992942	N	N/A	3800	LINE
2998699	N	N/A	2800	N/A
3065728	N	N/A	500	LINE
3079526	N	RP23-435I14, small gap remaining	30	LINE
3087715	N	NT_052827.1 (unmapped contig)		
3148914	Y			
3150544	N	AF513926		

Table 2.1 (previous pages) – Positions of gaps in the sequence and how they were filled in. The position of the 5' end of each gap is listed. If the gap was filled by using the analogous region of Chromosome 6 from NCBI, then there is a Y in the second column. An N in this column indicates that there was insufficient data from the genomic contig to unambiguously correct this region. If the sequence was corrected with the HTGS sequence data from NCBI, then the clone designation is listed in the next column. If multiple clones are listed, then each was used to correct the region designated. An N/A in this column indicates that there was still insufficient data to fill in the sequence of this region. The fourth column lists the approximate length in bp of the unknown sequence, based on the original length of the gap given by Celera, minus the length of the sequence used to fill in the gap. The final column lists the type of repetitive sequence in which the gap is located. N/A indicates the gap is not within an identified repeat region. If two types of repetitive sequences are listed, the gap is bounded by each of them.

Overall features of the mouse Igκ gene locus and its boundaries. The Igκ locus spans 3,207,797 bp on Chromosome 6, from position 67.8 to 71.2 Mb. The tumor-associated calcium-signal transducer 2 gene (Tacstd2) is located 19 kb upstream of the most 5' Vκ gene in the locus, and is transcribed in the opposite orientation with respect to Jκ-Cκ. Tacstd2 is a 1.7 kb intronless gene encoding a Type I membrane protein whose intracellular domain is involved calcium signaling pathways. Aberrant expression of the human ortholog M1S1 has been implicated in a number of human tumors, both lymphoid and non-lymphoid in origin. Most notably, its mutation can lead to gelatinous drop-like corneal dystrophy, an autosomal recessive disorder that eventually leads to blindness (Ren et al., 2002). This flanking gene is not conserved in the human genome with respect to position. In the 300 kb 5' to both the human and mouse Igκ genes there is no conservation of genes. In this 300 kb in the mouse, genes 5' to Tacstd2 are the interleukin-23 and interleukin-12b receptors. The previously identified 3' boundary gene, Rpia (Apel et al., 1995), is located immediately downstream of

the J-C region in both human and mouse. Interestingly, this 3' flanking gene is also transcribed in the opposite orientation as Jκ-Cκ.

The size of the Igκ locus makes it a challenging, yet interesting system in which to study gene regulation. However, to undertake these analyses, it is necessary to determine whether the Igκ gene locus contains any non-Igκ genes that might affect interpretation of results of any future studies. Therefore, I searched the sequence for predicted gene structures using the GENSCAN program. Despite the size of the locus, I failed to find any predicted genes. Several known processed pseudogenes and gene fragments were identified (Fig. 2.1), but it appears that Igκ is a single contiguous locus.

In an effort to fill in more of the remaining gaps as well as understand more about the genomic structure of the Igκ locus, I determined the distribution and positions of repeated DNA sequences within my assembly by submitting the sequence to the RepeatMasker server (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker) (Fig. 2.1, grey sections). The majority of the repeats, as expected, are LINEs, which constitute 41% of the sequence of the Igκ locus. Retroviral LTR elements comprise 11% of the sequence, and only 2.5% is made up of SINEs, B1 and B2 sequences. Individual elements can be found in the composite sequence submitted to GenBank. On average, a mouse autosome consists of approximately 15% LINE sequence, while the X chromosome is composed of 29% LINE sequence (Allen et al., 2003). Igκ therefore contains a much higher percentage of LINE sequence, and a correspondingly lower percentage of SINE sequence than does the genome as a whole. Analysis of the repeat regions also allowed me to conclude that approximately 90% of the unknown sequence is located in these repeat sequences (See Table 2.1).

Mapping V genes onto the mouse Igκ gene locus sequence. I mapped each of the known potentially functional V genes, pseudogenes and relics onto the assembled sequence. As a guide, I used the work of Zachau and colleagues (Thiebe et al., 1999) and an updated list at http://biochemie.web.med.uni-muenchen.de/zachau/kappa.htm. Each of these sequences was loaded into Vector NTI for mapping. The assembled sequence was then searched for the reported sequence of each known V gene. All of the V genes were found to greater than 99% identity in the composite sequence, but, as described below, there were some differences in the data published by Thiebe et al. (1999) (Fig. 2.1). Table 2.2 lists each of the Vκ genes in order of their location 5' to 3', following the nomenclature adopted for the human Ig loci (Lefranc, 2001a; Lefranc, 2001b), along with accession numbers for their sequence and 1 kb of both upstream and downstream flanking sequence. This nomenclature uses the logical system of naming a V gene with the family name followed by the gene's position numbered sequentially from the 3' to 5' end of the locus. This numbering is used for both functional and pseudogenes. Functional genes are defined as those with no apparent defects in their promoters, splice sites or RSSs. Genes that have reported cDNAs in the database are also classified as functional even if they have minor defects in their RSS. Pseudogenes are the genes with defects in the above regions that would prevent them being used in V-J joining. Relics have been identified through hybridization studies, but have significant divergence in normal gene structure. Because of the significant divergence of relics, they were named in a slightly different manner; the family name, followed by a letter designation (from a to r), also 3' to 5'. This new nomenclature is also related to that of the Zachau group, as well as other previously used names for the V genes in Table 2.2. Figure 2.1 shows the resulting physical map of this assembly.

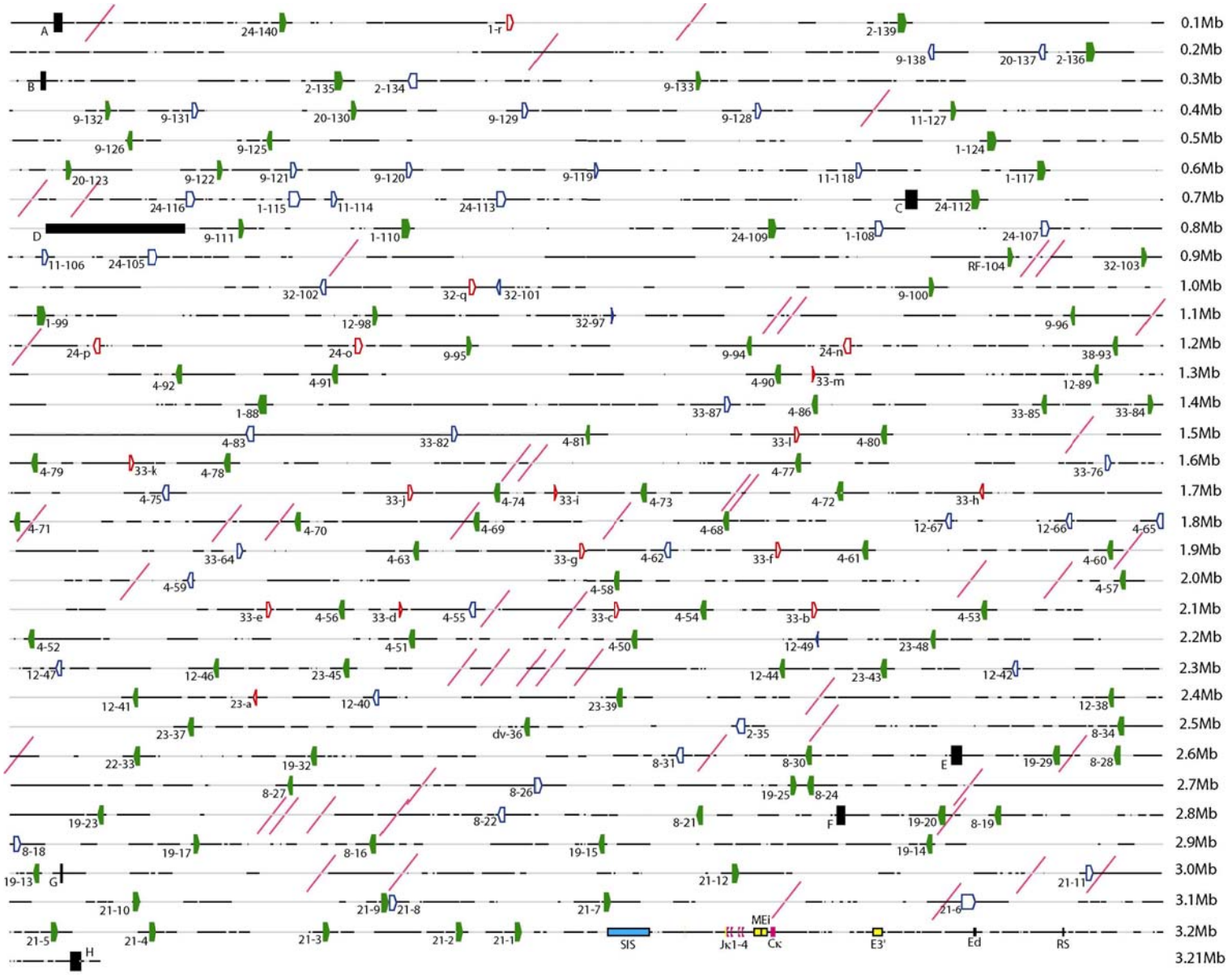
0.1Mb
0.2Mb
0.3Mb
0.4Mb
0.5Mb
0.6Mb
0.7Mb
0.8Mb
0.9Mb
1.0Mb
1.1Mb
1.2Mb
1.3Mb
1.4Mb
1.5Mb
1.6Mb
1.7Mb
1.8Mb
1.9Mb
2.0Mb
2.1Mb
2.2Mb
2.3Mb
2.4Mb
2.5Mb
2.6Mb
2.7Mb
2.8Mb
2.9Mb
3.0Mb
3.1Mb
3.2Mb
3.21Mb
A
B
C
D
E
F
G
H
24-140
1-r
2-139
9-138
20-137
2-136
2-135
2-134
9-133
9-132
9-131
20-130
9-129
9-128
11-127
9-126
9-125
1-124
20-123
9-122
9-121
9-120
9-119
11-118
1-117
24-116
1-115
11-114
24-113
24-112
9-111
1-110
24-109
1-108
24-107
11-106
24-105
RF-104
32-103
32-102
32-q
32-101
9-100
1-99
12-98
32-97
9-96
24-p
24-o
9-95
9-94
24-n
38-93
4-92
4-91
4-90
33-m
12-89
1-88
33-87
4-86
33-85
33-84
4-83
33-82
4-81
33-l
4-80
4-79
33-k
4-78
4-77
33-76
4-75
33-j
4-74
33-i
4-73
4-72
33-h
4-71
4-70
4-69
4-68
12-67
12-66
4-65
33-64
4-63
33-g
4-62
33-f
4-61
4-60
4-59
4-58
4-57
33-e
4-56
33-d
4-55
33-c
4-54
33-b
4-53
4-52
4-51
4-50
12-49
23-48
12-47
12-46
23-45
12-44
23-43
12-42
12-41
23-a
12-40
23-39
12-38
23-37
dv-36
2-35
8-34
22-33
19-32
8-31
8-30
19-29
8-28
8-27
8-26
19-25
8-24
19-23
8-22
8-21
19-20
8-19
8-18
19-17
8-16
19-15
19-14
19-13
21-12
21-11
21-10
21-9
21-8
21-7
21-6
MEI
21-5
21-4
21-3
21-2
21-1
SIS
Jκ1-4
Cκ
E3'
Ed
RS

Figure 2.1 (Previous page) – Schematic of the entire mouse Igκ locus, drawn to scale. For order and identification of the V genes, Table 2.2 indicates them from 5' to 3' with the accompanying GenBank accession numbers for the individual files, which have exons, TATA box, RSS and other binding sites labeled. Green arrows are potentially functional sequences, hollow blue arrows are presumed pseudogenes and hollow red arrows indicate relics. The direction of each arrow indicates the transcriptional orientation of each V gene. Black boxes represent gene fragments in the locus, as well as the boundary genes at either end. Also indicated in the far downstream J-C region are the positions of several *cis*-acting regulatory sequences (for details, see Liu et al. 2002) and the recombining sequence (RS) (Durdik et al. 1984). The adjacent MAR and Ei are labeled as MEi. Non-repetitive and repetitive DNA sequences are depicted as solid black and solid grey lines, respectively. Each pink diagonal line represents the approximate location of a gap. For reference in Table 2.2d, capital letters represent boundary genes and other gene fragments in the locus.

The sequence of 4-53 matched both ar4 and ko4 within one base. Because this sequence only occurs once in the locus, I conclude that they are the same V gene, 4-53. The orientation of three potentially functional and one pseudogene were found to be opposite of that originally published. The orientation of 9-125, 9-126 and 9-138, formerly cy9, cw9 and cx9 are in the reverse transcriptional orientation with respect to the J-C region. The 1-117 gene, previously cr1, was found to be in the forward orientation (Figure 2.1). These changes in orientation are on large contiguous segments of DNA. The previous assignments were made based on mapping of YACs and BACs, with much more unknown sequence. In all of the cases in which I found the orientation to differ from the published data, there are numerous adjacent V genes in the contiguous ungapped segment of DNA that fit the published orientations. This gives me complete confidence in my orientation assignments.

The placement, orientation and position of several genes identified after the original publications was also elucidated (Röschenthaler et al., 2000; Thiebe et

al., 1999). The more recently discovered genes bv9, bz9, and hk24, now called 9-122, 9-119, and 24-116 were all located in the 5' portion of the locus in the forward orientation (Figure 2.1).

To ensure that I had identified all potentially functional V genes, I searched the assembly for conserved heptamer and nonamer sites. I chose this method because this conserved feature would best indicate a V region (Gellert, 2002). While I found numerous sites of either a heptamer or nonamer, I identified only four new sites that contained both these consensus sequences separated by the requisite 12 bp spacer. Three of these sequences did not have any further similarity to V regions. The remaining sequence was determined by similarity matches to be a member of the Vκ4 gene family and has been named 4-57. The gene segment contains similar intron/exon structure to its other family members, including splice sites, exon lengths and promoter elements. This leaves the final count of V genes at 140, 95 of which are potentially functional. To ensure that the 5' boundary of Igκ was properly defined, I also searched 300 kb upstream of Vκ24-140 for potential V genes. This distance is 3 times greater than the largest distance between two V genes in the locus. No other potential V genes were found. I am confident that there are no other V genes contained in this locus.

Vκ segment	**Other names**	**Accession number**
24-140	hf24	AY591710
2-139	bd2, Vκ2(70/3)	AY591707
2-136	bj2, Vκ2(70/1)	AY591706
2-135	bi2	AY591705
9-133	cb9	AY591744
9-132	cj9, MOPC173C	AY591743
20-130	bw20	AY591704
11-127	if11, MSI-N17	AY591696
9-126	cw9, Igκ-V9a	AY591742
9-125	cy9, Igκ-V9b	AY591741
1-124	bl1, K18.1	AY591698

20-123	bt20	AY591703
9-122	bv9	AY591740
1-117	cr1, Vκ1C, K1A5	AY591697
24-112	hg24, $V\kappa_{167}$	AY591709
9-111	ba9, Vκ41	AY591739
1-110	bb1, Vκ1A	AY591695
24-109	he24, Vκ24B	AY591708
RF-104	RF	AY591748
32-103	gr32, 1-10	AY591711
9-100	cf9	AY591738
1-99	cv1	AY591702
12-98	ci12	AY591700
9-96	ce9, $V\kappa Id^{CR}$	AY591747
9-95	by9, 3386 5'V	AY591746
9-94	cp9, AJ2	AY591745
38-93	gj38c	AY591714
4-92	ay4	AY591737
4-91	kf4	AY591736
4-90	an4, R2	AY591735
12-89	fl12	AY591699
1-88	cs1	AY591701
4-86	kb4, X24, H6	AY591734
33-85	gn33, Igκ-34b	AY591713
33-84	gm33, Igκ-34c	AY591712
4-81	ag4, R11	AY591733
4-80	af4, 84con	AY591732
4-79	ae4, 4.68, 81con(1)	AY591731
4-78	ad4, H1	AY591730
4-77	aj4, H8	AY591729
4-74	ai4, 128con	AY591728
4-73	ah4, 72con	AY591727
4-72	am4, 70Z/3, 67con	AY591726
4-71	al4, 18con	
4-70	kn4, R9	AY591725
4-69	km4, H9	AY591724
4-68	aq4, H13	AY591723
4-63	ac4, 38con	AY591722
4-61	aa4, H4	AY591721
4-60	kk4, H3, pOx	AY591720
4-58	kj4, S107B, R1	AY591719
4-57	kp4	
4-56	ap4, 76con	AY591718
4-54	at4, np2.6.1	AY591717
4-53	ar4, ko4, 148con	AY591716
4-52	kh4, 4.58	AY591715
4-51	4-51, V-L8, L8	AY591668
4-50	4-50, κ(2154), R13	AY591667
23-48	23-48, L7	AY591639
12-46	12-46	AY591607

23-45	23-45, B1P8-7b-2	AY591638
12-44	12-44	AY591606
23-43	23-43	AY591637
12-41	12-41, VκRFT2	AY591604
23-39	23-39, 23.32	AY591636
12-38	12-38	AY591602
23-37	23-37	AY591635
dv-36	dv36	AY591694
8-34	8-34	AY591686
22-33	22-33, Vκ22G	AY591634
19-32	19-32, V-Ser	AY591620
8-30	8-30, H8 VL	AY591684
19-29	19-29	AY591619
8-28	8-28	AY591683
8-27	8-27	AY591682
19-25	19-25	AY591618
8-24	8-24	AY591680
19-23	19-23	AY591617
8-21	8-21	AY591678
19-20	19-20	AY591616
8-19	8-19	AY591677
19-17	19-17	AY591615
8-16	8-16	AY591675
19-15	19-15	AY591614
19-14	19-14	AY591613
19-13	19-13, V_{TNP}	AY591612
21-12	21-12, 1.5kb Vκ, Vκ21E	AY591624
21-10	21-10, 16.0kb Vκ, Vκ21B	AY591622
21-9	21-9, 4.0kb Vκ	AY591633
21-7	21-7, 1.6kb Vκ	AY591631
21-5	21-5, Vκ21C	AY591630
21-4	21-4, 8.5kb Vκ	AY591629
21-3	21-3, 18.5 Vκ	AY591628
21-2	21-2, Vκ21A	AY591626
21-1	21-1, Vκ21G	AY591625

Table 2.2a.

Pseudogene	**Other names**	**Accession number**
9-138	psi-cx9, Vκ9B (294A9)	AY591693
20-137	psi-bk20, Vκ20 (294A9)	AY591621
2-134	psi-bh2	AY591627
9-131	psi-cl9	AY591692
9-129	psi-co9	AY591691
9-128	psi-br9	AY591690
9-121	psi-bp9	AY591689
9-120	psi-bq9	AY591688
9-119	psi-bz9	AY591687
11-118	psi-ie11	AY591600
24-116	psi-hk24	

1-115	psi-cz1	AY591601
11-114	psi-ic11	AY591599
24-113	psi-ha24	AY591643
1-108	psi-cq1	
24-107	psi-hc24	AY591642
11-106	psi-ia11	AY591598
24-105	psi-hd24	AY591641
32-102	psi-gs32	AY591648
32-101	psi-gl32	AY591647
32-97	psi-gk32	AY591649
33-87	psi-gp33	AY591654
4-83	psi-kg4	AY591674
33-82	psi-gq33	AY591653
33-76	psi-ga33	AY591652
4-75	psi-ka4	AY591673
12-67	psi-fg12	AY591611
12-66	psi-fr12	AY591610
4-65	psi-ki4	AY591672
33-64	psi-gu33	AY591651
4-62	psi-ab4	AY591671
4-59	psi-kl4	AY591670
4-55	psi-ao4, 118con	AY591669
12-49	psi12-49	AY591609
12-47	psi12-47	AY591608
12-42	psi12-42	AY591605
12-40	psi12-40	AY591603
2-35	psi2-35	AY591640
8-31	psi8-31	AY591685
8-26	psi8-26	AY591681
8-22	psi8-22	AY591679
8-18	psi8-18	AY591676
21-11	psi21-11, 6.0kb Vκ	AY591623
21-8	psi21-8, 4.0kb Vκ	AY591632
21-6	psi21-6, 6.0kb Vκ	

Table 2.2b.

Relic	**Former name**	**Accession number**
1-r	r-bc1	
32-q	r-gt32	AY591650
24-p	r-hi24	AY591646
24-o	r-hb24	AY591645
24-n	r-hh24	AY591644
33-m	r-gf33	AY591666
33-l	r-ge33	AY591665
33-k	r-gb33	AY591664
33-j	r-gv33	AY591663
33-i	r-go33	AY591662

33-h	r-gd33	AY591661
33-g	r-gb33	AY591660
33-f	r-gh33	AY591659
33-e	r-gz33	AY591658
33-d	r-gg33	AY591657
33-c	r-gx33	AY591656
33-b	r-gy33	AY591655
23-a	r-fp23	

Table 2.2c.

Label	**Gene**
A	Tumor associated calcium signal transducer
B	alpha-tubulin related pseudogene
C	cortactinB related pseudogene
D	mouse mammary tumor virus-8
E	s-adenysylmethionine decarboxylase gene fragment
F	s-adenysylmethionine decarboxylase gene fragment
G	ornithine decarboxylase gene fragment
H	Ribose-5-phosphate isomerase exon 9

Table 2.2d.

Table 2.2 - A 5' to 3' listing of each of the significant elements in the Igκ locus. a) each of the potentially functional V gene segments displayed on Fig. 2.1, b) each of the presumed pseudogenes, c) each of the presumed relics within the boundaries of Igκ. These are followed by the GenBank Accession numbers of each of these gene segments with 1 kb on either side of the coding region. d) lists the other genes, pseudogenes and gene fragments located in and near Igκ.

Alignment of promoter regions reveals conserved transcription factor binding sites. Several intriguing studies demonstrate essential roles for specific transcription factors in regulating B cell development (reviewed in (Reya and Grosschedl, 1998; Schebesta et al., 2002)). It is particularly striking that the engineered co-expression of E2A or early B cell factor (Ebf) in conjunction with RAG1/2 can induce V(D)J rearrangement of IgL genes in non-lymphoid cells (Goebel et al., 2001; Romanow et al., 2000). Therefore, I wished to determine whether there are significantly conserved patterns of potential binding sites within or near V genes for E2A, Ebf, and other transcription factors involved in B cell

development and Ig gene expression. For this purpose, I aligned all of the sequences of the potentially functional Vκ genes, either upstream of their leader regions, within their coding and intron regions, or downstream of their RSSs. I searched these alignments for the positions of consensus binding sites for relevant transcription factors. These included: the Oct1/2 Octamer binding site (Falkner and Zachau, 1984; Wirth et al., 1987); an E2A preferred E box (Blackwell and Weintraub, 1990; Ephrussi et al., 1985); the Ikaros site (Georgopoulos et al., 1994; Georgopoulos et al., 1992); the Ebf site (Hagman et al., 1991; Lin and Grosschedl, 1995); the PU.1 site (Klemsz et al., 1990; Schwarzenback et al., 1995); the Pip/IRF4 site (Eisenbeis et al., 1995; Pongubala et al., 1992); the Stat-5 site (Liu et al., 1995); the Lef-1 HMG box (reviewed in (Schilham and Clevers, 1998); the NF-κB site (Sen and Baltimore, 1986a; Wirth and Baltimore, 1988); as well as a consensus sequence for forkhead proteins – which are involved in opening of chromatin in early development (Cirillo and Zaret, 1999). Table 2.3 shows the consensus sites used for searching the sequence, expected frequencies, and probabilities for the observed patterns occurring on a random basis. Although these probability calculations do not take into account that V genes were likely generated by gene duplication events and hence should share considerable sequence identity, analysis of sequences immediately flanking conserved binding sites reveals sequence divergence (see discussion).

