1-1-2014

The Rag2 C Terminus Participates in Repair Pathway Choice in Vivo and Suppresses Lymphomagenesis

Vered Gigi
University of Pennsylvania, vered.gigi@gmail.com

The Rag2 C Terminus Participates in Repair Pathway Choice in Vivo and Suppresses Lymphomagenesis

Abstract
THE RAG2 C TERMINUS PARTICIPATES IN REPAIR PATHWAY CHOICE IN VIVO AND SUPPRESSES LYMPHOMAGENESIS

Vered Gigi

Dr. David B Roth

DNA double-stranded breaks (DSBs) can be repaired by several mechanisms, including classical NHEJ (c-NHEJ) and a poorly defined, error-prone process termed alternative NHEJ (a-NHEJ). How cells choose between these alternatives to join physiologic DSBs remains unknown. Here we show that deletion of RAG2's C-terminus allows a-NHEJ to repair RAG-mediated DSBs in developing lymphocytes from both c-NHEJ-proficient and c-NHEJ-deficient mice, demonstrating that the V(D)J recombinase influences repair pathway choice in vivo. Analysis of V(D)J junctions revealed that, contrary to expectation, junctional characteristics alone do not reliably distinguish between a-NHEJ and c-NHEJ. These data suggest that a-NHEJ is not necessarily mutagenic, and may be more prevalent than previously appreciated. Whole genome sequencing of lymphomas arising in a p53-/- mouse bearing a C terminal RAG2 truncation reveals evidence of a-NHEJ and also of aberrant recognition of DNA sequences resembling RAG recognition sites. The ability to recognize these sites is not because of specificity relaxation due to the lack of the RAG2 C terminus but probably other potential mechanisms that should be further investigated.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Immunology

First Advisor
David B. Roth

Keywords
alternative NHEJ, classical NHEJ, DNA repair, RAG2, VDJ recombination

Subject Categories
Allergy and Immunology | Immunology and Infectious Disease | Medical Immunology

THE RAG2 C TERMINUS PARTICIPATES IN REPAIR PATHWAY CHOICE IN VIVO AND SUPPRESSES LYMPHOMAGENESIS

Vered Gigi

A DISSERTATION

in

Immunology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

2014

Supervisor of Dissertation

_____

David B. Roth
Simon Flexner Professor and Chair of Pathology and Laboratory Medicine

Graduate Group Chairperson

_____

Michael P. Cancro
Professor of Pathology and Laboratory Medicine

Dissertation Committee:

Craig Bassing, Associate Professor of Pathology and Laboratory Medicine
Avinash Bhandoola, Professor of Pathology and Laboratory Medicine
David Allman, Associate Professor of Pathology and Laboratory Medicine
Warren Pear, Gaylord P. and Mary Louise Harnwell Professor of Pathology and Laboratory Medicine

Dedication

I would like to dedicate this work first and foremost to my beautiful parents. My mom Ayala Gigi, and my Dad Moti Gigi (RIP) who always supported me, pushed me to be a better person and loved me unconditionally. I would also want to dedicate it to my loving husband Jon Tekac and the best siblings one can ask for Ziv Gigi and Adva Gigi. Their love, understanding and continued support throughout the years made my academic adventure feasible and a great experience.

ACKNOWLEDGMENT

A special big thank you to my mentor and advisor Dr. David Roth. Joining his lab in NYU and then moving with him to UPenn was the best decision I made during my graduate endeavor. I appreciate his support, scientific knowledge and vision that were instrumental for shaping my work. Most importantly, I am grateful for his confidence and belief in me to allow me explore and develop my own scientific thinking. His mentorship, guidance and advice extend beyond the science world and for that I am very thankful.

I would also like to say a special thank you to Dr. Susanna Lewis that had to endure my endless questioning. Her assistance, patience and deep scientific insights were very influential in interpreting this body of work.

I thank the chair of my thesis committee Dr. Craig Bassing for always being accessible and willing to help in any matter that arose. I appreciate his support and guidance when walking some fine lines in the world of academia and for always seeking the best solutions to enhance both my scientific work and myself. I am also grateful for my thesis committee: Dr. Dave Allman, Dr. Avinash Bhandoola and Dr. Warren Pear. They were always ready to assist me conceptually as well as technically to develop and deepen my scientific work.

Special thank you to my current and former lab members that shared this great experience with me. Dr. Bette Pancake and Dr. Olga Shestova who supported

me on a daily basis professionally and personally, it would not have been the same without you. Your help made my experience that much better and I could not ask for better bench mates. Also, Dr. Suzzette Arnal, Yi-Fan Chou, Dr. Marc Coussens, Dr. Ludovic, Deriano Dr. Cory Lindsay, Dr. Martina Mijuskovic, Dr. Jennifer Villa, Zoe Zampana and Alexis Lerro for their support and friendship.

I am thankful for my family, Ayala Gigi, Ziv Gigi, Adva Gigi and Janice & John Tekac for their emotional and logistic support throughout the years. My son, Eytan, for all the happiness, sunshine and meaning he brings into my life every day. Lastly, my husband, Jonathan Tekac for putting up with me, for his strength, for his unwavering love and his infinite belief in me. He is my best friend and my biggest advocate. I wouldn't have accomplished all of that without him.

ABSTRACT

THE RAG2 C TERMINUS PARTICIPATES IN REPAIR PATHWAY CHOICE IN VIVO AND SUPPRESSES LYMPHOMAGENESIS

Vered Gigi

Dr. David B Roth

DNA double-stranded breaks (DSBs) can be repaired by several mechanisms, including classical NHEJ (c-NHEJ) and a poorly defined, error-prone process termed alternative NHEJ (a-NHEJ). How cells choose between these alternatives to join physiologic DSBs remains unknown. Here we show that deletion of RAG2's C-terminus allows a-NHEJ to repair RAG-mediated DSBs in developing lymphocytes from both c-NHEJ-proficient and c-NHEJ-deficient mice, demonstrating that the V(D)J recombinase influences repair pathway choice *in vivo*. Analysis of V(D)J junctions revealed that, contrary to expectation, junctional characteristics alone do not reliably distinguish between a-NHEJ and c-NHEJ. These data suggest that a-NHEJ is not necessarily mutagenic, and may be more prevalent than previously appreciated. Whole genome sequencing of lymphomas arising in a $p53^{-/-}$ mouse bearing a C terminal RAG2 truncation reveals evidence of a-NHEJ and also of aberrant recognition of DNA sequences resembling RAG recognition sites. The ability to recognize these sites is not because of specificity relaxation due to the lack of the RAG2 C terminus but probably other potential mechanisms that should be further investigated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

INTRODUCTION

Misrepair of double-stranded breaks (DSBs) creates structural genomic lesions (deletions, chromosome translocations, duplications, and inversions), which can fuel oncogenic transformation (1,2). The need to inhibit such events and maintain genome integrity is very clear. However, how cells manage to do so is not. Resolution of DSBs is a complex cascade of events influenced by the stimulus that originated the break (e.g. nucleases vs radiation) and the environment in which it was generated (e.g. cell cycle phase, chromatin status). Such a degree of complexity is inherently error-prone, however, cells have evolved to be very efficient in faithful repair of DSBs. There are three main repair mechanisms to aid in joining the broken ends: Homologous Recombination (HR), classical Non Homologous End