I first aligned and analyzed the 500 bp region immediately upstream of each potentially functional Vκ gene's leader sequence. Sequences are grouped by V gene family, listed in the order that they occur in the locus 5' to 3' (Figure 2.2, top to bottom). The TATA box is in every V gene, and shares similar positioning within but not between each V gene family (Figure 2.2, open, black-bordered boxes). The known conservation of the octamer motif is also striking, with only

seven examples that do not contain a perfect consensus, but a variant octamer sequence (Figure 2.2, red boxes and open red-bordered boxes, respectively). When the pseudogenes and relics are examined, the most obvious initial observation is the mutation or altered position of the octamer site (Figure 2.3) This known conservation led to the calculation of a probability of observing this pattern by random chance to be approaching zero.

The next most noticeable conservation in V genes' upstream regions is the presence of an E2A preferred E box, either 5' or 3' of the octamer site in many of the V gene families (Figure 2.2, blue boxes). Several of these E boxes have been identified previously (Bemark et al., 1998), but I have detailed the conservation more completely here. With the exception of the Vκ2, Vκ1, VκRF, Vκ4 and Vκ23 gene families, each V gene family contains a conserved E box within 30 bases of the octamer sequence. Interestingly, most members of the Vκ21 and Vκ19 gene families contain E boxes flanking both sides of their octamer site (Figure 2.2, bottom), while the more 5' gene members of the Vκ19 family only have a single E box site, as do the Vκ22 and Vκ8 gene family members. However, the most favored position of E box sites in the V genes residing in the 5' region of the Igκ locus is upstream of the octamer, irrespective of their transcriptional orientation in the locus. The probability of observing at least one E box in these V regions is again, a very small number. E boxes are either completely absent or far removed with regard to the position of the octamer site in the pseudogenes residing within the 3' region of the locus (Figure 2.3).

Other interesting patterns emerge when one examines the locations of potential Lef-1 HMG boxes. Most striking is the Vκ4 gene family, where Lef-1 sites are located immediately downstream of the TATA box of 19/27 family

members (Figure 2.2, center, orange boxes). The Vκ9 and Vκ12 gene families also exhibit this site with a high level of conservation, at –250 to –350 bp from the transcription start site (Figure 2.2, upper half, orange boxes).

There are also a large number of both Ikaros and winged helix/forkhead consensus sites upstream of most Vκ genes. Only 8/95 do not exhibit at least one Ikaros site and only 2/95 do not have potential binding sites for either (Figure 2.2, pink and green boxes, respectively). Interestingly, these sites are present in a statistically significant greater frequency over the expected random values for their occurrence both near the V genes as well as in the rest of the locus (Table 2.3; data not shown). It is also of interest to note that the probabilities in Table 2.3 list the probabilities based on at least one of each of these factors, whereas there are actually multiple copies of each of these sites at each Vκ gene. In effect, the Igκ locus is enriched in potential binding sites for both of these factors.

Sequence searches for potential binding sites for Ebf (not found), NF-κB (dark red boxes), PU.1 (black boxes), Pip (data not shown, statistically insignificant), and Stat-5 (data not shown, statistically insignificant) did not yield any results of obvious interest. Nevertheless, it can be concluded that potential binding sites for Oct1/2, E2A, Lef-1, Ikaros and winged helix/forkhead proteins often exhibit highly significant non-random positioning in the upstream regions of many Vκ genes (Table 2.3).

Transcription Factor	Consensus Sequence	Expected Frequency (in bp, approx.)	V regions with pattern / total V regions	Probability of observed patterns of occurrence	Reference(s)	Color in Figs. 2-4
Oct1/2	ATGCAAAK	1/10,000	88/95 20/27	p(promoter)≈$4.8x10^{-143}$ p($V\kappa4_{DS}$)≈$7x10^{-26}$	(Falkner and Zachau, 1984; Wirth et al., 1987)	Red
E2A and AP-4	CASSTG	1/1,700	62/95 5/95 51/68	p(promoter)≈$4.6x10^{-37}$ p(intron)≈0.15 p(DS(excl Vκ4))≈$2.4x10^{-27}$	(Aranburu et al., 2001; Blackwell and Weintraub, 1990; Ephrussi et al., 1985)	Light blue
Ikaros	GGGAW	1/350	79/95 62/95 74/95	p(promoter)≈$2.7x10^{-5}$ p(intron)≈$2.2x10^{-23}$ p(DS)≈0.03	(Georgopoulos et al., 1994; Georgopoulos et al., 1992)	Pink
Ebf	TWCCCNNGGGAWT	1/1,150,000	15/27	p(intron,Vκ4)≈$1.9x10^{-57}$	(Hagman et al., 1991; Lin and Grosschedl, 1995)	Yellow
PU.1	TGRRGAAGT	1/31,000	5/10	p(Vκ9 promoter)≈$2.3x10^{-9}$	(Klemsz et al., 1990; Schwarzenback et al., 1995)	Black
Pip/IRF-4	GAASTGAA	1/13,000	8/95	p(V region)≈0.14	(Eisenbeis et al., 1995; Pongubala et al., 1992)	N/A
Stat-5	TTCYNRGAA	1/12,000	9/95	p(V region)≈0.15	(Liu et al., 1995)	N/A
Lef-1	WWCAAAG	1/1,300	53/95 13/18 22/95	p(promoter)≈$1.7x10^{-6}$ p(intron for Vκs8,19,22)≈$1.8x10^{-13}$ p(DS)≈0.11	(Schilham and Clevers, 1998)	Orange
NF-κB	GGGRNNYYCC	1/12,500	4/5 1/95	p(Vκ1 intron)≈$1.4x10^{-8}$ p(promoter)≈0.09	(Sen and Baltimore, 1986b; Wirth and Baltimore, 1988)	Maroon
Forkhead	RWAAAYAW	1/700	69/95 55/77 74/95	p(promoter)≈$1.2x10^{-13}$ p(intron excl Vκs8,19,22)≈$5.9x10^{-39}$ p(DS)≈$2.1x10^{-17}$	(Cirillo and Zaret, 1999)	Green
Pax-5	CANTGNNGCGKRACSR	1/5,800,000	0/95	p(V region)≈0.98	(Barberis et al., 1990)	N/A

Table 2.3

Table 2.3 (Previous page) – Consensus sequences used for searching the Igκ sequence for transcription factor motifs. Following the consensus sequence is the expected frequency of random occurrence of each motif based on the nucleotide composition of the locus. The next column lists the number of V genes possessing the motif of interest versus the number of V genes analyzed. The final column lists the probability for the number of V genes observed to contain at least one copy of the corresponding motif based on its random occurrence, calculated as described in the Materials and Methods section. For Ikaros and Forkhead promoter calculations, the sequences used were the 300 bp located from –500 to –200 bp upstream of the first exon. For all other calculations, promoter sequences corresponded to the 500 bp upstream regions. Intron length corresponds to the average intron length for the sequences considered. For downstream (DS) calculations, the length corresponded to the 300 bp region located immediately downstream of the RSS. For the probabilities of Pax-5, Stat-5, and Pip, the entire V region and its flanking sequence were used since there were so few of these sites present. For calculations where not all V regions were included, those families excluded are indicated by "excl." R=A or G; K=G or T; S=C or G; W=A or T; Y=C or T.

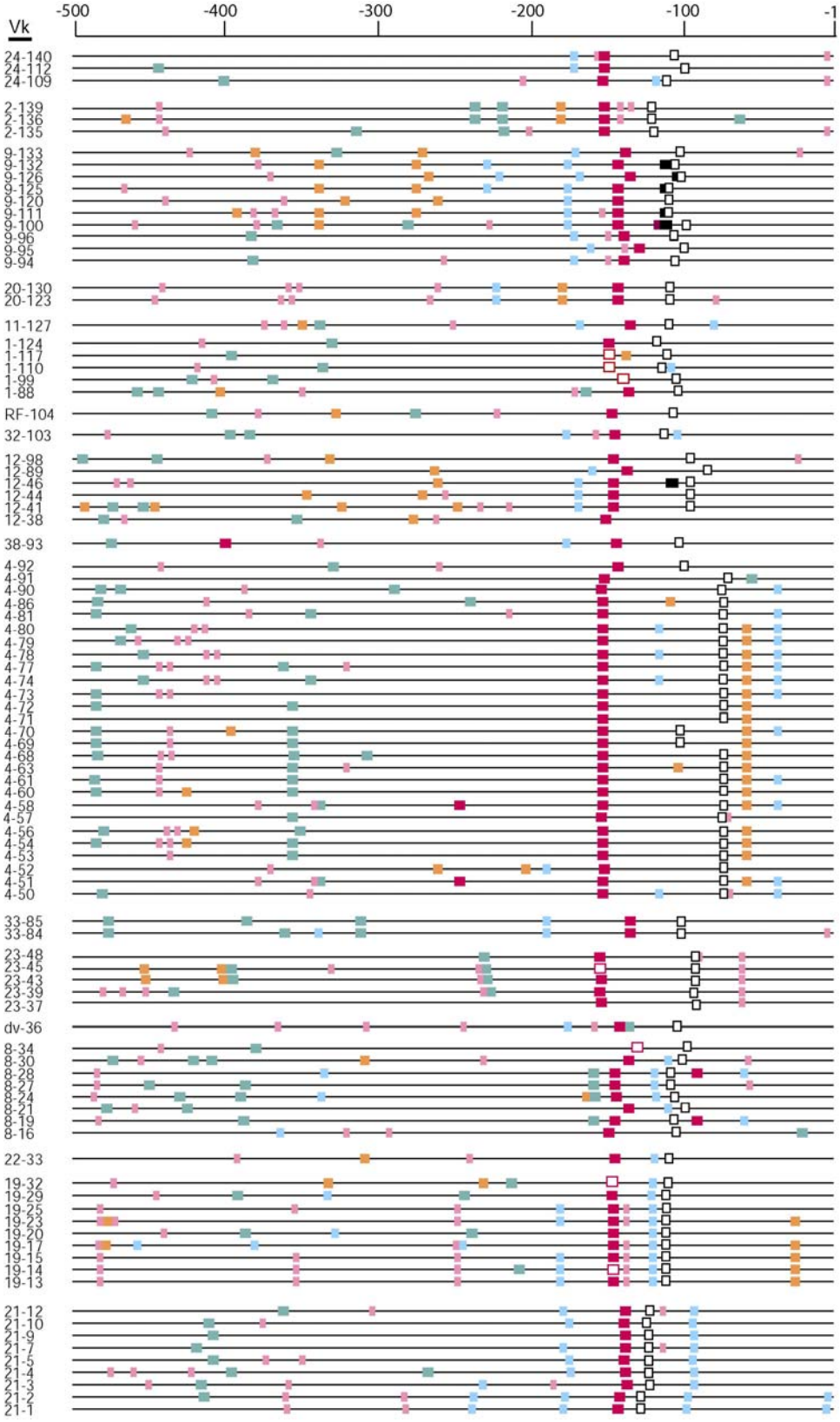
-500
-400
-300
-200
-100
-1
Vk
24-140
24-112
24-109
2-139
2-136
2-135
9-133
9-132
9-126
9-125
9-120
9-111
9-100
9-96
9-95
9-94
20-130
20-123
11-127
1-124
1-117
1-110
1-99
1-88
RF-104
32-103
12-98
12-89
12-46
12-44
12-41
12-38
38-93
4-92
4-91
4-90
4-86
4-81
4-80
4-79
4-78
4-77
4-74
4-73
4-72
4-71
4-70
4-69
4-68
4-63
4-61
4-60
4-58
4-57
4-56
4-54
4-53
4-52
4-51
4-50
33-85
33-84
23-48
23-45
23-43
23-39
23-37
dv-36
8-34
8-30
8-28
8-27
8-24
8-21
8-19
8-16
22-33
19-32
19-29
19-25
19-23
19-20
19-17
19-15
19-14
19-13
21-12
21-10
21-9
21-7
21-5
21-4
21-3
21-2
21-1

Figure 2.2 (Previous page) – Putative transcription factor binding sites within the 500 bp upstream regions of each potentially functional Vκ gene. Sequences are grouped by V gene family, listed in the order that they occur in the locus, 5' to 3', from top to bottom. The right side of the alignment corresponds to the 3' ends of the sequences, starts at -1 bp from the ATGs in the leader exons and extends upstream to the left for -500 bp. It should be noted that the 5' untranslated regions of Igκ mRNAs tend to be only a few nucleotides long (Kelley et al., 1982). Color coded boxes represent the following elements and potential binding sites for transcription factors: Open black-bordered, TATA box; red or open red-bordered, octamer and variant octamer, respectively; blue, E box; orange, Lef-1; pink, Ikaros; green, winged helix/forkhead; and black, PU.1. Ebf and NF-κB sites were not found.

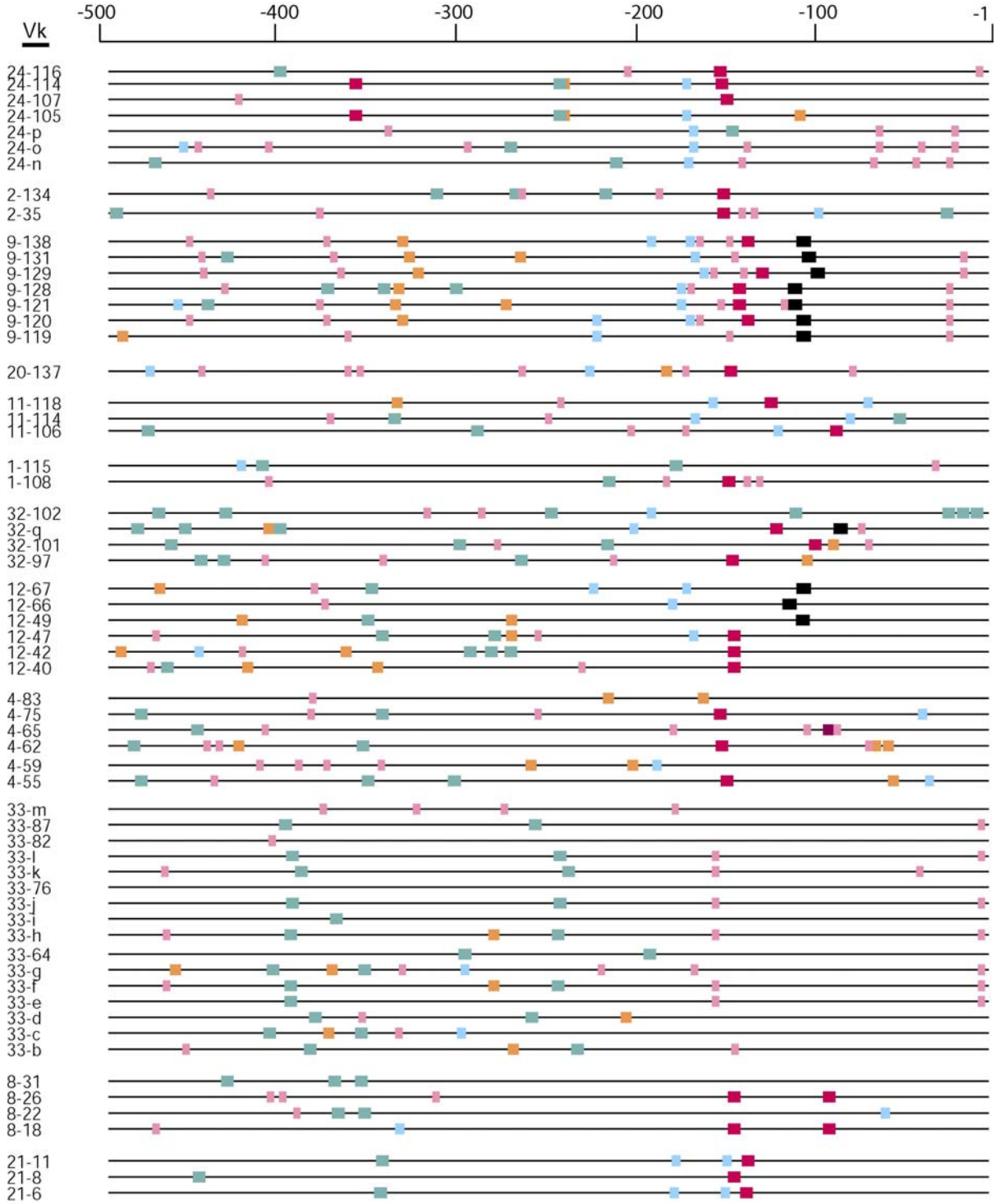
-500
-400
-300
-200
-100
-1
Vk
24-116
24-114
24-107
24-105
24-p
24-o
24-n
2-134
2-35
9-138
9-131
9-129
9-128
9-121
9-120
9-119
20-137
11-118
11-114
11-106
1-115
1-108
32-102
32-q
32-101
32-97
12-67
12-66
12-49
12-47
12-42
12-40
4-83
4-75
4-65
4-62
4-59
4-55
33-m
33-87
33-82
33-l
33-k
33-76
33-j
33-i
33-h
33-64
33-g
33-f
33-e
33-d
33-c
33-b
8-31
8-26
8-22
8-18
21-11
21-8
21-6

Figure 2.3 (previous page) – Putative transcription factor binding sites within the 500 bp immediately upstream of each Vκ pseudogene and relic. V regions are grouped by family in the order of appearance in the locus 5' to 3'. Octamer/variant octamer – red /open red, TATA – open white, HMG box – orange, E-box – blue, Ikaros – pink, Forkhead – green, PU.1 – black, NF-κB – maroon.

Alignment of coding and intron regions reveals interesting similarities in transcription factor binding sites. To determine whether there are conserved sites in the introns or exons of any of the potentially functional V gene families, I aligned these genes starting from their leader exons (Figure 2.4). Examination of this alignment demonstrates that variation occurs in the intron length even in highly conserved families, but this variation is generally less that that seen between different families (Figure 2.4, narrow lines).

There are a number of additional observations that are obvious upon visual inspection of sequence motifs within this alignment. With the exception of the intermingled Vκ8, 19, 22 and dv families, all V gene families have clusters of potential Ikaros binding sites in their introns, or conserved winged helix/forkhead binding sites, or both (Figure 2.4, pink and green boxes, respectively). All of the gene families that do not possess these patterns have potential Lef-1 binding sites in their introns, excepting the single gene Vκdv family (Figure 2.4, lower, orange boxes).

Two Vκ gene families contain unique clusters of putative binding sites for either Ebf or NF-κB in their introns. The introns of the Vκ4 gene family have conservation of potential Ebf binding sites (Figure 2.4, center, bright yellow boxes). This conservation occurs in most of the same members that contain a potential Lef-1 binding site in their upstream regions; 14 Vκ4 genes have both the Lef-1 binding site in their intron and the Ebf sequence in their

promoter. (Figure 2.4, center, orange boxes). Vκ4-56 displays only the promoter Lef-1 site and four Vκ4 genes have the Ebf intron site without the promoter Lef-1 site. Four of the five Vκ1 family members were found to have a potential NF-κB binding site in their intron (Figure 2.4, upper, dark red boxes). Interestingly, this statistically improbable and conserved positioning of potential Ebf and NF-κB binding sites is unique and not shared with any other Vκ gene families.

The nearly complete absence of E boxes in the intron sequences is very noticeable (3 out of 95), although they often are found in exons, frequently near either the 5' or 3' intron junctions (Figure 2.4, blue boxes). This can be seen in the two most 3' families, Vκ21 and Vκ19, where all but one gene has an E box at the intron exon boundary (Figure 2.4). In addition, there are no conserved clusters of octamer sites (Figure 2.4, red boxes), and only the 4-91 gene has a potential PU.1 binding site in its intron (Figure 2.4, center, black box). There is a complete absence of conserved Pip or Stat-5 binding sites in the introns of Vκ genes.

In conclusion, the only conservation that is shared between most Vκ genes is the high frequency of sites for chromatin modulating factors in their introns. Potential binding sites for Ikaros, winged helix/forkhead and/or Lef-1 proteins often exhibit highly significant non-random positioning in the introns of most Vκ genes (Table 2.3). In addition, the introns of Vκ4 and Vκ1 gene families possess potential Ebf and NF-κB binding sites, respectively.

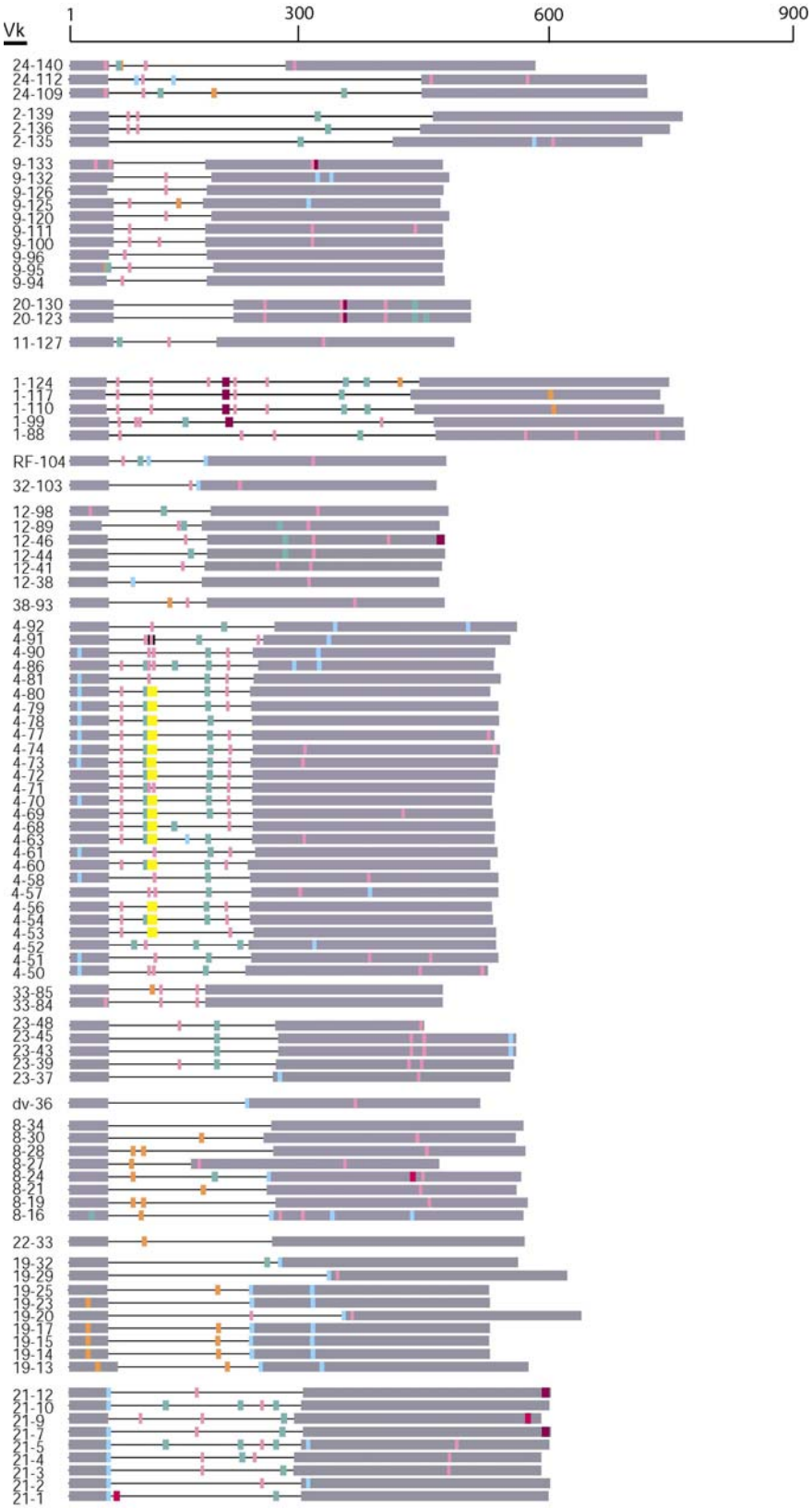
Vk
1
300
600
900
24-140
24-112
24-109
2-139
2-136
2-135
9-133
9-132
9-126
9-125
9-120
9-111
9-100
9-96
9-94
20-130
20-123
11-127
1-124
1-117
1-110
1-99
1-88
RF-104
32-103
12-98
12-89
12-46
12-44
12-41
12-38
38-93
4-92
4-91
4-90
4-86
4-81
4-80
4-79
4-78
4-77
4-74
4-73
4-72
4-71
4-70
4-69
4-68
4-63
4-61
4-60
4-58
4-57
4-56
4-54
4-53
4-52
4-51
4-50
33-85
33-84
23-48
23-45
23-43
23-39
23-37
dv-36
8-34
8-30
8-28
8-27
8-24
8-21
8-19
8-16
22-33
19-32
19-29
19-25
19-23
19-20
19-17
19-15
19-14
19-13
21-12
21-10
21-9
21-7
21-5
21-4
21-3
21-2
21-1

Figure 2.4 (Previous page) – Putative transcription factor binding sites located within the coding exons and intron of each potentially functional Vκ gene. Sequences are grouped by V gene family, listed in the order that they occur in the locus, 5' to 3', from top to bottom. The left side of the alignment corresponds to 5' ends of sequences and starts at the ATGs in the leader exons as +1 bp and extends downstream to the right to the RSSs. Exons are represented as light gray rectangles and introns as narrow lines. Color coded boxes represent the following potential binding sites for transcription factor: pink, Ikaros; green, winged helix/forkhead; orange, Lef-1; bright yellow, Ebf; dark red, NF-κB; blue, E box; black, PU.1; and red, octamer.

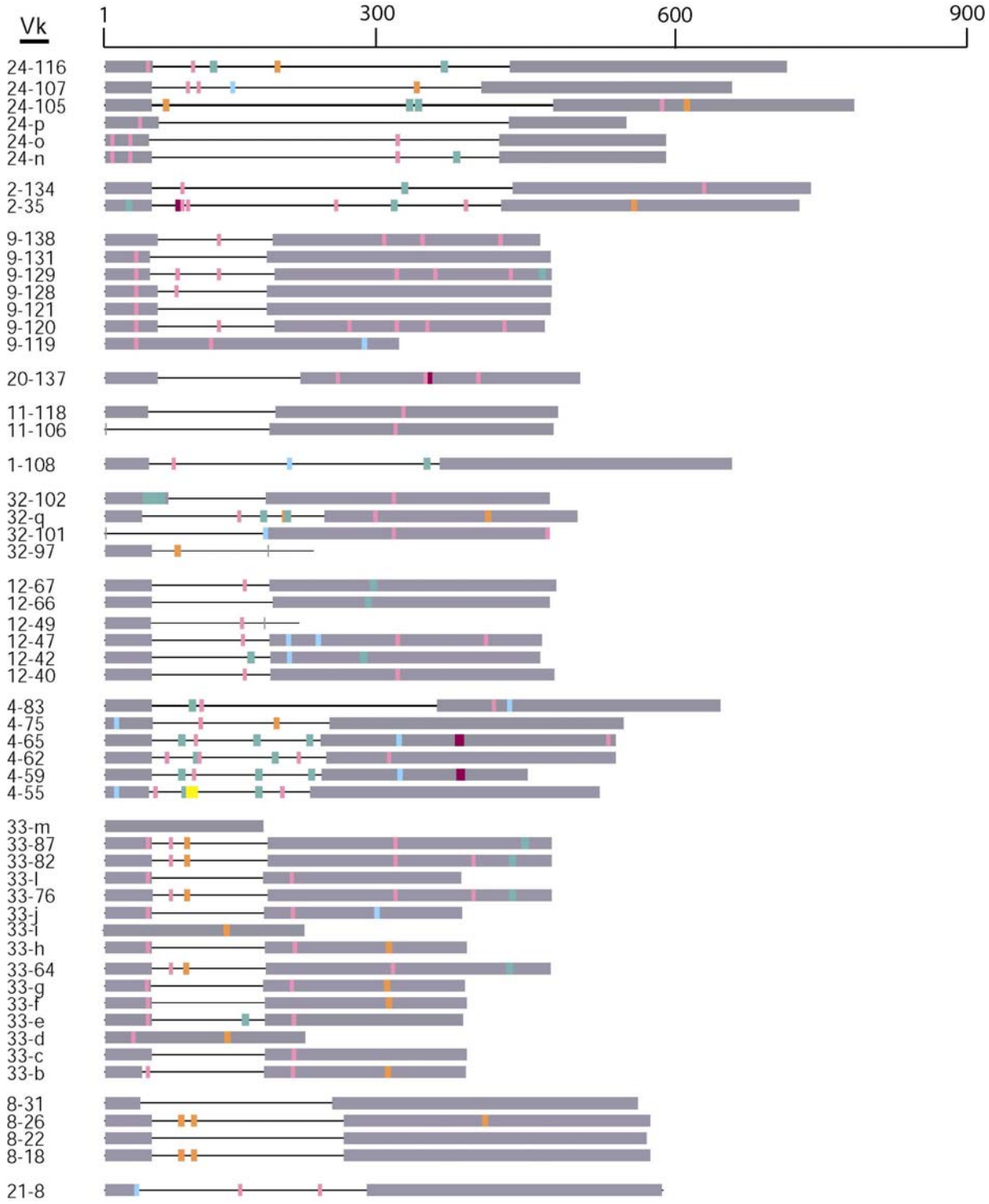
Vk
1
300
600
900
24-116
24-107
24-105
24-p
24-o
24-n
2-134
2-35
9-138
9-131
9-129
9-128
9-121
9-120
9-119
20-137
11-118
11-106
1-108
32-102
32-q
32-101
32-97
12-67
12-66
12-49
12-47
12-42
12-40
4-83
4-75
4-65
4-62
4-59
4-55
33-m
33-87
33-82
33-l
33-76
33-j
33-i
33-h
33-64
33-g
33-f
33-e
33-d
33-c
33-b
8-31
8-26
8-22
8-18
21-8

Figure 2.5 (previous page) – Putative transcription factor binding sites located within the coding exons and intron of Vκ pseudogenes and relics. Vκ segments are listed from 5' to 3' as in Fig 2.3. Exons are represented as light grey rectangles. Ebf – yellow, HMG box – orange, E-box – blue, Ikaros – pink, Forkhead – green, PU.1 – black, NF-κB – maroon.

Sequences downstream of the RSS also exhibit conserved transcription factor binding sites. Localized transcription factor binding has been associated with recruitment of chromatin remodeling complexes in other systems, and could be involved in generating accessibility of RSSs to the recombination apparatus (for reviews, see Felsenfeld and Groudine, 2003; Hampsey and Reinberg, 2003). To explore this idea, I determined whether there were conserved sites of potential importance downstream of the RSSs. I performed the same type of alignment as for the upstream regions and introns. However, this time I aligned the RSSs and the 500 bp immediately downstream. Conservation of sites decreased dramatically after 300 bp, therefore I present results only 300 bp downstream of the RSS.

The most striking and completely unexpected finding that I made is a family-specific conservation in octamer sites. There are no published reports of this conservation. Eighteen members of the centrally located cluster of Vκ4 genes contain conserved octamer sites in their downstream regions ~100 bp from their RSSs (Figure 2.6, center, red boxes). Fourteen of these same genes also contain potential Lef-1 binding sites in their promoters (Figure 2.2, orange boxes) and have the conserved potential Ebf binding sites in their introns (Figure 2.4, bright yellow boxes). Neither the upstream Lef-1 site, the intronic Ebf site, nor the downstream octamer site, however, is conserved in the Vκ4 pseudogenes that are interspersed in this region (Figures 2.3, 2.5, and 2.7, respectively).

Another intriguing finding that I made by this analysis is the statistically significant conservation of E boxes downstream of the RSSs. With the exception of the Vκ4, Vκ24 and Vκ21 families, there is a significant conservation of E boxes downstream of the RSS (Figure 2.7, blue boxes). Curiously, these E box deficient families are located in the center and at either end of the locus. Vκ1 genes, which do not have any noticeable level of conservation of E boxes in their promoters, contain one completely conserved E box in their downstream regions (Figure 2.6, upper, blue boxes).

There also is a high frequency of potential Ikaros and winged helix/forkhead binding sites downstream of RSSs (Figure 2.6, pink and green boxes, respectively). Winged helix/forkhead sites are present at a very high frequency; most Vκ genes contain 2-4 consensus sequences within the 300 bp region downstream of the RSSs.

As in the upstream region, there is a lack of conservation of potential PU.1 (Figure 2.6, black boxes), Ebf (not found), Pip (data not shown, statistically insignificant) or Stat-5 (data not shown, statistically insignificant) binding sites in the region downstream of the RSSs. In addition, the only Vκ genes possessing downstream potential NF-κB binding sites are several Vκ1 gene family members; interestingly, this family is also the only one exhibiting such sites in the genes' introns (Figures 2.4 and 2.6, upper, dark red boxes). Potential Lef-1 binding sites are not clustered downstream of any Vκ gene family, although they appeared numerous times downstream of different Vκ genes (Figure 2.6, orange boxes).

In conclusion, I uncovered the significant conservation of potential Oct1/2 and E2A transcription factor binding sites in the region immediately downstream of many Vκ genes' RSSs.

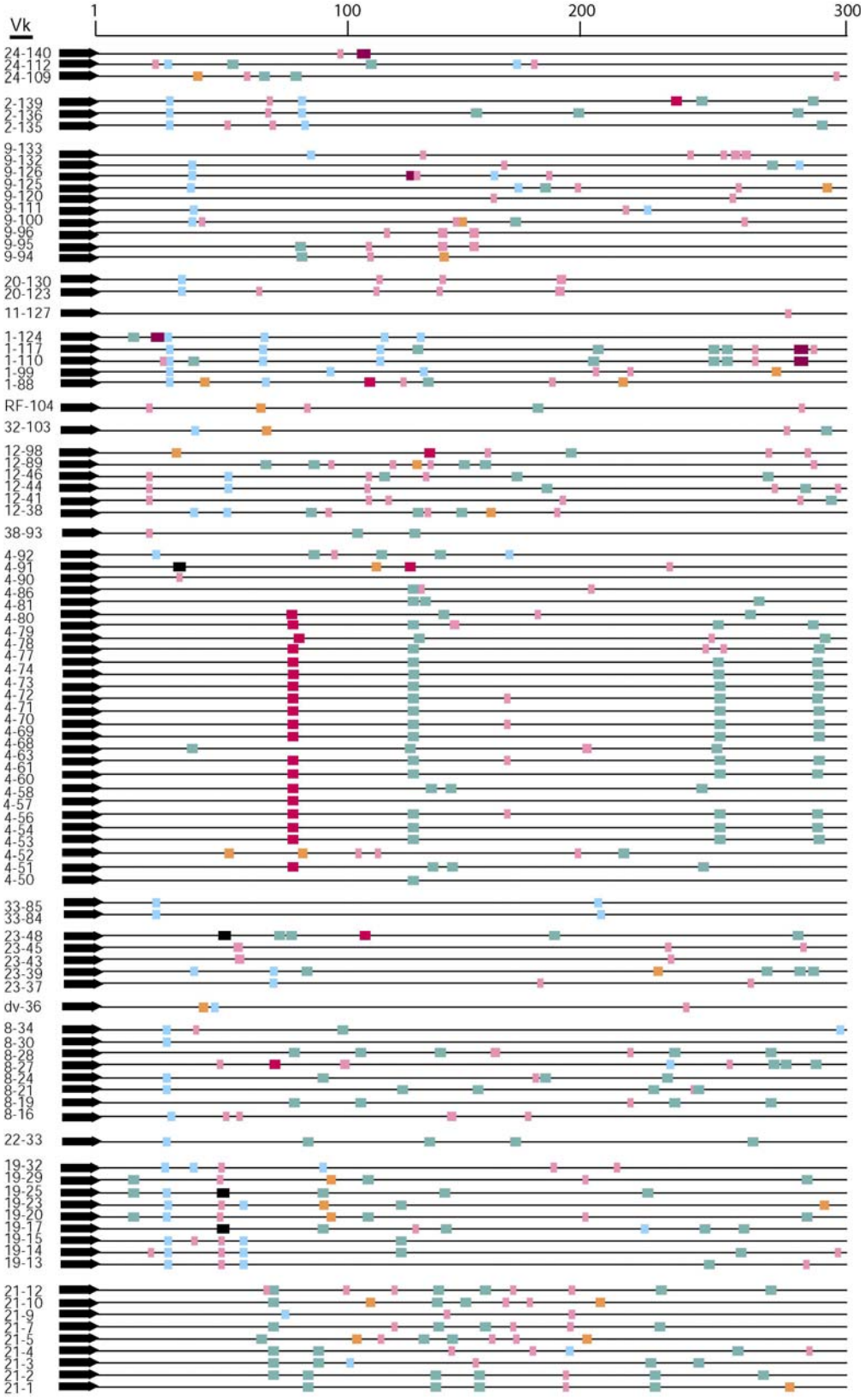
Vk
1
100
200
300
24-140
24-112
24-109
2-139
2-136
2-135
9-133
9-132
9-126
9-125
9-120
9-111
9-100
9-96
9-95
9-94
20-130
20-123
11-127
1-124
1-117
1-110
1-99
1-88
RF-104
32-103
12-98
12-89
12-46
12-44
12-41
12-38
38-93
4-92
4-91
4-90
4-86
4-81
4-80
4-79
4-78
4-77
4-74
4-73
4-72
4-71
4-70
4-69
4-68
4-63
4-61
4-60
4-58
4-57
4-56
4-54
4-53
4-52
4-51
4-50
33-85
33-84
23-48
23-45
23-43
23-39
23-37
dv-36
8-34
8-30
8-28
8-27
8-24
8-21
8-19
8-16
22-33
19-32
19-29
19-25
19-23
19-20
19-17
19-15
19-14
19-13
21-12
21-10
21-9
21-7
21-5
21-4
21-3
21-2
21-1

Figure 2.6 (Previous page) – Putative transcription factor binding sites within the 300 bp downstream regions of each potentially functional Vκ gene. Sequences are grouped by V gene family, listed in the order that they occur in the locus, 5' to 3', from top to bottom. The left side of the alignment starts corresponds to the 5' ends of sequences and starts at the RSSs (black arrows), and extends downstream 3' for 300 bp. Color coded boxes represent the following potential binding sites for transcription factors: red, octamer; blue, E box; pink, Ikaros; green, winged helix/forkhead; dark red, NF-κB; orange, Lef-1; and black, PU.1. Ebf was not found.

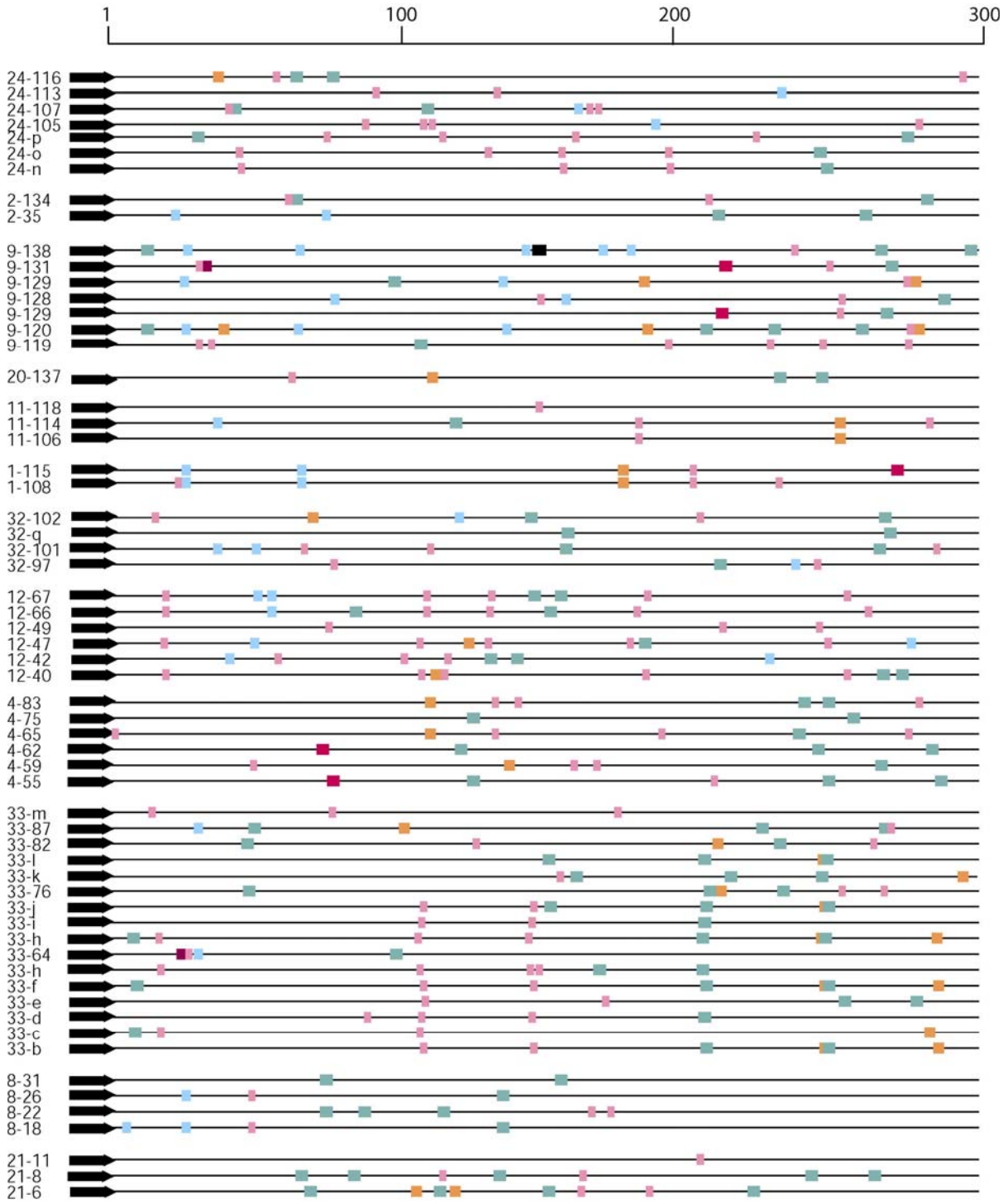
1
100
200
300
24-116
24-113
24-107
24-105
24-p
24-o
24-n
2-134
2-35
9-138
9-131
9-129
9-128
9-129
9-120
9-119
20-137
11-118
11-114
11-106
1-115
1-108
32-102
32-q
32-101
32-97
12-67
12-66
12-49
12-47
12-42
12-40
4-83
4-75
4-65
4-62
4-59
4-55
33-m
33-87
33-82
33-l
33-k
33-76
33-j
33-i
33-h
33-64
33-h
33-f
33-e
33-d
33-c
33-b
8-31
8-26
8-22
8-18
21-11
21-8
21-6

Figure 2.6 (Previous page) – Putative transcription factor binding sites in the 300 bp immediately downstream of Vκ pseudogenes and relics. Sequences are organized as in Fig 2.2 and 2.4. Octamer – red, HMG box – orange, E-box – blue, Ikaros – pink, Forkhead – green, PU.1 – black, NF-κB – maroon.

Discussion

The most significant finding of this study is that there are several regions in which potential binding sites for transcription factors cluster around Vκ genes. This occurs either in the upstream region, the intron, or downstream of the RSS. This clustering of sites appears to be family specific, with the majority of the members of a given family containing common sites of clustering. Most families only contain such clustering in either one or two of the three regions we analyzed. The only obvious exception to this is the Vκ4/5 family, which possesses clusters of potential binding sites for Lef-1, Ebf, and Oct1/2 in the upstream, intron and downstream regions, respectively. These overall clustering results are novel as they point to regions in addition to Vκ gene promoters as being potentially functionally important in targeting Vκ genes for recombination, namely introns and the region downstream of the RSS. I predict that this positioning will have functional significance in V-J joining. We have observed that the RSS region of the germline Vκ4-60 gene is enriched for Oct-1 at this same point in development (unpublished results from Zhe Liu). These results suggest different combinatorial ways in which Vκ genes might be targeted for recombination. Chromatin opening by transcription factor binding need not require the process of transcription to occur, as it is well established that many such factors simply recruit various chromatin remodeling proteins (for review, see Felsenfeld and Groudine, 2003).

Some Vκ genes do not obey the above family rule of clustering. Exceptions may exist for at least three reasons. First, some binding sites that I analyzed may mismatch the

consensus sequence by only one or two bp. For certain transcription factors, depending on the context, this degree of mismatch may be functionally tolerated. If such variant conserved sites do exist, I would have missed them in my analysis. The second reason for exception to the family rule of clustering could be that a Vκ gene that is an outlier in its family may be non-functional. The criterion for classifying a Vκ gene as potentially functional was the presence of a matching cDNA sequence in the NCBI's database (Thiebe et al., 1999). This database consists of sequences derived from numerous strains of mice. It is very likely that each inbred mouse strain has its own classification for functional- and pseudo-Vκ genes and that mutations over time contribute to strain-specific differences in this classification. This has been shown to be the case for some of the Vκ1 gene family members in different strains of mice (Ng et al., 1989). Therefore, some of the functional Vκ regions in the NCBI database that we assigned to be potentially functional could be pseudogenes in varying strains of mice**.** A third possibility for deviation from family member clustering is that it is possible that other transcription factors, not studied here, are important in the initiation of V(D)J recombination at the level of targeting.

The Vκ1 gene family is interesting because of its non-conserved nature. It is a very highly used family, especially in initial rearrangements (Medina and Teale, 1993; Ramsden et al., 1994b); it undergoes germline transcription and has an altered chromatin structure during the pre-B cell stage of development (See Chapter 3), but most of its members have a variant octamer site in their promoters. These observations support our idea that regions other than a gene's promoter may be important for the targeting of the recombination machinery. In this regard, Vκ1 genes comprise the only family that possesses potential NF-κB binding

sites in their intronic and downstream regions. Furthermore, Vκ1 genes also possess multiple E boxes in their downstream regions.

It is perhaps puzzling that the only conserved Ebf sites uncovered in this study were in the introns of a group of Vκ4 genes. Ebf has been implicated to play a role in V(D)J recombination of both IgH and IgL genes (Goebel et al., 2001; Sigvardsson et al., 1996; Sigvardsson et al., 1995). The paucity of Ebf binding sites would suggest that either Ebf functions through binding at sites in the locus other than at Vκ genes, or Ebf may regulate other non-Igκ genes whose products act on Vκ genes, or Ebf may simply be less important in generating accessibility to the Igκ locus. It is also possible as mentioned above that Ebf sites were not picked up in this analysis because the consensus sequence searched for was either imperfect or too stringent. However, it has been shown that the Vκ4 gene family is preferentially rearranged in both Abelson transformed pre-B cell lines and in the fetal livers of embryonic mice (Kalled and Brodeur, 1990; Ramsden et al., 1994b), therefore Ebf may play an important role in these initial rearrangement events.

The centrally located Vκ4 gene family is an enigma that exhibits features strikingly different from other families. Its members are the most conserved and possess unique patterns of potential binding sites for several transcription factors. Within the Vκ4 family, the patterns of transcription factor binding sites are conserved more so than the surrounding sequence. For example, if each of the Vκ4 genes' promoters is aligned using the Oct site as a guide, there is 100% identity between all members at the eight bp site. However, the eight bp 5' to the site are only 80% identical and the 3' side is only 75% identical. This is interpreted

to indicate that the conservation is due to a functional requirement and that the sites would have diverged evolutionarily without selection.

Several factors that are known to play important roles in regulating B cell development and V(D)J recombination are likely not involved as major players in directly regulating the accessibility of Vκ genes. Besides NF-κB and Ebf as discussed above, factors such as PU.1, Stat-5 and Pip/IRF4 are also likely to have most of their duties in the process elsewhere as evidenced by the lack of sites within or around most Vκ genes. PU.1 sites are only present in the promoters of a few Vκ genes. The involvement of PU.1 and NF-κB is likely to be principally attributed to the enhancers in the J-C region, since these sites are nearly random elsewhere throughout the entire Igκ locus.

CHAPTER THREE

Identification and Characterization of DNase I Hypersensitive Sites Surrounding Igκ Variable Gene Segments

Summary

Regulation of chromatin accessibility is thought to play a major role in determining the specificity of the V(D)J recombination process. This process is tightly developmentally regulated in both a cell type and locus specific manner. Here I provide *in vivo* evidence that the accessibility of the chromatin surrounding immunoglobulin kappa (Igκ) variable (V) gene segments alters during the pre-B cell stage of development. This accessibility is demonstrated by the presence of DNase I hypersensitive sites (HSs) that localize at or near the promoters and recombination signal sequences (RSS) of the V gene segments. The HS_P are present only during the pre-B cell stage of development and do not correlate with germline transcription from the same V regions. HS_P are present in all functional germline Vκ gene segments tested. HS_P correlate with the conserved transcription factor binding sites discussed in Chapter Two. Differences in the patterns of sites for specific transcription factors among different V gene families suggest that there is a variety of combinatorial pathways to achieve an open chromatin configuration of Vκ genes.

Introduction

It has been proposed that V(D)J recombination is regulated by modulation of chromatin accessibility to the recombination machinery (Yancopoulos and Alt, 1986). Accessibility can be measured in a number of ways. Germline transcription from promoters near Vκ and Jκ genes often precedes or accompanies the rearrangement process (Casellas et al., 2002; Schlissel and Baltimore, 1989; Schlissel et al., 1991b). Therefore this has been a traditional measure of accessibility of chromatin to transcription and other protein machinery. Monoallelic demethylation and early DNA replication have also been implicated as changes that occur in the active allele of the Igκ locus (Mostoslavsky et al., 1998; Mostoslavsky et al., 2001). Histone acetylation and methylation status also changes as genes become "accessible." Changes in acetylation status have been observed in Ig and TCR systems (Agata et al., 2001; Cherry and Baltimore, 1999; McBlane and Boyes, 2000; McMurry and Krangel, 2000), but the importance of histone methylation is not known, although a role in silencing has been suggested for this phenomenon (Ng et al., 2003). Interestingly, methylation of DNA leads to the recruitment of histone deacetylases that inhibit recombination by closing chromatin (Ji et al., 2003; Nan et al., 1998). One other method for examining accessibility is the use of *in vivo* nuclease cleavage assays (Aronow et al., 1995; Asenbauer and Klobeck, 1996; Chung et al., 1983; Ellis et al., 1996; Gross and Garrard, 1988; Haines and Brodeur, 1998; Li and Wrange, 1997; Liu et al., 2002; Madisen and Groudine, 1994; Maes et al., 2001; Mather and Perry, 1983; Mymryk et al., 1997; Persiani and Selsing, 1989; Roque et al., 1996; Storb et al., 1981). Support for this proposal comes

from more recent studies in model *in vitro* systems with nucleosomes reconstituted on recombination substrates have demonstrated that accessibility of the RSS depends on nucleosome positioning, HMG1, histone acetylation and SWI/SNF remodeling (Kwon et al., 1998; Kwon et al., 2000).

The chromatin accessibility hypothesis is supported by the observation that transiently transfected recombination substrates can be induced to undergo rearrangement upon expression of RAG-1/2 proteins in non-B cells (Oettinger et al., 1990; Schatz et al., 1989), but the corresponding endogenous alleles remain in the unrearranged, germline state (Romanow et al., 2000). Additionally, the chromatin of mouse Vκ and Jκ genes becomes generally sensitive to DNase I at the pro- to pre-B cell transition (Maes et al., 2001). Other direct support for the chromatin accessibility hypothesis come from the demonstration that appropriate RSS in early B and T cells are targets for RAG-1/2 cleavage *in vivo* (Constantinescu and Schlissel, 1997) or in isolated nuclei (Goldmit et al., 2002; Stanhope-Baker et al., 1996).

I chose to examine the chromatin structure surrounding germline Vκ genes, using the data from Chapter II as a guide. Chromatin structure changes can by monitored by examining the sensitivity of the chromatin to nucleases, germline transcription in the regions of interest, studies of methylation status and acetylation status of histones. I chose to use DNase I hypersensitivity assays as well as test for the presence of germline transcription from individual Vκ genes. These studies demonstrated that Vκ genes exhibit HS_P and sometimes HS_{RSS} specifically at the pre-B cell stage of development. These sites are not caused by RAG expression and are not inducible by traditional "chromatin activating"

factors. I have followed the initial observation of HSs by attempting to determine the cause of the sites and have observed the germline transcription status from the Vκ genes of interest. I have also demonstrated that germline transcription of a particular Vκ gene is being neither necessary nor sufficient for maintenance of the HSs.

Materials and Methods

Cell lines. BASC6 C2, AH7, 38B9, T7, S7, 63-12, and B10 were maintained in RPMI 1640 supplemented with 10% FBS and 50μM β-mercaptoethanol at 37°C, 5% CO_2. 103Bcl2 cells were maintained in RPMI 1640 as above except at 34°C. S194 cells were maintained in Iscove's MDM supplemented with 5% FBS and 50μM β-mercaptoethanol and grown at 37°C and 10% CO_2. The BASC6 C2, T7, and B10 cell lines were a generous gift from Eugene Oltz (Vanderbilt University). The AH7 cell line was a generous gift from Christopher Roman and David Baltimore (Caltech). For drug treatments, cells were treated for 6 days with samples taken for DNase I analysis at 24 hour increments. Cell lines were treated with varying concentrations of Trichostatin-A (TSA) (WAKO Pharmaceuticals) and 5-Aza-2-deoxycytidine (Aza-C) (Sigma), both stocks dissolved in ethanol. For TSA, 25nM, 50nM, 100nM; Aza-C, 1μM, 5μM, 10μM. Cells were also mock treated in parallel with ethanol. These concentrations were selected based on the manufacturer's recommendation and personal experience of Michael Hale in the laboratory.

DNase I treatment of cells. Cells were grown to a density of approximately 10^6 per ml, counted and spun at 1000xg for 5 min. 12 x 10^7 cells were used for each experiment. The cell pellet was washed in 1X PBS and permeabilized in Solution I (150mM sucrose, 80 mM KCl, 35mM HEPES pH 7.4, 5mM K_2HPO_4, 5mM $MgCl_2$, 0.5mM $CaCl_2$, 0.02% NaAzide) + 0.05% hen egg white lysolecithin (Sigma, St. Louis, MO) at 37°C and incubated at 37°C for 1 min. The cells were then spun at 1000xg for 5 minutes, washed in Solution I

without lysolecithin and transferred to microfuge tubes. The pellet was then slurried and resuspended in Solution II (150mM sucrose, 80mM KCl, 35mM HEPES pH7.4, 5mM K_2HPO_4, 5mM $MgCl_2$, 1mM $CaCl_2$, 0.02% NaAzide). DNase I (Worthington) was added in varying concentrations (from 0.125 U/ml to 8.0 U/ml) and incubated at room temperature for 5 min. The samples were then incubated on ice for 10 min., spun at 13,000rpm for 5 min. and lysed by the addition of 200µl of 20mM Tris-HCl pH 8.0, 20mM NaCl, 20mM EDTA, 1% SDS, 600µg/ml proteinase K. After overnight incubation, the samples were diluted with an equal volume of 150mM NaCl, 5mM EDTA, phenol:chloroform extracted and ethanol precipitated. The resuspended samples were treated with DNase I-free, RNase A (Roche) at a final concentration of 100µg/ml at 37°C for 2 hours, phenol:chloroform extracted, ethanol precipitated, resuspended in a final volume of 200µl of TE, and quantitated. The average DNA yield ranged from 100µg to 200µg depending on the extent of DNase I digestion.

*Generation of probes for Southern analysis***.** A 684 bp single copy DNA fragment was amplified from 3.5 kb upstream of Vκ2-139. A BAC clone containing this upstream region was used as template. PCR primers (5' to 3') were: GTC TCT TCT AGC CCA CTG ACC ATA G and CTT GAA TGG CCA AGA AGC ACT TAG AGA, with an annealing temperature of 55°C. For detection of HS near Vκ1-110 and –117, a 1.4 kb *Bgl*II, *Hin*dIII fragment, 3' to a *Bam*HI site was excised from a the plasmid p1.2.8 that contains the Vκ1-110 gene. The p1.2.8 plasmid was a generous gift of Dr. David Gibson.

Southern analysis. 15μg of each DNase I treated sample was digested with restriction enzymes as noted in figure legends and electrophoresed in 0.8% SeaKem GTG (FMC Bioproducts) gels with 1X TAE running buffer for 10 hours at 60 volts. Gels were transferred overnight in 0.4N NaOH, 0.25M NaCl to Zeta-Probe GT filters (Bio-Rad). Filters were hybridized overnight in 7% SDS, 1mM EDTA, 250mM sodium phosphate (pH 7.4), and 1% BSA at 65°C. Probes were generated by random priming with Rediprime II (AP Biotech) according to the manufacturer's instructions. Filters were washed 2 times at 65°C with 2% SDS, 0.5X SSC and 20mM sodium phosphate and 1X in 2% SDS, 0.2 SSC and 20mM sodium phosphate. After washing, filters were either exposed to film (Kodak XAR-5) for 2-5 days or exposed overnight to PhosphorImager screens and analyzed by PhosphorImaging (Molecular Dynamics).

RNA isolation and preparation of cDNA from cell lines. 10ml of cell lines above were grown to a concentration of approximately 10^6 cells/ml. Total cellular RNA was isolated using RNA-Stat 60 (Tel-Test, Friendswood, TX) according to the manufacturer's instructions. RNA was resuspended in 50μl of 1mM EDTA and quantitated. OD260/280 ratios were greater than 1.8 and the average concentration of RNA was between 0.5 and 1μg/μl. 1-5μl of the isolated RNA was used to generate cDNA with the Superscript II system (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions.

Germline transcription assays. PCR reactions were carried out as follows: 1μl cDNA from above, 10 pmol of each appropriate primer, and Taq polymerase in 50μl

reactions. Annealing temperatures were 65°C for the GAPDH and Jκ reactions and 58°C for all others. PCR conditions (30 cycles) were as follows: 94°C, 2 min; 94°C 30 sec; 58°C or 65°C, 1 min; 72°C, 30 sec and a final 10 minute extension at 72°C. Primers were designed to be specific to the particular unrearranged Vκ region being analyzed. 1/10 of each reaction was analyzed by ethidium bromide staining and Southern analysis as above.

	5' primer (5' to 3')	3' primer (5' to 3')	°C
Vκ2-139	GTTACCATTGGACAACCAGCCTC	GTCAGCTGGGCTGCCCCAGGC	58
Vκ1-117	CAGTAATGGAAACACCTATTTAC	GCTGTGATACCCAAAGTAAGAC	58
Vκ1-110	TAGTAATGGAAACACCTATTTAG	GCTGTGATACCCAAAGTAAGAC	58
Jκ	GAGGGGGTTAAGCTTTCGCC-TACCCAC	CTGTATCTTTGCCTTGGAGAG-TGCCAGAATCTGG	65
RAG-1	CCAAGCTGCAGACATTCTAGCACT-CTGGCCG	CAACATCTGCCTTCAC-GTCGATCC	58
RAG-2	GAGATCCACAAGCAGGAAGTACACT-TCATAC	GGTTCAGGGACATCTCCTAC-TAAGTCTTTCTC	58
GAPDH	TTTGTGATGGGTGTGAACCA	TGCCAGTGAGCTTCCCGTTC	65

Table 3.1 – Primer pairs and PCR conditions used for RT-PCR

Results

Vκ gene segments possess HS that localize around the RSS and promoter prior to the onset of Vκ-Jκ joining. Alterations in chromatin structure are commonly monitored by the appearance of DNase I hypersensitive sites (HS). Here I monitored the presence of HS during various stages of B cell development in the chromatin surrounding a germline Vκ2 (2-139) gene segment. This V gene is located in the most 5' region of the 3.2 Mb locus. I found that there are HS that form specifically in cell lines poised to, or undergoing V-J joining in culture (Table 3.2). These HS were localized to the promoter region and near the RSS and are referred to as HS_P and HS_{RSS}, respectively (Figure 3.1).

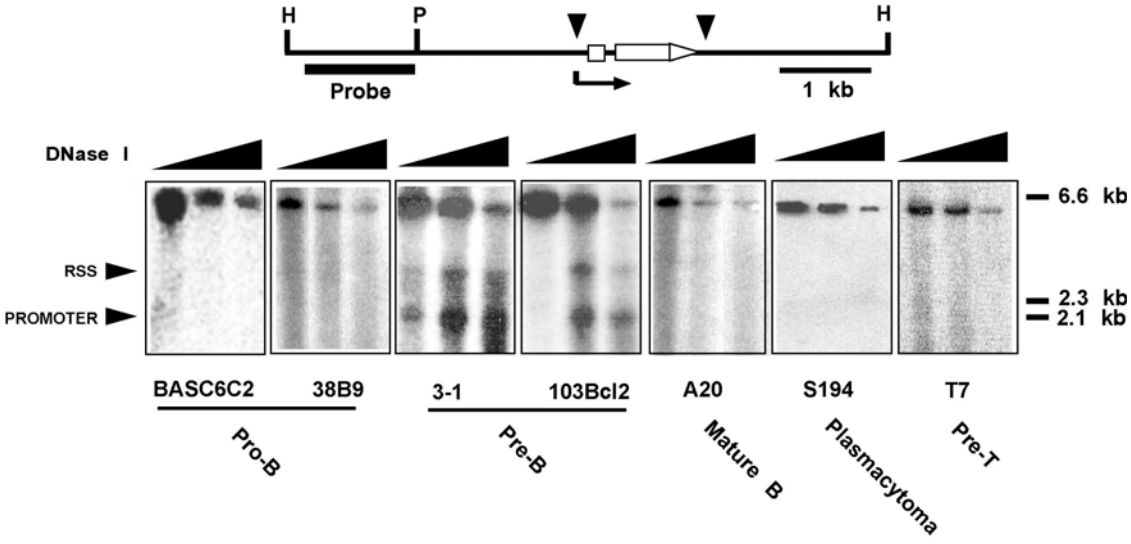

Figure 3.1. HSs specifically form at the promoter and RSS of a germline Vκ2 gene during the pre-B cell stage of development. Cell lines were permeabilized and treated with DNase I. Purified DNA was digested with HindIII and Southern hybridization was carried out with a probe adjacent to the 5' HindIII cut site for indirect end-labeling. Table 3.2 indicates the developmental stage and rearrangement status of each of the cell lines used in this chapter.

Cell line	Stage	IgH	Igκ	RAG-1	RAG-2	HS_{P-2}	HS_{P-1}	HS_{RSS-2}	HS_{RSS-1}	Vκ2 txn	Vκ1 txn	Jκ txn	Ref(s)
BASC6C2	Pro-B	D to J	G	Yes	Yes	No	No	No	No	Yes	Yes	No	(Lieber et al., 1987)
38B9	Pro-B	V to DJ	G	Yes	Yes	No	No	No	No	No	Yes	No	(Lieber et al., 1987)
103Bcl2	Pre-B	R	V to J	Yes	Yes	Yes	Yes	Yes	No	Yes	No	No	(Chen et al., 1994; Klug et al., 1994)
3-1	Pre-B	R	V to J	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	(Lieber et al., 1987)
A20	Mature B	R	R	No	No	No	No	No	No	No	Yes	No	(Kim et al., 1979)
S194	Plasmacytoma	R	R	No	No	No	No	No	No	Yes	No	No	(Horibata and Harris, 1970)
T7	Pre-T	G	G	Yes	No	No	No	No	No	Yes	No	No	(Ferrier et al., 1990)
AH7	Pro-B	G	G	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	(Roman et al., 1997)
63-12	Pro-B	G	G	Yes	No	Yes	Yes	Yes	No	Yes	Yes	No	(Shinkai et al., 1992)
EL-4	Mature T	G	G	No	No	No	No	No	No	No	No	No	(Shevach et al., 1972)
P815	mastocytoma	G	G	No	No	No	No	No	No	No	No	No	(Lundak and Raidt, 1973)

Table 3.2 (previous page) – Characteristics of cell lines used in this study. Rearrangement status is indicated as either G (germline), R (rearranged), or by the segments which can be induced to undergo rearrangement preferentially in culture. HS_{RSS-1} and HS_{RSS-2} indicate the presence of HS at or near the RSS of either Vκ1 or Vκ2 genes, respectively. txn – transcription.

Vκ gene segments exhibit HSRSS in the absence of RAG-1 and –2. Because RAG-1 and –2 are the only proteins currently known to bind to the RSS, I examined the Vκ gene segments in cell lines derived from mice in which either RAG-1 or RAG-2 has been deleted (Roman et al., 1997; Shinkai et al., 1992). I found for both RAG-1 and RAG-2 -/- cells that both HS_P and HS_{RSS} were present at Vκ2-139 as in the wild-type cells (Figure 3.2 and data not shown). I also found that there were no detectable HS_{RSS} in the Jκ regions of the chromatin of these cells, see Figure 3.5. To determine whether this was a localized phenomenon, I carried out HS mapping at two other V regions, Vκ1–110 and Vκ1-117 located more centrally in the locus (Ng et al., 1989). In this case, I found only HS_P (Figure 3.2 and data not shown).

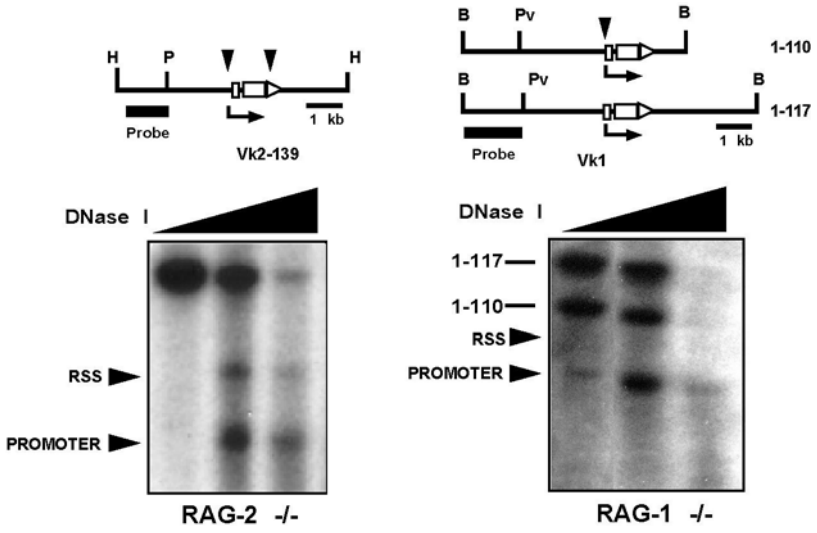

Figure 3.2 - RAG proteins are not responsible for generating HS. RAG-1-/- and RAG-2-/- cell lines were treated with DNase I as above and purified DNA was analyzed as in Fig. 3.1

Since opening of chromatin surrounding active genes has been correlated with increases in acetylation of histones and demethylation of DNA, I attempted to mimic these effects in cell lines to generate the open chromatin configuration in pro-B cells. I was unable to generate the HS by treating pro-B cells (38B9 and BASC6 C2) with bacterial lipopolysaccharide (LPS), the histone deacetylase inhibitor Trichostatin-A, or the non-methylatable cytosine analog, 5-aza-2-deoxycytidine (Figure 3.3). These cells were treated with varying concentrations of LPS, TSA or 5-Aza, alone or in combination in an attempt to generate HS_P and HS_{RSS}. However, these drugs had no measurable effect on the chromatin structure, suggesting that other factors are involved in generating this novel state.

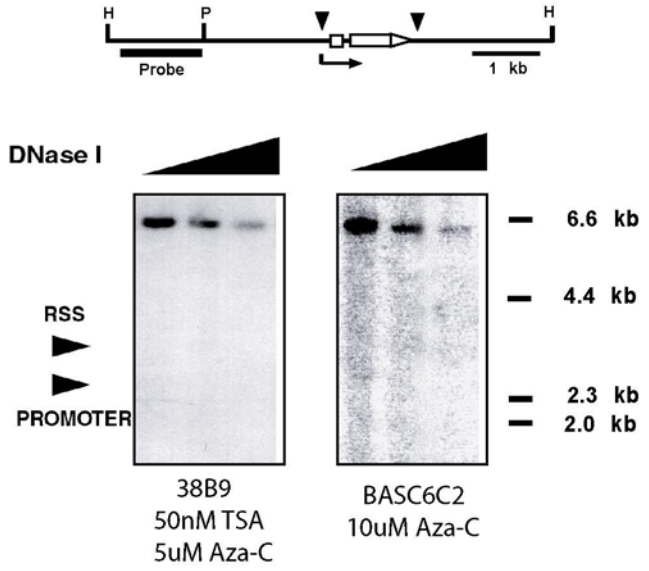

Figure 3.3 – DNase I HS_P and HS_{RSS} are not induced in pro-B cells by TSA or 5-AzaC. Cells were treated as mentioned in the methods with these drugs. Chromatin was treated with DNase I as before and DNA was isolated and subjected to Southern analysis for HS at Vκ2-139 as above.

Because of the importance of RSS in the joining reaction, I wanted to determine whether the Jκ RSS formed DNase I HS. I was unable to detect any HS_{RSS} in the Jκ regions of any cell lines tested (Figure 3.4).

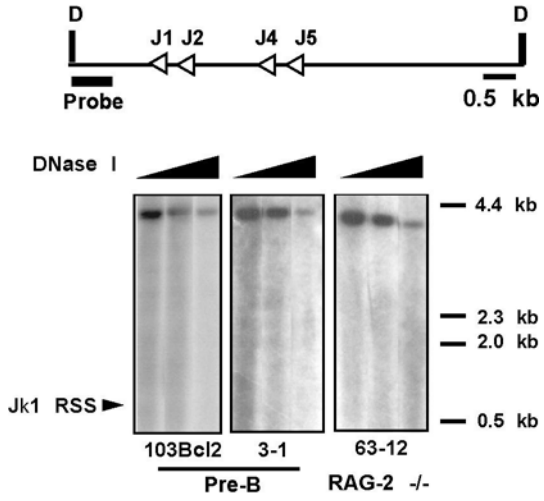

Figure 3.4. HSs are not found at Jκ gene segment RSSs. The indicated cell lines were treated with DNase I as above, and purified DNA was digested with *Dra*I for Southern hybridization using a probe adjacent to the 3' *Dra*I cut site for indirect end-labeling.

Transcription of Vκ segments is not correlated with the appearance of either HS_{RSS} or HS_P. Germline transcription has been well correlated with the onset of rearrangement of Ig and TCR loci (Schlissel and Baltimore, 1989; Schlissel et al., 1991b). It has been postulated that this transcription is either the mechanism by which cells open their chromatin, or the way that it remains in an accessible state for the recombinase to bind and cleave the RSS. I tested the correlation of germline transcription with the presence of HS_P and HS_{RSS} (Figure 3.6). Interestingly, I found there were significant differences in "accessibility" as measured by semi-quantitative RT-PCR versus DNase I HS mapping.

The first observation that can be made is that there are cell line specific differences in the transcription of different Vκ gene segments. For example Vκ2-139, but neither of the Vκ1 genes, is transcribed in 103Bcl2 cells (Fig. 3.5), but both gene segments possess HS_P (Fig. 3.1 and data not shown). Secondly, there is robust transcription from alleles that have

no detectable HS, such as Vκ1 genes in BASC6 C2 and 38B9 pro-B cell lines (Figs 3.1, 3.5, and data not shown). Finally, there is significant LPS-inducibility of transcription from each of the alleles tested in the 38B9 cell line (Figure 3.5). However, LPS treatment was unable to induce either HS (data not shown). These observations suggest that germline transcription and HS formation have distinct mechanisms for forming accessible chromatin structures.

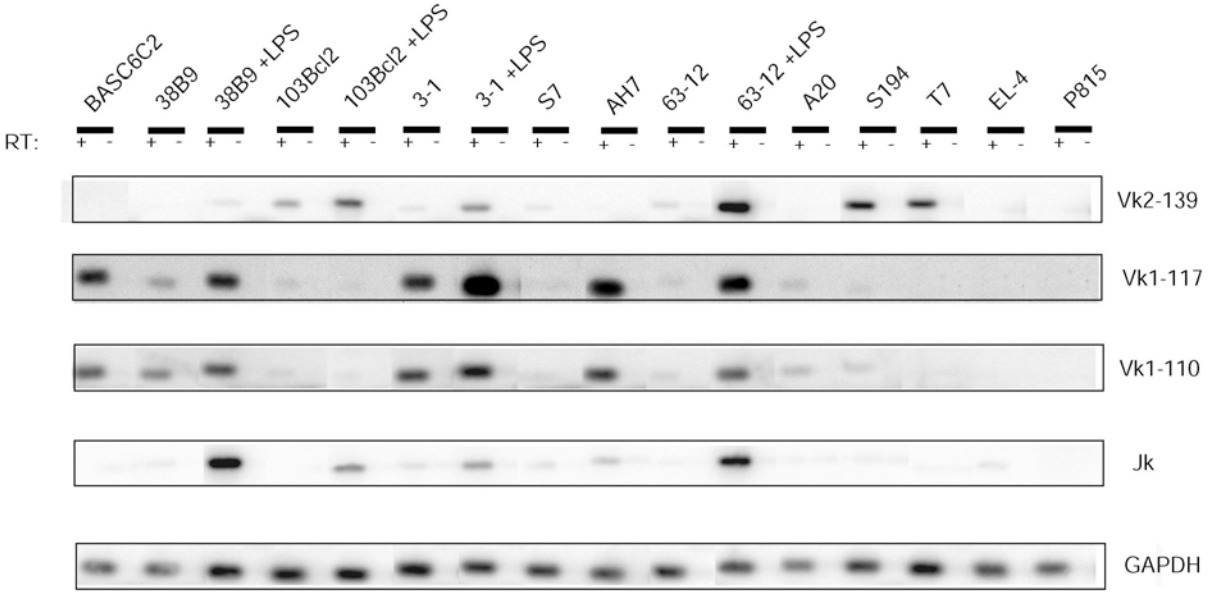

Figure 3.5 - Transcription of germline Vκ genes and Jκ gene segments. Cell lines represent a variety of stages of B cell development (See Table 3.2). +/- refers to whether total RNA was treated with reverse transcriptase before PCR amplification. For Vκ genes, one primer was in the coding exon and the other downstream of the RSS to ensure products were generated only from unrearranged alleles.

Analysis of Rag-1 and Rag-2 expression in cell lines by RT-PCR. The cell lines used in this study were examined by RT-PCR to confirm the absence or presence of transcripts for the V(D)J recombinase (Figure 3.6). The expression pattern was as would be expected, with no detectable expression in more mature cell lines that have ceased the recombination

process (A20, S194, and EL-4). The expression of Rag-1 and Rag-2 in the pro-B cell lines further supports the idea that the recombinase is not responsible for generating HS_P and HS_{RSS}.

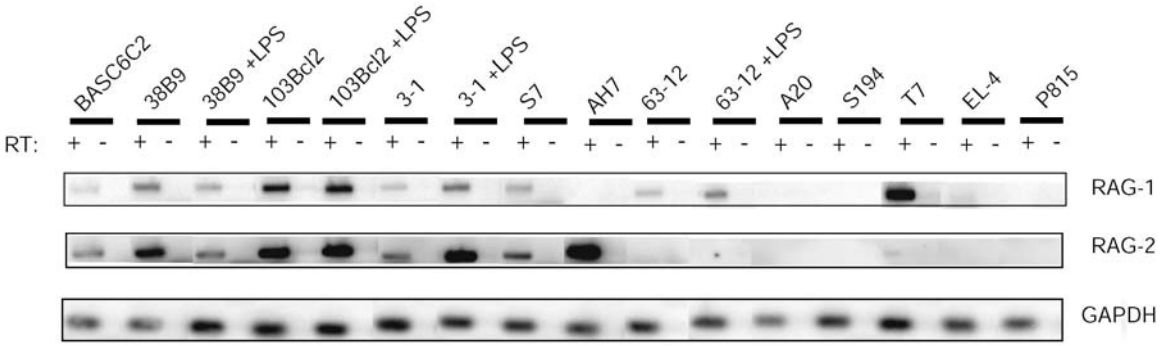

Figure 3.6 - Transcription of Rag1/2 in selected cell lines. Cell lines represent a variety of stages of B cell development (See Table 3.2). +/- refers to whether total RNA was treated with reverse transcriptase before PCR amplification.

Discussion

In this chapter I have used the data that I collected in Chapter Two to begin a study of the accessibility of the chromatin of Vκ regions in cell lines. For this purpose, I measured accessibility both in DNase I hypersensitivity and germline transcription from the regions of interest. I have found that HSs form in Vκ promoters at the stage in B cell development immediately preceding their usage in recombination. These HS_P disappear at later stages of development. In some families of Vκ genes, there are HS_{RSS} present prior to the V-J joining event as well.

All V(D)J recombination uses a common set of protein machinery; therefore a locus targeting factor must be involved in the generation of HS because IgH, Igλ and TCR loci are not chromatin accessible at the pre-B cell stage of development. This can be seen by the absence of HS in the T7 cell line (Fig. 3.1) as well as the absence of HS at the Igλ locus in cell lines with unrearranged Igκ loci (Hagman, 1989). This specificity lead me to hypothesize that determining the factors responsible for generating these HS will allow determination of the factors controlling targeting of the Igκ locus for V-J joining. These sites are well correlated with the presence of putative transcription factor binding sites. In the case of the Vκ1 and Vκ2 families, there are multiple E2A sites in the vicinity immediately downstream of the RSS (Figure 2.5). I have also shown that these sites are not inducible by LPS, HDAC inhibitors, or demethylation (Figure 3.3). RAG-1 and –2 are also not responsible for generating these sites (Figure 3.2).

There was no detectable HS_{RSS} in the Vκ1 genes in any of the cell lines tested although these two V regions are highly expressed in mice as well as used frequently in cell line rearrangements. Intriguingly, the conserved promoter region in these V gene segments lacks a perfect consensus match to the octamer binding sequence, but these two genes have strong HS_P. This suggests that other factors are responsible for targeting the recombinase and/or transcription machinery to the genes for expression and rearrangement.

Germline transcription has been implicated in generating chromatin accessibility to the recombinase (Casellas et al., 2002; Delpy et al., 2003; Goebel et al., 2001). However, these results give direct evidence that germline transcription is neither necessary nor sufficient to maintain localized HS_{RSS} or HS_P. Other studies have supported this observation as well in both mouse and human Ig genes (Angelin-Duclos and Calame, 1998; Casellas et al., 2002; Delpy et al., 2003; Goebel et al., 2001). This is in agreement with other recent studies that have disconnected germline transcription and V(D)J recombination (Angelin-Duclos and Calame, 1998; Delpy et al., 2003; Goebel et al., 2001). It is possible that the initial formation of accessible chromatin requires germline transcription and that is a brief phenomenon that is not required to maintain this "open" state, since it has been observed that HS can be propagated through DNA replication once established (Groudine and Weintraub, 1982).

Another novel aspect of these results is that although RAG-1 and –2 knockouts arrest lymphocyte development at the pro-B stage, prior to any V(D)J recombination (Roman et al., 1997; Shinkai et al., 1992), the chromatin phenotype of these Abelson-virus transformed lines is identical to the pre-B cell stage. This suggests a developmental pathway separate

from V(D)J recombination status exists to open the Vκ chromatin. It may be a type of "SOS" response, a last chance effort to initiate V-J joining, exhibited at the level of chromatin structure. There are other examples of alterations of Ig chromatin structure in the absence of RAG protein expression in the literature. During normal B cell development, Ig genes move from heterochromatin in the nuclear periphery to the center of the nucleus, prior to V(D)J recombination. This movement occurs normally in pro-B cells expanded from the bone marrow of RAG knockout mice as well (Kosak et al., 2002).

The lack of HS at the Jκ regions suggests three possibilities: 1) formation of Jκ HS_{RSS} is the rate-limiting step in V-J joining – this is consistent with the finding that RAG1/2 first assemble on 12 bp spacers, like those of Vκ genes (Jones and Gellert, 2002); 2) Jκ HS_{RSS} occurs in a minority of cells such that this assay is not sensitive enough to detect them; and 3) the downstream region of the locus already has an extremely altered chromatin structure and is too open to detect DNase I HSs (Goldmit et al., 2002).

Other recent studies have found that the RSSs of Vκ genes are not accessible to the RAG proteins in *in vitro* assembled chromatin or restriction endonucleases in both *in vitro* assembled chromatin and pre-B cells. (Baumann et al., 2003b; Golding et al., 1999). The study by Baumann *et al.* found that nucleosomes are precisely positioned over the RSS *in vitro*. This positioning is mediated by the nonamer sequence. They also found that the sequence of the nonamer was crucial for this positioning. Consensus nonamers positioned nucleosomes preferentially, while the non-consensus sequences are not protected from restriction enzyme cleavage. The Vκ4, 9 and 1 genes used in early rearrangement events all have consensus nonamer sequences in their RSS. This positioning may be another aspect of

control for rearrangement, or may not actually exist *in vivo*. The fact that the RSS are resistant to restriction endonuclease cutting but sensitive to DNase I digestion could be attributed to two possibilities. First, it is possible that the HS_{RSS} does not extend into the RSS, but reside nearby it. In this case HS_{RSS} may serve as an entry site for the recombination machinery to the adjacent RSS. Secondly, both RAG proteins and DNase I target DNA in the minor groove. Of the restriction enzymes used in Baumann *et al*. study, the enzymes with known structures for binding, bind the DNA in the major groove. Differential accessibility to the major and minor grooves could explain these differences in results.

Two other studies have found that nucleosomes completely inhibit cleavage by the RAG proteins *in vitro*. Neither hyperacetylated histones, nor HMG-1 were sufficient to overcome this inhibition (Golding et al., 1999; Kwon et al., 1998). This suggests that other proteins are required to overcome the repression caused by nucleosome structure. Other studies using the more *in vivo* approach of isolated nuclei, have observed that the RSS of Jκ genes become accessible in the pre-B cell stage of development (Stanhope-Baker et al., 1996). In the isolated nuclei, the Jκ RSS from pro-B cells are inaccessible to cleavage by the RAG proteins, but accessible in pre-B cells. This suggests that there are other factors present in these nuclei modulating the cleavage. Histone acetylation does play a role in this. In *in vitro* assays, the use of acetylated histones does not alter the ability of the nuclease to cleave the DNA, but *in vivo* histone acetylation stimulates recombination in a cell and stage specific manner (McBlane and Boyes, 2000). My observation of HS near the RSS of recombining genes provides an assay to determine the proteins in question. The factors that generate

HS_{RSS} could recruit the histone remodeling complexes that lead to this differential accessibility.

CHAPTER FOUR

Discussion

In summary, I have assembled the sequence of the mouse Igκ locus and used this sequence to draw novel conclusions about the regulation of the locus. This sequence assembly has allowed me to make a number of novel and important observations and predictions concerning the regulation of the locus. These observations include the analysis of transcription factor binding sites throughout the Igκ locus. The importance of these sites has also been linked to *in vivo* function by the observation of DNase I HS near the promoter and RSSs of Vκ genes in at least three positions spanning the locus. These HS appear before V-J joining, and disappear after the locus has rearranged. The presence of these sites is not coupled to the germline transcription of each of the V genes. The correlations between the potential binding sites and chromatin structure have been studied for two Vκ1 genes and a Vκ2 gene.

The observation of E boxes near the RSS of nearly half of the Vκ genes is quite striking, since the loss of E2A in mice leads to a complete block in B cell development (Bain et al., 1994; Zhuang et al., 1994). The loss of only one of the splice variants, E47, leads to a block at the pre-B cell stage, prior to V-J joining. However, loss of only E12 allows B cells to proceed completely through development (Bain et al., 1997). However, at the current time, the three enhancers in the downstream region are the only known functional binding sites in the Igκ locus. Clusters of sites downstream of most RSS provide a base for intriguing future studies on their importance. In addition to their importance in B cell development, E

proteins have also been implicated in modulating chromatin structure. They recruit histone acetylases and chromatin remodeling complexes to their binding sites (Bradney et al., 2003; Eckner et al., 1996; Massari et al., 1999; Qiu et al., 1998).

Octamer sites present downstream of the RSS of the members of the Vκ4 family are quite intriguing. Recently, it has been observed that mice with a knockout of the OcaB gene, an Oct1/2 coactivator, leads to a striking alteration in the usage of families for recombination (Casellas et al., 2002). However, in the preceding study, the Vκ4 family was still used at a high frequency. Below, I propose a model in which Vκ4 is targeted for the initial rearrangement events. The finding that Vκ4 rearrangement was not affected may suggest that OcaB is not required for the activity resulting from binding to the downstream octamer site. OcaB has been shown to provide specificity to the binding of the ubiquitous Oct-1 factor (Cepek et al., 1996). This may be important for transcription or other Oct-1 activities, but not for binding to the downstream site. Interestingly, E boxes are not observed downstream of the RSSs of the Vκ4 family members. This can be coupled with the observation that E2A and EBF activate different families for recombination in humans (Goebel et al., 2001) to lend support to a model that suggests accessibility of different V gene families is modulated by different factors.

An attractive model is to suggest a crucial role for the Vκ4 families for being central to the accessibility and recombination process. This family forms a core in the center of the locus. Although most families are interspersed with other families, the Vκ4 and Vκ21 families are exceptions to this. Vκ21 genes are the family closest to the Jκ array and are all found in the forward transcriptional orientation relative to Cκ. Vκ4 genes, on the other hand,

are all in the reverse orientation relative to Cκ. This is also the largest family of genes by far. Recombination of a Vκ4 gene takes place by inverting the sequences upstream of the gene selected, bringing them closer to the Jκ array for repeated recombination events. I suggest that the unique binding sites surrounding the Vκ4 genes are present to target the recombinase first to this central region of the locus. I would suggest that this region would become accessible before the other Vκ regions. Support for this observation comes from the observation that Vκ4 family members are consistently rearranged preferentially at the earliest detectable stages in fetal liver (Ramsden et al., 1994b), as well as in cell lines (Kalled and Brodeur, 1990; Kaushik et al., 1989).

In the 1994 Ramsden study, V regions were isolated from day 14 and day 16 fetal livers and the changes in repertoire were studied. The day 14 clones were the earliest rearrangements that were detectable from the Igκ locus. Studies such as this, in the absence of selection, provide an ideal system for looking at the preferences for V gene usage based purely on the DNA sequences or location of the gene. Of the day 14 rearrangements that were cloned, only 33% were productive. Of these productive rearrangements, I was able to look at the sequences and determine the V gene of origin (sequences from non-productive alleles were not reported). 33 out of the 65 genes were from the Vκ4 family. This is more than 50% representation from a family that makes up approximately 30% of the locus. Out of these 33 Vκ4 genes, 22 of them are from the cluster of Vκ4 genes that contain the unique pattern of potential Lef-1, Ebf and Octamer sites (Figure 4.1). 20 of the remaining 32 genes in these analyses were products of rearrangements of two Vκ9 genes. These two Vκ9 genes, 9-96 and 9-94, are the first two genes that are 5' to the Vκ4 cluster that are in the reverse

orientation. The only other large group of Vκ genes identified in this study were two genes from the Vκ1 family, 1-110 and 1-117. These genes are both located toward the 5' end of the locus. They both have variant octamer sequences in their promoters and both have three conserved E-boxes in the regions downstream of the RSS (Figure 4.1). The Vκ1-110 and 1-117 genes are located in the forward orientation. They also contain HS_P during the pre-B cell stage of development. Further studies on transcription factor binding surrounding the Vκ1, 9 and 4 genes observed early in development will provide a basis to determine the importance of these potential regulatory elements surrounding V genes. This can then lead to a detailed analysis of the importance of these factors in the development of repertoire and changes in repertoire that occur in autoimmune disease states.

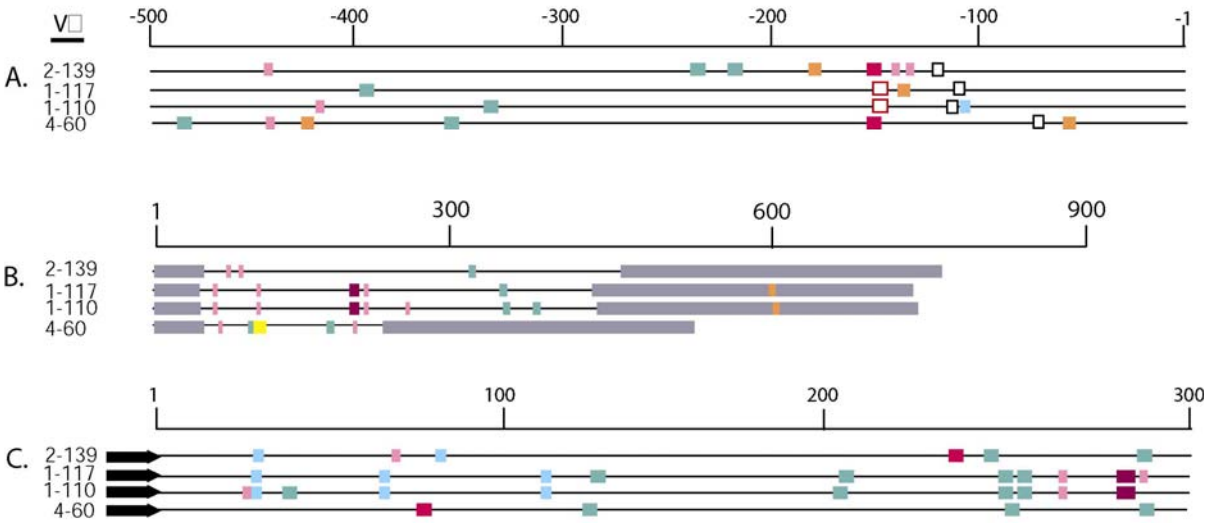

Figure 4.1 – Conserved binding sites from representative V genes. A. Promoter regions of V genes. B. Coding and intron regions of V genes. C. Downstream region of V genes. Scale for each segment is represented above the group. The two Vκ1 segments lack a conserved octamer sequence in the promoter region. The Vκ4 segment contains an octamer site in the region downstream of the RSS. The Vκ4 segment also does not contain any E boxes. Octamer/variant octamer – red / open

red, TATA – black outline, Ebf – yellow, HMG box – orange, E-box – blue, Ikaros – pink, Forkhead – green, PU.1 – black, NF-κB – maroon.

Chromatin immunoprecipitation experiments carried out by Zhe Liu in our laboratory have also supported this idea. These studies have shown that there is enrichment for Oct-1 binding near the RSS of a Vκ4 gene, 4-60, specifically in pre-B cells. These results provide further evidence that combinatorial binding of transcription factors leads to accessibility of individual Vκ genes.

Further studies to test this hypothesis include the deletion of E boxes or Octamer sites in the Vκ genes studied here in mice. It would be expected that if these sites are crucial for regulating V-J joining, there would be a reduction in the level of recombination of the particular Vκ gene of interest. It would also be expected that the HS presumed to be associated with the binding of the factor of interest would disappear in the mutant B cells. Other possible experiments to test these ideas include overexpression of the factors in question in pro-B cells. Expression of E2A in conjunction with RAGs in non-B cells induces recombination (Goebel et al., 2001). Therefore, it would be expected that if the factor is altering accessibility, it would induce the appropriate HS at the Vκ gene in question. Another possible study would be to examine the sequences of the spacer of the RSS. Several intriguing studies have demonstrated that differences in recombination frequency can be attributed to variations in the spacer sequences (Fanning et al., 1996; Larijani et al., 1999; Montalbano et al., 2003; Nadel et al., 1998; Posnett et al., 1994; Ramsden et al., 1994a) None of these previous studies have examined the Igκ locus in mouse. Therefore, it would

be worthwhile to align the spacer sequences from the mouse V genes that are most used in initial rearrangement events.

REFERENCES

Agata, Y., Katakai, T., Ye, S.-K., Sugai, M., Gonda, H., Honjo, T., Ikuta, K., and Shimizu, A. (2001). Histone acetylation determines the developmentally regulated accessibility for T cell receptor γ gene recombination. J Exp Med *193*, 873-879.

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

Aidinis, V., Bonaldi, T., Beltrame, M., Santagata, S., Bianchi, M. E., and Spanopoulou, E. (1999). The RAG1 homeodomain recruits HMG1 and HMG2 to facilitate recombination signal sequence binding and to enhance the intrinsic DNA-bending activity of RAG1-RAG2. Mol Cell Biol *19*, 6532-6542.

Akira, S., Okazaki, K., and Sakano, H. (1987). Two pairs of recombination signals are sufficient to cause immunoglobulin V-(D)-J joining. Science *238*, 1134-1138.

Allen, E., Horvath, S., Tong, F., Kraft, P., Spiteri, E., Riggs, A. D., and Marahrens, Y. (2003). High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes. Proc Natl Acad Sci U S A *100*, 9940-9945.

Alt, F., Rosenberg, N., Lewis, S., Thomas, E., and Baltimore, D. (1981). Organization and reorganization of immunoglobulin genes in A-MuLV-transformed cells: rearrangement of heavy but not light chain genes. Cell *27*, 381-390.

Alt, F. W., Yancopoulos, G. D., Blackwell, T. K., Wood, C., Thomas, E., Boss, M., Coffman, R., Rosenberg, N., Tonegawa, S., and Baltimore, D. (1984). Ordered rearrangement of immunoglobulin heavy chain variable region segments. EMBO J *3*, 1209-1219.

Angelin-Duclos, C., and Calame, K. (1998). Evidence that immunoglobulin VH-DJ recombination does not require germ line transcription of the recombining variable gene segment. Mol Cell Biol *18*, 6253-6264.

Annweiler, A., Muller-Immergluck, M., and Wirth, T. (1992). Oct2 transactivation from a remote enhancer position requires a B-cell-restricted activity. Mol Cell Biol *12*, 3107-3116.

Apel, T. W., Scherer, A., Adachi, T., Auch, D., Ayane, M., and Reth, M. (1995). The ribose 5-phosphate isomerase encoding gene is located immediately downstream from that encoding murine immunoglobulin kappa. Gene *156*, 191-197.

Aranburu, A., Carlsson, R., Persson, C., Leanderson, T. (2001). Transcription factor AP-4 is a ligand for immunoglobulin κ promoter E-box elements. Biochem J *353*, 431-438.

Aronow, B. J., Ebert, C. A., Valerius, M. T., Potter, S. S., Wiginton, D. A., Witte, D. P., and Hutton, J. J. (1995). Dissecting a locus control region: facilitation of enhancer function by extended enhancer-flanking sequences. Mol Cell Biol *15*, 1123-1135.

Asenbauer, H., and Klobeck, H.-G. (1996). Tissue-specific deoxyribonuclease I hypersensitive sites in the vicinity of the immunoglobulin Cλ cluster of man. Eur J Immunol *26*, 142-150.

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

Atchison, M. L., and Perry, R. P. (1988). Complementation between two cell lines lacking κ enhancer activity: implications for the developmental control of immunogloubulin transcription. EMBO J *7*, 4213-4220.

Baeuerle, P. A., and Baltimore, D. (1988). I kappa B: a specific inhibitor of the NF-kappa B transcription factor. Science *242*, 540-546.

Bailin, T., Mo, X., and Sadofsky, M. J. (1999). A RAG1 and RAG2 tetramer complex is active in cleavage in V(D)J recombination. Mol Cell Biol *19*, 4664-4671.

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

Bain, G., Maandag, E. C. R., Riele, H. P. J. t., Feeney, A. J., Sheehy, A., Schlissel, M., Shinton, S. A., Hardy, R. R., and Murre, C. (1997). Both E12 and E47 allow commitment to the B cell lineage. Immunity *6*, 145-154.

Baldwin, A. S., Jr. (1996). The NF-kappa B and I kappa B proteins: new discoveries and insights. Annu Rev Immunol *14*, 649-683.

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

Bassing, C. H., Swat, W., and Alt, F. W. (2002). The mechanism and regulation of chromosomal V(D)J recombination. Cell *107*, S45-S55.

Bauer, S. R., and Scheuermann, R. H. (1993). Expression of the VpreB/λ5/μ pseudo-Ig complex correlates with downregulated RAG-1 expression and V(D)J type recombination: a mechanism for allelic exclusion at the IgH locus. Transgene *1*, 33-45.

Baumann, M., Mamais, A., McBlane, F., Xiao, H., and Boyes, J. (2003a). Regulation of V(D)J recombination by nucleosome positioning at recombination signal sequences. EMBO J *22*, 5197-5207.

Baumann, M., Mamais, A., McBlane, F., Xiao, H., and Boyes, J. (2003b). Regulation of V(D)J recombination by nucleosome positioning at recombination signal sequences. Embo J *22*, 5197-5207.

Beg, A. A., Sha, W. C., Bronson, R. T., Ghosh, S., and Baltimore, D. (1995). Embryonic lethality and liver degeneration in mice lacking the RelA component of NF-kappa B. Nature *376*, 167-170.

Behrens, J., von Kries, J. P., Kuhl, M., Bruhn, L., Wedlich, D., Grosschedl, R., and Birchmeier, W. (1996). Functional interaction of beta-catenin with the transcription factor LEF-1. Nature *382*, 638-642.

Bemark, M., Liberg, D., and Leanderson, T. (1998). Conserved sequence elements in K promoters from mice and humans: implications for transcriptional regulation and repertoire expression. Immunogenetics *47*, 183-195.

Bergman, Y., Rice, D., Grosschedl, R., and Baltimore, D. (1984). Two regulatory elements for immunoglobulin κ light chain gene expression. Proc Natl Acad Sci U S A *81*, 7041-8045.

Besmer, E., Mansilla-Soto, J., Cassard, S., Sawchuk, D. J., Brown, G., Sadofsky, M., Lewis, S. M., Nussenzweig, M. C., and Cortes, P. (1998). Hairpin coding end opening is mediated by RAG1 and RAG2 proteins. Mol Cell *2*, 817-828.

Blackwell, T. K., and Alt, F. W. (1989). Mechanism and developmental program of immunoglobulin gene rearrangement in mammals. Annu Rev Genet *23*, 605-636.

Blackwell, T. K., and Weintraub, H. (1990). Differences of similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science *250*, 1104-1110.

Blasquez, V. C., Xu, M., Moses, S. C., and Garrard, W. T. (1989). Immunoglobulin k gene expression after stable integration. J Biol Chem *264*, 21183-21189.

Bogue, M. A., Jhappan, C., and Roth, D. B. (1998). Analysis of variable (diversity) joining recombination in DNA-dependent protein kinase (DNA-PK)-deficient mice reveals DNA-PK-independent pathways for both signal and coding joint formation. Proc Natl Acad Sci U S A *95*, 15559-15564.

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

Bruhn, L., Munnerlyn, A., and Grosschedl, R. (1997). ALY, a context-dependent coactivator of LEF-1 and AML-1, is required for TCRalpha enhancer function. Genes Dev *11*, 640-653.

Bucchini, D., Reynaud, C. A., Ripoche, M. A., Grimal, H., Jami, J., and Weill, J. C. (1987). Rearrangement of a chicken immunoglobulin gene occurs in the lymphoid lineage of transgenic mice. Nature *326*, 409-411.

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. J Mol Biol *268*, 78-94.

Burkly, L., Hession, C., Ogata, L., Reilly, C., Marconi, L. A., Olson, D., Tizard, R., Cate, R., and Lo, D. (1995). Expression of relB is required for the development of thymic medulla and dendritic cells. Nature *373*, 531-536.

Burrows, P. D., and Cooper, M. D. (1997). B cell development and differentiation. Curr Op Immunol *9*, 239-244.

Casellas, R., Jankovic, M., Meyer, G., Gazumyan, A., Luo, Y., Roeder, R. G., and Nussenzweig, M. C. (2002). OcaB is required for normal transcription and V(D)J recombination of a subset of immunoglobulin κ genes. Cell *110*, 575-585.

Casellas, R., Shih, T.-A. Y., Kleinewietfeld, M., Rakonjac, J., Nemanzee, D., Rajewsky, K., and Nussenweig, M. C. (2001). Contribution of receptor editing to the antibody repertoire. Science *291*, 1541-1544.

Cepek, K. L., Chasman, D. I., and Sharp, P. A. (1996). Sequence-specific DNA binding of the B-cell-specific coactivator OCA-B. Genes Dev *10*, 2079-2088.

Chen, H. T., Bhandoola, A., Difilippantonio, M. J., Zhu, J., Brown, M. J., Tai, X., Rogakou, E. P., Brotz, T. M., Bonner, W. M., Ried, T., and Nussenzweig, A. (2000). Response to RAG-mediated V(D)J cleavage by NBS1 and γ-H2AX. Science *290*, 1962-1964.

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

Cherry, S. R., and Baltimore, D. (1999). Chromatin remodeling directly activates V(D)J recombination. Proc Natl Acad Sci U S A *96*, 10788-10793.

Choi, J. K., Shen, C.-P., Radomska, H. S., Eckhardt, L. A., and Kadesch, T. (1996). E47 activates the Ig-heavy chain and TdT loci in non-B cells. EMBO J *15*, 5014-5021.

Chung, S.-Y., Folsom, V., and Wooley, J. (1983). DNase I hypersensitive sites in the chromatin of immunoglobulin κ light chain genes. Proc Natl Acad Aci U S A *80*, 2427-2431.

Cirillo, L. A., and Zaret, K. S. (1999). An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. Mol Cell *4*, 961-969.

Clerc, R. G., Corcoran, L. M., LeBowitz, J. H., Baltimore, D., and Sharp, P. A. (1988). The B-cell-specific Oct-2 protein contains POU box- and homeo box-type domains. Genes Dev *2*, 1570-1581.

Cobb, B. S., Morales-Alcelay, S., Kleiger, G., Brown, K. E., Fisher, A. G., and Smale, S. T. (2000). Targeting of Ikaros to pericentromeric heterochromatin by direct DNA binding. Genes Dev *14*, 2146-2160.

Cockerill, P. N., and Garrard, W. T. (1986). Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. Cell *44*, 273-282.

Coleclough, C., Perry, R. P., Karjalainen, K., and Weigert, M. (1981). Aberrant rearrangements contribute significantly to the allelic exclusion of immunoglobulin gene expression. Nature *290*, 372-380.

Constantinescu, A., and Schlissel, M. S. (1997). Changes in locus-specific V(D)J recombinase activity induced by immunoglobulin gene products during B cell development. J Exp Med *185*, 609-620.

Corcoran, L. M., Karvelas, M., Nossal, G. J., Ye, Z. S., Jacks, T., and Baltimore, D. (1993). Oct-2, although not required for early B-cell development, is critical for later B-cell maturation and for postnatal survival. Genes Dev *7*, 570-582.

Crepieux, P., Coll, J., and Stehelin, D. (1994). The Ets family of proteins: weak modulators of gene expression in quest for transcriptional partners. Crit Rev Oncog *5*, 615-638.

Critchlow, S. E., Bowater, R. P., and Jackson, S. P. (1997). Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. Curr Biol *7*, 588-598.

Daitch, L. E., Morre, M. W., Persiani, D. M., Durdik, J. M., and Selsing, E. (1992). Transcription and recombination of the murine RS element. J Immunol *19*, 832-840.

Davis, R. L., Weintraub, H., and Lassar, A. B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987-1000.

Delpy, L., Le Bert, M., Cogne, M., and Khamlichi, A. A. (2003). Germ-line transcription occurs on both the functional and the non-functional alleles of immunoglobulin constant heavy chain genes. Eur J Immunol *33*, 2108-2113.

Doi, T. S., Takahashi, T., Taguchi, O., Azuma, T., and Obata, Y. (1997). NF-kappa B RelA-deficient lymphocytes: normal development of T cells and B cells, impaired production of IgA and IgG1 and reduced proliferative responses. J Exp Med *185*, 953-961.

Durdik, J., Moore, M. W., and Selsing, E. (1984). Novel k light-chain gene rearrangements in mouse l light chain -producing B lymphocytes. Nature *307*, 749-752.

Early, P., Huang, H., Davis, M., Calame, K., and Hood, L. (1980). An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: Vh, D, and Jh. Cell *19*, 981-992.

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

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

Ellis, J., Tan-Un, K. C., Harper, A., Michalovich, D., Yannoutsos, N., Philipsen, S., and Grosveld, F. (1996). A dominant chromatin-opening activity 5' hypersensitive site 3 of the human β-globin locus control region. EMBO J *15*, 562-568.

Emorine, L., Kuehl, M., Weir, L., Leder, P., and Max, E. E. (1983). A conserved sequence in the immunoglobulin Jκ–Cκ intron: possible enhancer element. Nature *304*, 447-449.

Ephrussi, A., Church, G. M., Tonegawa, S., and Gilbert, W. (1985). B lineage-specific interactions of an immunoglobulin enhancer with cellular factors in vivo. Science *227*, 134-140.

Ernst, P., Hahm, K., Trinh, L., Davis, J. N., Roussel, M. F., Turck, C. W., and Smale, S. T. (1996). A potential role for Elf-1 in terminal transferase gene regulation. Mol Cell Biol *16*, 6121-6131.

Falkner, F. G., and Zachau, H. G. (1984). Correct transcription of an immunoglobulin κ gene requires an upstream fragment containing conserved sequence elements. Nature *310*, 71-74.

Fanning, L., Connor, A., Baetz, K., Ramsden, D., and Wu, G. E. (1996). Mouse RSS spacer sequences affect the rate of V(D)J recombination. Immunogenetics *44*, 146-150.

Feddersen, R. M., Martin, D. J., and VanNess, B. G. (1990). The frequency of multiple recombination events occurring at the human IgK L chain locus. J Immunol *144*, 1088-1093.

Felsenfeld, G., Boyes, J., Chung, J., Clark, D., and Studitsky, V. (1996). Chromatin structure and gene expression. Proc Natl Acad Sci U S A *93*, 9384-9388.

Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. Nature *421*, 448-453.

Fernandez, L. A., Winkler, M., and Grosschedl, R. (2001). Matrix attachment region-dependent function of the immunoglobulin μ enhancer involves histone acetylation at a distance without changes in enhancer occupancy. Mol Cell Biol *21*, 196-208.

Ferradini, L., Gu, H., Smet, A. D., Rajewsky, K., Reynaud, C.-A., and Weillb, J.-C. (1996). Rearrangement-enhancing element upstream of the mouse immunoglobulin kappa chain J cluster. Science *271*, 1416-1420.

Ferrier, P., Krippl, B., Blackwell, T. K., Furley, A. J. W., Suh, H., Winoto, A., Cook, W. D., Hood, L., Costanini, F., and Alt, F. W. (1990). Separate elements control DJ and VDJ rearrangement in a transgenic recombination substrate. EMBO J *9*, 117-125.

Fugmann, S. D., Lee, A. I., Shockett, P. E., Villey, I. J., and Schatz, D. G. (2000). The RAG proteins and V(D)J recombination: complexes, ends, and transposition. Annu Rev Immunol *18*, 495-527.

Gallo, M. L., Pergola, F., Daniels, G. A., and Lieber, M. R. (1994). Distinct roles for RAG-1 in the initiation of V(D)J recombination and in the resolution of coding ends. J Biol Chem *269*, 22188-22192.

Gao, Y., Chaudhuri, J., Zhu, C., Davidson, L., Weaver, D. T., and Alt, F. W. (1998). A targeted DNA-PKcs-null mutation reveals DNA-PK-independent functions for KU in V(D)J recombination. Immunity *9*, 367-376.

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

George, J. B., Li, S., and Garrard, W. T. (1995). Yeast artificial chromosome contigs reveal that distal variable-region genes reside at least 3 megabases from the joining regions in the murine immunoglobulin κ locus. Proc Natl Acad Sci U S A *92*, 12421-12425.

George, J. B., Li, S., and Garrard, W. T. (1997). The Murine immunoglobulin kappa locus: new insights into genome organization and expression. In *The Antibodies*, J. D. Capra, and M. Zanetti, eds. (San Diego, Gordon and Breach/Harwood Academic Publisher), pp. 41-62.

Georgopoulos, K., Bigby, M., Wang, J.-H., Molnar, A., Wu, P., Winandy, S., and Sharpe, A. (1994). The Ikaros gene is required for the development of all lymphoid lineages. Cell *79*, 143-156.

Georgopoulos, K., Moore, D., and Derfler, B. (1992). Ikaros, an early lymphoid-specific transcription factor and a putative mediator for T cell commitment. Science *258*, 808-812.

Ghia, P., tenBoekel, E., Rolkink, A. G., and Melchers, F. (1998). B-cell development: a comparison between mouse and man. Immunology Today *19*, 480-485.

Giese, K., Kingsley, C., Kirshner, J. R., and Grosschedl, R. (1995). Assembly and function of a TCR alpha enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. Genes Dev *9*, 995-1008.

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

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

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

Goodhardt, M., Cavelier, P., Akimenko, M. A., Lutfalla, G., Babinet, C., and Rougeon, F. (1987). Rearrangement and expression of rabbit immunoglobulin kappa light chain gene in transgenic mice. Proc Natl Acad Sci U S A *84*, 4229-4233.

Graninger, W. B., Goldman, P. L., Morton, C. C., O'Brien, S. J., and Korsmeyer, S. J. (1988). The kappa-deleting element. Germline and rearranged, duplicated and dispersed forms. J Exp Med *167*, 488-501.

Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T. E., Mann, M., and Lieber, M. R. (1997). Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. Nature *388*, 492-495.

Grilli, M., Chiu, J. J., and Lenardo, M. J. (1993). NF-kappa B and Rel: participants in a multiform transcriptional regulatory system. Int Rev Cytol *143*, 1-62.

Gross, D. S., and Garrard, W. T. (1988). Poising chromatin for transcription. Poising chromatin for transcription *12*, 293-297.

Groudine, M., and Weintraub, H. (1982). Propagation of globin DNAase I-hypersensitive sites in absence of factors required for induction: a possible mechanism for determination. Cell *30*, 131-139.

Gstaiger, M., Knoepfel, L., Georgiev, O., Schaffner, W., and Hovens, C. M. (1995). A B-cell coactivator of octamer-binding transcription factors. Nature *373*, 360-362.

Gu, Y., Jin, S., Gao, Y., Weaver, D. T., and Alt, F. W. (1997). Ku70-deficient embryonic stem cells have increased ionizing radiosensitivity, defective DNA end-binding activity, and inability to support V(D)J recombination. Proc Natl Acad Sci U S A *94*, 8076-8081.

Hagman, J., Belanger, C., Travis, A., Turck, C. W., and Grosschedl, R. (1993). Cloning and functional characterization of early B cell factor, a regulator of lymphocyte specific gene expression. Genes Dev *7*, 760-773.

Hagman, J., Gutch, M. J., Lin, H., and Grosschedl, R. (1995). EBF contains a novel zinc coordination motif and multiple dimerization and transcriptional activation domains. EMBO J *14*, 2907-2916.

Hagman, J., Travis, A., and Grosschedl, R. (1991). A novel lineage-specific nuclear factor regulates mb-1 gene transcription at the early stages of B cell differentiation. EMBO J *10*, 3409-3417.

Hagman, J. R. (1989) Murine immunoglobulin lambda genes are regulated by downstream enhancer elements, University of Washington, Seattle.

Hahm, K., Cobb, B. S., McCarty, A. S., Brown, K. E., Klug, C. A., Lee, R., Akashi, K., Weissman, I. L., Fisher, A. G., and Smale, S. T. (1998). Helios, a T cell-restricted Ikaros family member that quantitatively associates with Ikaros at centromeric heterochromatin. Genes Dev *12*, 782-796.

Hahm, K., Ernst, P., Lo, K., Kim, G. S., Turck, C., and Smale, S. T. (1994). The lymphoid transcription factor LyF-1 is encoded by specific, alternatively spliced mRNAs derived from the Ikaros gene. Mol Cell Biol *14*, 7111-7123.

Haines, B. B., and Brodeur, P. H. (1998). Accessibility changes across the mouse IgH-V locus during B cell development. Eur J Immunol *28*, 4228-42358.

Hale, M. A., and Garrard, W. T. (1998). A targeted κ immunoglobulin gene containing a deletion of the nuclear matrix association region exhibits spontaneous hyper-recombination in pre-B cells. Mol Immunol *35*, 609-620.

Hampsey, M., and Reinberg, D. (2003). Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. Cell *113*, 429-432.

Hartley, K. O., Gell, D., Smith, G. C., Zhang, H., Divecha, N., Connelly, M. A., Admon, A., Lees-Miller, S. P., Anderson, C. W., and Jackson, S. P. (1995). DNA-dependent protein kinase catalytic subunit: a relative of phosphatidylinositol 3-kinase and the ataxia telangiectasia gene product. Cell *82*, 849-856.

Henderson, A., and Calame, K. (1998). Transcriptional regulation during B cell development. Ann Rev Immunol *16*, 163-200.

Hesse, J. E., Lieber, M. R., Mizuuchi, K., and Gellert, M. (1989). V(D)J recombination: a functional definition of the joining signals. Genes Dev *3*, 1053-1061.

Hiom, K., and Gellert, M. (1997). A stable RAG1-RAG2-DNA complex that is active in V(D)J cleavage. Cell *88*, 65-72.

Horibata, K., and Harris, A. W. (1970). Mouse myelomas and lymphomas in culture. Exp Cell Res *60*, 61-77.

Horwitz, B. H., Scott, M. L., Cherry, S. R., Bronson, R. T., and Baltimore, D. (1997). Failure of lymphopoiesis after adoptive transfer of NF-kappaB-deficient fetal liver cells. Immunity *6*, 765-772.

Hsieh, C.-L., and Lieber, M. R. (1992). CpG methylated minichromosomes become inaccessible for V(D)J recombination after undergoing replication. EMBO J *11*, 315-325.

Huang, J., Durum, S. K., and Muegge, K. (2001). Histone acetylation and recombination at the TCRγ locus follows IL-7 induction. J Immunol *167*, 6073-6077.

Huang, S.-Y., Barnard, M. B., Xu, M., Matsui, S.-I., Rose, S. M., and Garrard, W. T. (1986). The active immunoglobulin k chain gene is packaged by non-ubiquitin-conjugated nucleosomes. Proc Natl Acad Sci U S A *83*, 3738-3742.

Huber, O., Korn, R., McLaughlin, J., Ohsugi, M., Herrmann, B. G., and Kemler, R. (1996). Nuclear localization of beta-catenin by interaction with transcription factor LEF-1. Mech Dev *59*, 3-10.

Ji, Y., Zhang, J., Lee, A. I., Cedar, H., and Bergman, Y. (2003). A multistep mechanism for the activation of rearrangement in the immune system. Proc Natl Acad Sci U S A *100*, 7557-7562.

Johnson, K., Angelin-Duclos, C., Park, S., and Calame, K. L. (2003). Changes in histone acetylation are associated with differences in accessibility of V(H) gene segments to V-DJ recombination during B-cell ontogeny and development. Mol Cell Biol *23*, 2438-2450.

Jones, J. M., and Gellert, M. (2002). Ordered assembly of the V(D)J synaptic complex ensures accurate recombination. EMBO J *21*, 4162-4171.

Kalled, S. L., and Brodeur, P. H. (1990). Preferential rearrangement of Vκ4 gene segments in pre-B cell lines. J Exp Med *172*, 559-566.

Kamps, W. A., and Cooper, M. D. (1982). Microenvironmental studies of pre-B and B cell development in human and mouse fetuses. J Immunol *129*, 526-531.

Karanjawala, Z. E., Adachi, N., Irvine, R. A., Oh, E. K., Shibata, D., Schwarz, K., Hsieh, C. L., and Lieber, M. R. (2002). The embryonic lethality in DNA ligase IV-deficient mice is rescued by deletion of Ku: implications for unifying the heterogeneous phenotypes of NHEJ mutants. DNA Repair (Amst) *1*, 1017-1026.

Karasuyama, H., Kudo, A., and Melchers, F. (1990). The proteins encoded by the VpreB and lambda 5 pre-B cell-specific genes can associate with each other and with mu heavy chain. J Exp Med *172*, 969-972.

Kaushik, A., Schulze, D. H., Bona, C., and Kelsoe, G. (1989). Murine Vκ gene expression does not follow the V_H paradigm. J Exp Med *169*, 1859-1864.

Kelley, D. E., Coleclough, C., and Perry, R. P. (1982). Functional significance and evolutionary development of the 5' terminal regions of immunoglobulin variable region genes. Cell *29*, 681-689.

Kim, J., Sif, S., Jones, B., Jackson, A., Koipally, J., Heller, E., Winandy, S., Viel, A., Sawyer, A., Ikeda, T.*, et al.* (1999). Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes. Immunity *10*, 345-355.

Kim, K. J., Kanellopoulos-Langevin, C., Merwin, R. M., Sachs, D. H., and Asofsky, R. (1979). Establishment and characterization of BALB/c lymphoma lines with B cell properties. J Immunol *122*, 549-554.

Kim, U., Qin, X.-J., Gong, S., Stevens, S., Luo, Y., Nussenzweig, M., and Roeder, R. G. (1996). The B-cell-specific transcription coactivator OCA-B/OBF-1/Bob-1 is essential for normal production of immunoglobulin isotypes. Nature *383*, 542-547.

Kirch, S. A., Rathburn, G. A., and Oettinger, M. A. (1998). Dual role of RAG2 in V(D)J recombination: catalysis and regulation of ordered Ig gene assembly. EMBO J *17*, 4881-4886.

Kirschbaum, T., Jaenichen, R., and Zachau, H. G. (1996). The mouse immunoglobulin kappa locus contains about 140 variable gene segments. Eur J Immunol *26*, 1613-1620.

Kirschbaum, T., Pourrajabi, S., Zöcher, I., Schwendinger, J., Heim, V., Röschenthaler, F., Kirschbaum, V., and Zachau, H. G. (1998). The 3' part of the immunoglobulin κ locus of the mouse. Eur J Immunol *28*, 1458-1466.

Kirschbaum, T., Röschenthaler, F., Bensch, A., Hölscher, B., Lautner-Rieske, A., Ohnrich, M., Pourrajabi, S., Schwendinger, J., Zöcher, I., and Zachau, H. G. (1999). The central part of the mouse immunoglobulin κ locus. Eur J Immunol *29*, 2057-2064.

Klemsz, M., McKercher, S., Celada, A., VanBeveren, C., and Maki, R. (1990). The macrophage and B cell-specific transcription factor PU.1 is related to the ets oncogene. Cell *61*, 113-124.

Klug, C. A., Gerety, S. J., Shah, P. C., Chen, Y.-Y., Rice, N. R., Rosenberg, N., and Singh, H. (1994). The v-abl tyrosine kinase negatively regulated NF-kB / Rel factors and blocks kappa gene transcription in pre-B lymphocytes. Genes Dev *8*, 678-687.

Klug, C. A., Morrison, S. J., Masek, M., Hahm, K., Smale, S. T., and Weissman, I. L. (1998). Hematopoietic stem cells and lymphoid progenitors express different Ikaros isoforms, and Ikaros is localized to heterochromatin in immature lymphocytes. Proc Natl Acad Sci U S A *95*, 657-662.

Kofler, R., Geley, S., Kofler, H., and Helmberg, A. (1992). Mouse variable region gene families: complexity, polymorphism, and use in non-autoimmune responses. Immunol Rev *128*, 5-21.

Koipally, J., Renold, A., Kim, J., and Georgopoulos, K. (1999). Repression by Ikaros and Aiolos is mediated through histone deacetylase complexes. EMBO J *18*, 3090-3100.

Komori, T., Okada, A., Stewart, V., and Alt, F. W. (1993). lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes. Science *260*, 1171-1174.

Köntgen, F., Grumont, R. J., Strasser, A., Metcalf, D., Li, R., Tarlinton, D., and Gerondakis, S. (1995). Mice lacking the c-rel proto-oncogene exhibit defects in lymphocyte proliferation, humoral immunity, and interleukin-2 expression. Genes Dev *9*, 1965-1977.

Kopp, E. B., and Ghosh, S. (1995). NF-kappa B and rel proteins in innate immunity. Adv Immunol *58*, 1-27.

Korinek, V., Barker, N., Willert, K., Molenaar, M., Roose, J., Wagenaar, G., Markman, M., Lamers, W., Destree, O., and Clevers, H. (1998). Two members of the Tcf family implicated in Wnt/beta-catenin signaling during embryogenesis in the mouse. Mol Cell Biol *18*, 1248-1256.

Kosak, S. T., Skok, J. A., Medina, K. L., Riblet, R., LeBeau, M. M., Fisher, A. G., and Singh, H. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. Science *296*, 158-162.

Kudo, A., and Melchers, F. (1987). A second gene, VpreB in the lambda 5 locus of the mouse, which appears to be selectively expressed in pre-B lymphocytes. EMBO J *6*, 2267-2272.

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

Kwon, J., Morshead, K. B., Guyon, J. R., Kingston, R. E., and Oettinger, M. A. (2000). Histone acetlyation and hSWI/SNF remodeling act in concert to stimulate V(D)J cleavage of nucleosomal DNA. Mol Cell *6*, 1039-1048.

Lansford, R., Okada, A., Chen, J., Oltz, E. M., Blackwell, T. K., Alt, F. W., and Rathbun, G. (1996). Mechanism and control of immunoglobulin gene rearrangement. In Molecular

Immunology, B. D. Hames, and D. M. Glover, eds. (Oxford, Oxford University Press), pp. 1-100.

Larijani, M., Yu, C. C., Golub, R., Lam, Q. L., and Wu, G. E. (1999). The role of components of recombination signal sequences in immunoglobulin gene segment usage: a V81x model. Nucleic Acids Res *27*, 2304-2309.

Lassar, A. B., Buskin, J. N., Lockshon, D., Davis, R. L., Apone, S., Hauschka, S. D., and Weintraub, H. (1989). MyoD is a sequence-specific DNA binding protein requiring a region of myc homology to bind to the muscle creatine kinase enhancer. Cell *58*, 823-831.

Lassar, A. B., Davis, R. L., Wright, W. E., Kadesch, T., Murre, C., Voronova, A., Baltimore, D., and Weintraub, H. (1991). Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. Cell *66*, 305-315.

Leclerq, L., Butkeraitis, P., and Reth, M. (1989). A novel germ-line Jκ transcript starting immediately upstream of Jκ1. Nucleic Acids Res *17*, 6809-6819?

Lefranc, M. P. (2001a). Nomenclature of the human immunoglobulin heavy (IGH) genes. Exp Clin Immunogenet *18*, 100-116.

Lefranc, M. P. (2001b). Nomenclature of the human immunoglobulin kappa (IGK) genes. Exp Clin Immunogenet *18*, 161-174.

Leiden, J. M. (1993). Transcriptional regulation of T cell receptor genes. Annu Rev Immunol *11*, 539-570.

Lenardo, M., Pierce, J. W., and Baltimore, D. (1987). Protein-binding sites in Ig gene enhancers determine transcriptional activity and inducibility. Science *236*, 1573-1577.

Leu, T. M. J., Eastman, Q. M., and Schatz, D. G. (1997). Coding joint formation in a cell-free V(D)J recombination system. Immunity *7*, 303-314.

Levitt, D., and Cooper, M. D. (1980). Mouse pre-B cells synthesize and secrete mu heavy chains but not light chains. Cell *19*, 617-625.

Li, Q., and Wrange, O. (1997). Assays for transcription factors access to nucleosomal DNA. Methods: A Companion to Methods in Enzymology *12*, 96-104.

Li, S., and Garrard, W. T. (2003). The kinetics of V-J joining throughout 3.5 megabases of the mouse Ig kappa locus fit a constrained diffusion model of nuclear organization. FEBS Lett *536*, 125-129.

Li, Z., Otevrel, T., Gao, Y., Cheng, H.-L., Seed, B., Stamato, T. D., Taccioli, G. E., and Alt, F. W. (1995). The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. Cell *83*, 1079-1089.

Liang, H.-E., Hsu, L.-Y., Cado, D., and Schlissel, M. S. (2004). Variegated Transcriptional Activation of the Immunoglobulin κ Locus in Pre-B Cells Contributes to the Allelic Exclusion of Light-Chain Expression. Cell *118*, 19-29.

Lichtenstein, M., Keini, G., Cedar, H., and Bergman, Y. (1994). B cell specific demethylation: a novel role for the intronic k chain enhancer sequence. Cell *76*, 913-923.

Lieber, M. R. (1992). The mechanism of V(D)J recombination: a balance of diversity, specificity and stability. Cell *70*, 873-976.

Lieber, M. R., Hesse, J. E., Mizuuchi, K., and Gellert, M. (1987). Developmental stage specificity of the lymphoid V(D)J recombination activity. Genes Dev *1*, 751-761.

Lin, H., and Grosschedl, R. (1995). Failure of B-cell differentiation in mice lacking the transcription factor EBF. Nature *376*, 263-267.

Liou, H. C., Sha, W. C., Scott, M. L., and Baltimore, D. (1994). Sequential induction of NF-kappa B/Rel family proteins during B-cell terminal differentiation. Mol Cell Biol *14*, 5349-5359.

Liu, X., Robinson, G. W., Gouilleux, F., Groner, B., and Hennighausen, L. (1995). Cloning and expression of Stat5 and an additional homologue (Stat5b) involved in prolactin signal transduction in mouse mammary tissue. Proc Natl Acad Sci U S A *92*, 8831-8835.

Liu, X., and VanNess, B. (1999). Gene targeting of the KI-KII sequence elements in a model pre-B cell line: effects on germline transcription and rearrangement on the κ locus. Mol Immunol *36*, 461-469.

Liu, Z.-m., George-Raizen, J. B., Li, S., Meyers, K. C., Chang, M. Y., and Garrard, W. T. (2002). Chromatin structural analyses of the mouse Igκ gene locus reveal new hypersensitive sites specifying a transcriptional silencer and enhancer. J Biol Chem *277*, 32640-32649.

Liu, Z. and Garrard, W.T. (2005). Long-range interactions between three transcriptional enhancers, active Vkappa gene promoters, and a 3' boundary sequence spanning 46 kilobases. Mol Cell Biol *25*, 3220-3231.

Loffert, D., Schaal, S., Ehlich, A., Hardy, R. R., Zou, Y.-R., Muller, W., and Rajewsky, K. (1994). Early B cell development in the mouse: insights from mutations introduced by gene targeting. Immunol Rev *137*, 135-153.

Lundak, R. L., and Raidt, D. J. (1973). Cellular immune resonse against tumor cells. Cell Immunol *9*, 60-66.

Luo, Y., and Roeder, R. G. (1995). Cloning, functional characterization, and mechanism of action of the B-cell-specific transcriptional coactivator OCA-B. Mol Cell Biol *15*, 4115-4124.

Ma, Y., Pannicke, U., Schwarz, K., and Lieber, M. R. (2002). Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. Cell *108*, 781-794.

Madisen, L., and Groudine, M. (1994). Identification of a locus control region in the immunoglobulin heavy-chain locus that deregulated c-myc expression in plasmacytoma and Burkitt's lymphoma cells. Genes Dev *8*, 2212-2226.

Madisen, L., Krumm, A., Hebbes, T. R., and Groudine, M. (1998). The immunoglobulin heavy chain locus control region increases histone acetylation along linked c-myc genes. Mol Cell Biol *18*, 6281-6292.

Maes, J., O'Neill, L. P., Cavelier, P., Turner, B. M., Rougeon, F., and Goodhardt, M. (2001). Chromatin remodeling at the Ig loci prior to V(D)J recombination. J Immunol *167*, 866-874.

Martin, D., and VanNess, B. (1990). Initiation and processing of two kappa immunoglobulin germline transcripts in mouse B cell. Mol Cell Biol *10*, 1950-1958.

Massari, M. E., Grant, P. A., Pray-Grant, M. G., Berger, S. L., Workman, J. L., and Murre, C. (1999). A conserved motif present in a class of helix-loop-helix proteins activates transcription by direct recruitment of the SAGA complex. Mol Cell *4*, 63-73.

Mather, E. L., and Perry, R. P. (1983). Methylation status and DNase I sensitivity of immunoglobulin genes: changes associated with rearrangement. Proc Natl Acad Sci U S A *80*, 4689-4693.

Mathieu, N., Hempel, W. M., Spicuglia, S., Verthuy, C., and Ferrier, P. (2000). Chromatin remodeling by the T cell receptor (TCR) - β gene enhancer during early T cell development: implications for the control of TCR - β locus recombination. J Exp Med *192*, 625-636.

Max, E. E., Maizel, J. V., Jr., and Leder, P. (1981). The nucleotide sequence of a 5.5-kilobase DNA segment containing the mouse kappa immunoglobulin J and C region genes. J Biol Chem *256*, 5116-5120.

McBlane, F., and Boyes, J. (2000). Stimulation of V(D)J recombination by histone acetylation. Curr Biol *10*, 483-486.

McBlane, J. F., Gent, D. C. v., Ramsden, D. A., Romeo, C., Cuomo, C. A., Gellert, M., and Oettinger, M. A. (1995). Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. Cell *83*, 387-395.

McElhinny, N. S. A., Snowden, C. M., McCarville, J., and Ramsden, D. A. (2000). Ku recruits the XRCC4-ligase IV complex to DNA ends. Mol Cell Biol *20*, 2996-3003.

McIntire, K. R., and Rouse, A. M. (1970). Mouse Ig light chains alteration of κ/λ ratio. Fed Proc *29*, 704.

McKercher, S. R., Torbett, B. E., Anderson, K. L., Henkel, G. W., Vestal, D. J., Baribault, H., Klemsz, M., Feeney, A. J., Wu, G. E., Paige, C. J., and Maki, R. A. (1996). Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities. EMBO J *16*, 5647-5658.

McMurry, M. T., Hernandez-Munain, C., Lauzurica, P., and Krangel, M. S. (1997). Enhancer control of local accessibility to V(D)J recombinase. Mol Cell Biol *147*, 4553-4561.

McMurry, M. T., and Krangel, M. S. (2000). A role for histone acetylation in the developmental regulation of VDJ recombination. Science *287*, 495-498.

Medina, C. A., and Teale, J. M. (1993). Restricted κ chain expression in early ontogeny: biased utilization of Vκ exons and preferential Vκ–Jκ recombination. J Exp Med *177*, 1317-1330.

Meffre, E., Casellas, R., and Nussenzweig, M. C. (2000). Antibody regulation of B cell development. Nat Immunol *1*, 379-385.

Melchers, F., Karasuyama, H., Haasner, D., Bauer, S., Kudo, A., Sakaguchi, N., Jameson, B., and Rolink, A. (1993). The surrogate light chain in B-cell development. Immunology Today.

Meyer, K. B., and Neuberger, M. S. (1989). The immunoglobulin k locus contains a second, stronger B cell specific enhancer which is located downstream of the constant region. EMBO J *8*, 1959-1964.

Miyamoto, S., Schmitt, M. J., and Verma, I. M. (1994). Qualitative changes in the subunit composition of kB binding complexes during murine B cell differentiation. Proc Natl Acad Sci U S A *91*, 5056-5060.

Modesti, M., Hesse, J. E., and Gellert, M. (1999). DNA binding of Xrcc4 protein is associated with V(D)J recombination but not with stimulation of DNA ligase IV activity. EMBO J *18*, 2008-2018.

Molnar, A., and Georgopoulos, K. (1994). The Ikaros gene encodes a family of functionally divers zinc finger DNA-binding proteins. Mol Cell Biol *14*, 8292-8303.

Montalbano, A., Ogwaro, K. M., Tang, A., Matthews, A. G., Larijani, M., Oettinger, M. A., and Feeney, A. J. (2003). V(D)J recombination frequencies can be profoundly affected by changes in the spacer sequence. J Immunol *171*, 5296-5304.

Moore, M. W., Durdik, J., Persiani, D. M., and Selsing, E. (1985). Deletions of k chain constant regiongenes in mouse l chain producing B cells involve intrachromosomal DNA recombinations similar to V-J joining. Proc Natl Acad Sci U S A *82*, 6211-6215.

Morgan, B., Sun, L., Avitahl, N., Andrikopoulos, K., Ikeda, T., Gonzales, E., Wu, P., Neben, S., and Georgopoulos, K. (1997). Aiolos, a lymphoid restricted transcription factor that interacts with Ikaros to regulate lymphocyte differentiation. EMBO J *16*, 2004-2013.

Moshous, D., Callebaut, I., deChasseval, R., Corneo, B., Cavazzana-Calvo, M., LeDeist, F., Tezcan, I., Sanal, O., Bertrand, Y., Noel Philippe, A. F., and deVillartay, J.-P. (2001). Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutation in human severe combined immune deficiency. Cell *105*, 177-186.

Mostoslavsky, R., Singh, N., Kirillov, A., Pelanda, R., Cedar, H., Chess, A., and Bergman, Y. (1998). κ chain monoallelic demethylation and the establishment of allelic exclusion. Genes Dev *12*, 1801-1811.

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

Muller, M. M., Ruppert, S., Schaffner, W., and Matthias, P. (1988). A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells. Nature *336*, 544-551.

Murre, C., McCaw, P. S., and Baltimore, D. (1989a). A New DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless, MyoD,* and *myc* proteins. Cell *56*, 777-783.

Murre, C., McCaw, P. S., Vassin, H., Caudy, M., Jan, L. Y., Jan, Y. N., Cabrera, C. V., Buskin, J. N., Hauschka, S. D., Lassar, A. B.*, et al.* (1989b). Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell *58*, 537-544.

Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., Kravitz, S. A., Mobarry, C. M., Reinert, K. H., Remington, K. A.*, et al.* (2000). A whole-genome assembly of Drosophila. Science *287*, 2196-2204.

Mymryk, J. S., Fryer, C. J., Jung, L. A., and Archer, T. K. (1997). Analysis of chromatin structure in vivo. Methods: A Companion to Methods in Enzymology *12*, 105-114.

Nadel, B., Cazenave, P.-A., and Sanchez, P. (1990). Murine lambda gene rearrangements: the stochastic model prevails over the ordered model. EMBO J *9*, 435-440.

Nadel, B., Tang, A., Escuro, G., Lugo, G., and Feeney, A. J. (1998). Sequence of the spacer in the recombination signal sequence affects V(D)J rearrangement frequency and correlates with nonrandom Vκ usage in vivo. J Exp Med *187*, 1495-1503.

Nan, X., Ng, H.-H., Johnson, C. A., Laherty, C. D., Turner, B. M., and Eisenman, R. N. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386-389.

Ng, H. H., Ciccone, D. N., Morshead, K. B., Oettinger, M. A., and Struhl, K. (2003). Lysine-79 of histone H3 is hypomethylated at silenced loci in yeast and mammalian cells: a potential mechanism for position-effect variegation. Proc Natl Acad Sci U S A *100*, 1820-1825.

Ng, K. H., Lavigueur, A., Ricard, L., Boivrette, M., Maclean, S., Cloutier, D., and Gibson, D. M. (1989). Characterization of allelic Vκ1 region genes in inbred strains of mice. J Immunol *143*, 638-648.

Nielsen, P. J., Georgiev, O., Lorenz, B., and Schaffner, W. (1996). B lymphocytes are impaired in mice lacking the transcriptional co-activator Bob1/OCA-B/OBF1. Eur J Immunol *26*, 3214-3218.

Nutt, S. L., Morrison, A. M., Dorfler, P., Rolink, A., and Busslinger, M. (1998). Identification of BSAP (Pax-5) target genes in early B-cell development by loss- and gain-of-function experiments. EMBO J *17*, 2319-2333.

Nutt, S. L., Urbanek, P., Rolink, A., and Busslinger, M. (1997). Essential functions of Pax5 (BSAP) in pro-B cell development: difference between fetal and adult B lymphopoiesis and reduced V-to-*DJ* recombination at the *IgH* locus. Genes Dev *11*, 476-491.

O'Brien, D. P., Oltz, E. M., and Ness, B. G. V. (1997). Coordinate transcription and V(D)J recombination of the kappa immunoglobulin light-chain locus: NF-κB-dependent and -independent pathways of activation. Mol Cell Biol *17*, 3477-3487.

Oettinger, M. A. (1992). Activation of V(D)J recombination by RAG1 and RAG2. Trends Genet *8,* 413-416.

Oettinger, M. A., Schatz, D. G., Gorka, C., and Baltimore, D. (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. Science *248*, 1517-1523.

Okabe, T., Watanabe, T., and Kudo, A. (1992). A pre-B and B cell specific DNA binding protein, EBB-1, which binds to the promoter of the VpreB1 gene. Eur J Immunol *22*, 37-43.

Oosterwegel, M., van de Wetering, M., Dooijes, D., Klomp, L., Winoto, A., Georgopoulos, K., Meijlink, F., and Clevers, H. (1991). Cloning of murine TCF-1, a T cell-specific transcription factor interacting with functional motifs in the CD3-epsilon and T cell receptor alpha enhancers. J Exp Med *173*, 1133-1142.

Papvasiliou, F., Jankovic, M., Gong, S., and Nussenzweig, M. C. (1997). Control of immunoglobulin gene rearrangements in developing B cells. Curr Op Immunol *9*, 233-238.

Papvasiliou, F., Jankovic, M., and Nussenweig, M. C. (1996). Surrogate or conventional light chains are required for membrane immunoglobulin mu to activate the precursor B cell transition. J Exp Med *184*, 2025-2030.

Parslow, T. G., Blair, D. L., Murphy, W. J., and Granner, D. K. (1984). Structure of the 5' ends of immunoglobulin genes: a novel conserved sequence. Proc Natl Acad Sci U S A *81*, 2650-2654.

Paull, T. T., and Gellert, M. (2000). A mechanistic basis for Mre11-directed DNA joining at microhomologies. Proc Natl Acad Sci U S A *97*, 6409-6414.

Paull, T. T., Rogakou, E. P., Yamazaki, V., Kirchgessner, C. U., Gellert, M., and Bonner, W. M. (2000). A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. Curr Biol *10*, 886-895.

Pergola, F., Zdzienicka, M. Z., and Lieber, M. R. (1993). V(D)J recombination in mammalian cell mutants defective in DNA double-strand break repair. Mol Cell Biol *13*, 3464-3471.

Perry, R. P., Kelley, D. E., Coleclough, C., and Kearney, J. F. (1981). Organization and expression of immunoglobulin genes in fetal liver hybridomas. Proc Natl Acad Sci U S A *78*, 247-251.

Persiani, D. M., and Selsing, E. (1989). DNase I sensitivity of immunoglobulin light chain genes in Abelson murine leukemia virus transformed pre-B cell lines. Nucl Acids Res *17*, 5339-5348.

Picard, D., and Schaffner, W. (1984). A lymphocyte-specific enhancer in the mouse immunoglobulin κ gene. Nature *307*, 81-82.

Pongubala, J. M. R., Nagulapalli, S., Klemsz, M. J., McKercher, S. R., Maki, R. A., and Atchison, M. L. (1992). PU.1 recruits a second nuclear factor to a site important for immunoglobulin k 3' enhancer activity. Mol Cell Biol *12*, 368-378.

Posnett, D. N., Vissinga, C. S., Pambuccian, C., Wei, S., Robinson, M. A., Kostyu, D., and Concannon, P. (1994). Level of human TCRBV3S1 (V beta 3) expression correlates with

allelic polymorphism in the spacer region of the recombination signal sequence. J Exp Med *179*, 1707-1711.

Qiu, Y., Sharma, A., and Stein, R. (1998). p300 mediates transcriptional stimulation by the basic helix-loop-helix activators of the insulin gene. Mol Cell Biol *18*, 2957-2964.

Queen, C., and Baltimore, D. (1983). Immunoglobulin gene transcription is activated by downstream sequence elements. Cell *33*, 741-748.

Ramsden, D. A., Baetz, K., and Wu, G. E. (1994a). Conservation of sequence in recombination signal sequence spacers. Nucleic Acids Res *22*, 1785-1796.

Ramsden, D. A., and Gellert, M. (1998). Ku protein stimulates DNA end joining by mammalian DNA ligases: a direct role for Ku in repair of DNA double-strand breaks. Embo J *17*, 609-614.

Ramsden, D. A., Paige, C. J., and Wu, G. E. (1994b). κ light chain rearrangement in mouse fetal liver. J Immunol *153*, 1150-1160.

Ramsden, D. A., Paull, T. T., and Gellert, M. (1997). Cell-free V(D)J recombination. Nature *388*, 488-491.

Ren, Z., Lin, P., Klintworth, G., Iwata, F., Munier, F., Schorderet, D., El-Matri, L., Theenkadara, V., Basti, S., Reddy, M., and Hejtmancik, J. (2002). Allelic and locus heterogeneity in autosomal recessive gelatinous drop-like corneal dystrophy. Hum Genet *110*, 568-577.

Reya, T., and Grosschedl, R. (1998). Transcriptional regulation of B-cell differentiation. Curr Op Immunol *10*, 158-165.

Reya, T., O'Riordan, M., Okamura, R., Devaney, E., Willert, K., Nusse, R., and Grosschedl, R. (2000). Wnt signaling regulates B lymphocyte proliferation through a LEF-1 dependent mechanism. Immunity *13*, 15-24.

Rolink, A., Andersson, J., Ghia, P., Grawunder, U., Haasner, D., Karasuyama, H., Boekel, E. t., Winkler, T. H., and Melchers, F. (1995). B cell development in mouse and man. The Immunologist *3/4*, 125-151.

Rolink, A., Haasner, D., Nishikawa, S., and Melchers, F. (1993). Changes in frequencies of clonable pre B cells during life in different lymphoid organs of mice. Blood *81*, 2290-2300.

Rolink, A., Kudo, A., and Melchers, F. (1991). Long-term proliferating early pre-B-cell lines and clones with the potential to develop to surface immunoglobulin-positive, mitogen-reactive B-cells in vitro and in vivo. Biochem Soc Trans *19*, 275-276.

Roman, C. A. J., Cherry, S. R., and Baltimore, D. (1997). Complementation of V(D)J recombination deficiency in RAG-1$^{-/-}$ B cells reveals a requirement for novel elements in the N-terminus of RAG-1. Immunity *7*, 13-24.

Romanow, W. J., Langerak, A. W., Goebel, P., Wolvers-Tettero, I. L. M., vanDongen, J. J. M., Feeney, A. J., and Murre, C. (2000). E2A and EBF act in synergy with the V(D)J recombinase to generate a diverse immunoglobulin repertoire in nonlymphoid cells. Mol Cell *5*, 343-353.

Roque, M. C., Smith, P. A., and Blasquez, V. C. (1996). A developmentally modulated chromatin structure at the mouse immunoglobulin kappa 3' enhancer. Mol Cell Biol *16*, 3138-3155.

Röschenthaler, F., Hameister, H., and Zachau, H. G. (2000). The 5' part of the mouse immunoglobulin κ locus as a continuously cloned structure. Eur J Immunol *30*, 3349-3354.

Rukumara, R., Araki, R., Fujimori, A., Tsutsumi, Y., Kurimasa, A., Li, G. C., Chen, D. C., Tatsumi, K., and Abe, M. (2000). Signal joint formation is also impaired in DNA-dependent protein kinase catalytic subunit knockout cells. J Immunol *165*, 3883-3889.

Sakaguchi, N., and Melchers, F. (1986). Lambda 5, a new light-chain-related locus selectively expressed in pre-B lymphocytes. Nature *324*, 579-582.

Sakano, H., Kurosawa, Y., Weigert, M., and Tonegawa, S. (1981). Identification and nucleotide sequence of a diversity DNA segment (D) of immunoglobulin heavy-chain genes. Nature *290*, 562-565.

Schable, K. F., Thiebe, R., Bench, A., Brensing-Kuppers, J., Heim, V., Kirschbaum, T., Mitlöhner, H., Ohnrich, M., Pourrajabi, S., Röschenthaler, F., and Zachau, H. G. (1999). Characteristics of the immunoglobulin Vκ genes, pseudogenes, relics and orphons in the mouse genome. Eur J Immunol *29*, 2082.

Schatz, D. G., Oettinger, M. A., and Baltimore, D. (1989). The V(D)J recombination activating gene, RAG-1. Cell *59*, 1035-1048.

Schebesta, M., Heavey, B., and Busslinger, M. (2002). Transcriptional control of B cell development. Curr Op Immunol *14*, 216-223.

Scheidereit, C., Cromlish, J. A., Gerster, T., Kawakami, K., Balmaceda, C. G., Currie, R. A., and Roeder, R. G. (1988). A human lymphoid-specific transcription factor that activates immunoglobulin genes is a homoeobox protein. Nature *336*, 551-557.

Scherer, D. C., Brockman, J. A., Bendall, H. H., Zhand, G. M., Ballard, D. W., and Oltz, E. M. (1996). Corepression of RelA and c-rel inhibits immunoglobulin kappa gene transcription and rearrangement in precursor B lymphocytes. Immunity *5*, 563-574.

Schilham, M. W., and Clevers, H. (1998). HMG box containing transcription factors in lymphocyte differentiation. Sem Immunol *10*, 127-132.

Schlissel, M., Voronova, A., and Baltimore, D. (1991a). Helix-loop-helix transcription factor E47 activates germline immunoglobulin heavy chain gene transcription and rearrangement in a pre-T cell line. Genes Dev *5*, 1367-1376.

Schlissel, M. S. (2002). Does Artemis end the hunt for the hairpin-opening activity in V(D)J recombination? Cell *109*, 1-4.

Schlissel, M. S., and Baltimore, D. (1989). Activation of immunoglobulin kappa gene rearrangement correlates with induction of germline kappa gene transcription. Cell *58*, 1001-1007.

Schlissel, M. S., Corcoran, L. M., and Baltimore, D. (1991b). Virus-transformed pre-B cells show ordered activation but not inactivation of immunoglobulin gene rearrangement and transcription. J Exp Med *173*, 711-720.

Schubart, D. B., Rolink, A., Kosco-Vilbois, M. H., Botteri, F., and Matthias, P. (1996). B-cell-specific coactivator OBF-1/OCA-B/Bob1 required for immune response and germinal centre formation. Nature *383*, 538-542.

Schubart, K., Massa, S., Schubart, D., Corcoran, L. M., Rolink, A. G., and Matthias, P. (2001). B cell development and immunoglobulin gene transcription in the absence of Oct-2 and OBF-1. Nat Immunol *2*, 69-74.

Schupp, I. W., Schlake, T., Kirshbaum, T., Zachau, H. G., and Boehm, T. (1997). A yeast artificial chromosome contig spanning the mouse immunoglobulin kappa light chain locus. Immunogenetics *45*, 180-187.

Schwarzenback, H., Newell, J. W., and Matthias, P. (1995). Involvement of the Ets family factor PU.1 in the activation of immunoglobulin promoters. J Biol Chem *270*, 898-907.

Scott, E. W., Simon, M. C., Anastasi, J., and Singh, H. (1994). Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. Science *265*, 1573-1577.

Sen, R., and Baltimore, D. (1986a). Inducibility of kappa immunoglobulin enhancer-binding protein Nf-kappa B by a posttranslational mechanism. Cell *47*, 921-928.

Sen, R., and Baltimore, D. (1986b). Multiple nuclear factors interact with the immunoglobulin enhancer sequences. Cell *46*, 705-716.

Sha, W. C., Liou, H.-C., Tuomanen, E. I., and Baltimore, D. (1995). Targeted disruption of the p50 subunit of NF-kB leads to multifocal defects in immune response. Cell *80*, 321-330.

Shaffer, A. L., Peng, A., and Schlissel, M. S. (1997). In vivo occupancy of the κ light chain enhancers in primary pro- and pre-B cells: a model for κ locus activation. Immunity *6*, 131-143.

Shapiro, M. A., and Weigert, M. (1987). How immunoglobulin Vk genes rearrange. J Immunol *139*, 3834-3839.

Shen, C.-P., and Kadesch, T. (1995). B-cell-specific DNA binding by an E47 homodimer. Mol Cell Biol *15*, 4518-4524.

Shevach, E. M., Stobo, J. D., and Green, I. (1972). Immunoglobulin and θ-bearing murine leukemias and lymphomas. J Immunol *108*, 1146-1151.

Shimizu, T., Iwasato, T., and Yamagishi, H. (1991). Deletions of Immunoglobulin Cκ region characterized by the circular excision products in mouse splenocytes. J Immunol *173*, 1065-1107.

Shinkai, Y., Rathbun, G., Lam, K.-P., Oltz, E. M., Stewart, V., Mendelsohn, M., Charron, J., Datta, M., and Young, F. (1992). RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. Cell *68*, 855-867.

Shirakata, M., Huppi, K., Usada, S., Okazaki, K., Yoshida, K., and Sakano, H. (1991). HMG1-related DNA-binding protein isolated with V(D)J recombination signal probes. Mol Cell Biol *11*, 4528-4536.

Siden, E., Alt, F. W., Shinefeld, L., Sato, V., and Baltimore, D. (1981). Synthesis of immunoglobulin mu chain gene products precedes synthesis of light chains during B-lymphocyte development. Proc Natl Acad Sci U S A *78*, 1823-1827.

Sigvardsson, M., Akerblad, P., and Leanderson, T. (1996). Early B cell factor interacts with a subset of κ promoters. J Immunol *156*, 3788-3796.

Sigvardsson, M., Bemark, M., and Leanderson, T. (1995). Stimulation of kappa transcription by a decamer-dependent, synergistic mechanism. Eur J Immunol *25*, 298-301.

Siminovitch, K. A., Bakhshi, A., Goldman, P., and Korsmeyer, S. J. (1985). A uniform deleting element mediate the loss of κ genes in human B cells. Nature *316*, 260-262.

Singh, N., Bergman, Y., Cedar, H., and Chess, A. (2003). Biallelic germline transcription at the κ immunoglobulin locus. J Exp Med *197*, 743-750.

Sleckman, B. P., Gorman, J. R., and Alt, F. W. (1996). Accessibility control of antigen-receptor variable-region gene assembly: role of *cis*-acting elements. Ann Rev Immunol *14*, 459-481.

Sloan, S. R., Shen, C.-P., McCarrick-Walmsley, R., and Kadesch, T. (1996). Phosphorylation of E47 as a potential determinant of B-cell-specific activity. Mol Cell Biol *16*, 6900-6908.

Souabni, A., Cobaleda, C., Schebesta, M., and Busslinger, M. (2002). Pax5 promotes B lymphopoiesis and blocks T cell development by repressing Notch1. Immunity *17*, 781-793.

Sperry, A. O., Blasquez, V. C., and Garrard, W. T. (1989). Dysfunction of chromosomal loop attachment sites: illegitimate recombination linked to matrix association regions and topoisomerase II. Proc Natl Acad Sci U S A *86*, 5498-5501.

Spiegelberg, H. L. (1974). Biological activities of immunoglobulins of different classes and subclasses. Adv Immunol *19*, 259-294.

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

Storb, U., Wilson, R., Selsing, E., and Walfield, A. (1981). Rearranged and germline immunoglobulin k genes: different states of DNase I sensitivity of constant k genes in immunocompetent and nonimmune cells. Biochemistry *20*, 990-996.

Strubin, M., Newell, J. W., and Matthias, P. (1995). OBF-1, a novel B cell-specific cofactor that stimulates immunoglobulin promoter activity through association with octamer-binding proteins. Cell *80*, 497-506.

Sturm, R. A., Das, G., and Herr, W. (1988). The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain. Genes Dev *2*, 1582-1599.

Sun, L., Liu, A., and Georgopoulos, K. (1996). Zinc finger-mediated protein interactions modulate Ikaros activity, a molecular control of lymphocyte development. EMBO J *15*, 5358-5369.

Swanson, P. C., and Desiderio, S. (1998). V(D)J recombination signal recognition: distinct, overlapping DNA-protein contacts in complexes containing RAG1 with and without RAG2. Immunity *9*, 115-125.

Taccioli, G. E., Amatucci, A. G., Beamish, H. J., Gell, D., Xiang, X. H., Arzayus, M. I. T., Priestley, A., Jackson, S. P., Rothstein, A. M., Jeggo, P. A., and Herrera, V. L. M. (1998). Targeted disruption of the catalytic subunit of the DNA-PK gene in mice confers severe combined immunodeficiency and radiosensitivity. Immunity *9*, 355-366.

Taccioli, G. E., Rathbun, G., Oltz, E., Stamato, T., Jeggo, P. A., and Alt, F. W. (1993). Impairment of V(D)J recombination in double-strand break repair mutants. Science *260*, 207-210.

Takeda, S., Zou, Y.-R., Bluethmann, H., Kitamura, D., Muller, U., and Rajewsky, K. (1993). Deletion of the immunoglobulin κ chain intron enhancer abolishes κ chain gene rearrangement in cis but not λ chain gene rearrangement in trans. EMBO J *12*, 2329-2336.

ten Boekel, E., Melchers, F., and Rolink, A. (1995). The status of Ig loci rearrangements in single cells from different stages of B cell development. Int Immunol *7*, 1013-1019.

Thiebe, R., Schable, K. F., Bensch, A., Brensing-Kupers, J., Heim, V., Kirchbaum, T., Mitlohner, H., Ohnrich, M., Pourrajabi, S., Roschenthaler, F*., et al.* (1999). The variable genes and gene families of the mouse immunoglobulin κ locus. Eur J Immunol *29*, 2072-2081.

Tian, J., Okabe, T., Miyazaki, T., Takeshita, S., and Kudo, A. (1997). Pax-5 is identical to EBB-1/KLP and binds to the VpreB and lambda5 promoters as well as the KII and KII sites upstream of the Jkappa genes. Eur J Immunol *27,* 750-755

Tonegawa, S. (1983). Somatic generation of antibody diversity. Nature *302*, 575-581.

Travis, A., Amsterdam, A., Belanger, C., and Grosschedl, R. (1991). LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function [corrected]. Genes Dev *5*, 880-894.

Urbanek, P., Wang, Z.-Q., Getka, I., Wagner, E. F., and Busslinger, M. (1994). Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. Cell *79*, 901-912.

van de Wetering, M., Oosterwegel, M., Dooijes, D., and Clevers, H. (1991). Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box. EMBO J *10*, 123-132.

van Genderen, C., Okamura, R. M., Farinas, I., Quo, R. G., Parslow, T. G., Bruhn, L., and Grosschedl, R. (1994). Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice. Genes Dev *8*, 2691-2703.

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

vanGent, D. C., Ramsden, D. A., and Gellert, M. (1996). The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. Cell *85*, 107-113.

VanNess, B., Weigert, M., Coleclough, C., Mather, E. L., Kelle, D. E., and Perry, R. P. (1981). Transcription of the unrearranged mouse C kappa locus: sequence of the initiation region and comparison of activity with a rearranged V kappa - C kappa gene. Cell *27*, 593-602.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A.*, et al.* (2001). The sequence of the human genome. Science *291*, 1304-1351.

Verma, I. M., Stevenson, J. K., Schwarz, E. M., Van Antwerp, D., and Miyamoto, S. (1995). Rel/NF-kappa B/I kappa B family: intimate tales of association and dissociation. Genes Dev *9*, 2723-2735.

Wang, J. H., Nichogiannopoulou, A., Wu, L., Sun, L., Sharpe, A. H., Bigby, M., and Georgopoulos, K. (1996). Selective defects in the development of the fetal and adult lymphoid system in mice with an Ikaros null mutation. Immunity *5*, 537-549.

Weaver, D., and Baltimore, D. (1987). B lymphocyte-specific protein binding near an immunoglobulin k-chain gene J segment. Proc Natl Acad Sci U S A *84*, 1516-1520.

Wei, Z., and Lieber, M. R. (1993). Lymphoid V(D)J recombination. Functional analysis of the spacer sequence within the recombination signal. J Biol Chem *268*, 3180-3183.

Weih, F., Carrasco, D., Durham, S. K., Barton, D. S., Rizzo, C. A., Ryseck, R.-P., Lira, S. A., and Bravo, R. (1995). Multiorgan inflammation and hematopoietic abnormalities in mice with a targeted disruption of RelB, a member of the NF-kB/Rel family. Cell *80*, 331-340.

West, R. B., and Lieber, M. R. (1998). The RAG-HMG1 complex enforces the 12/23 rule of V(D)J recombination specifically at the double-hairpin formation step. Mol Cell Biol *18*, 6408-6415.

Whitehurst, C., Henney, H. R., Max, E. E., Schroeder, H. W., Jr., Stuber, F., Siminovitch, K. A., and Garrard, W. T. (1992). Nucleotide sequence of the intron of the germline human kappa immunoglobulin gene connecting the J and C regions reveals a matrix association region (MAR) next to the enhancer. Nucl Acids Res *20*, 4929-4930.

Wirth, T., and Baltimore, D. (1988). Nuclear factor NF-kappa B can interact functionally with its cognate binding site to provide lymphoid-specific promoter function. EMBO J *7*, 3109-3113.

Wirth, T., Staudt, L., and Baltimore, D. (1987). An octamer oligonucleotide upstream of a TATA motif is sufficient for lymphoid-specific promoter activity. Nature *329*, 174-178.

Xu, M., Hammer, R. E., Blasquez, V. C., Jones, S. L., and Garrard, W. T. (1989). Immunoglobulin κ gene expression after stable integration. J Biol Chem *264*, 21190-21195.

Xu, Y., Davidson, L., Alt, F. W., and Baltimore, D. (1996). Deletion of the Ig kappa light chain intronic enhancer/matrix attachment region impairs but does not abolish V kappa J kappa rearrangement. Immunity *4*, 377-385.

Yamagami, T., tenBoekel, E., Andersson, J., Rolink, A., and Melchers, F. (1999). Frequencies of multiple IgL chain gene rearrangements in single normal or κL chain-deficient B lineage cells. Immunity *11*, 317-327.

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

Yancopoulos, G. D., and Alt, F. W. (1986). Regulation of the assembly and expression of variable region genes. Ann Rev Immunol *4*, 339-368.

Yi, M., Wu, P., Trevorrow, K. W., Claflin, L., and Garrard, W. T. (1999). Evidence That the Igκ Gene MAR Regulates the Probability of Premature V-J Joining and Somatic Hypermutation. J Immunol *162*, 6029-6039.

Zachau, H. G. (1989). Immunoglobulin light-chain genes of the κ type in man and mouse. In Immunoglobulin genes, T. Honjo, F. W. Alt, and T. H. Rabbits, eds. (New York, Academic Press), pp. 91.

Zachau, H. G. (2004). Immunoglobulin κ Genes of Human and Mouse. In Molecular Biology of B Cells, F. W. Alt, T. Honjo, and M. Neuberger, eds. (London, Elsevier Science), pp. 27-36.

Zhuang, Y., Kim, C. G., Bartelmez, S., Cheng, P., Groudine, M., and Weintraub, H. (1992). Helix-loop-helix transcription factors E12 and E47 are not essential for skeletal or cardiac myogenesis, erythropoiesis, chondrogenesis, or neurogenesis. Proc Natl Acad Sci U S A *89*, 12132-12136.

Zhuang, Y., Soriano, P., and Weintraub, H. (1994). The helix-loop-helix gene E2A is required for B cell formation. Cell *79*, 875-884.

VITAE

Katherine Cherry Meyers Brekke was born in Charlotte, North Carolina, on March 25, 1974, the daughter of Cherry Sampson Meyers and Bruce McKinnon Meyers. In August of 1990, she entered the North Carolina School of Science and Mathematics, Durham, North Carolina. While there, she conducted research concerning water quality of freshwater and marine estuaries in conjunction with the Environmental Research Project in Chapel Hill, North Carolina. After receiving her degree in 1992, she entered the University of North Carolina at Chapel Hill, Chapel Hill, North Carolina. During the years of 1995 and 1996, she worked in the laboratory of Dr. Oliver Smithies at the University of North Carolina at Chapel Hill. Her research there involved creation of single copy targeted Bcl2 transgenic mice. She received the degree of Bachelor of Science with a major in Biology and a minor in Chemistry from the University of North Carolina at Chapel Hill in June, 1996. In August, 1996, she entered the Graduate School of Biomedical Sciences at the University of Texas Southwestern Medical Center at Dallas. She joined the laboratory of Dr. William Garrard in 1997 where she completed her dissertation research. In 2002, she married Monty Howard Brekke of Carrollton, Texas.

Permanent Address: 1371 Farrington Rd.
Apex, NC 27523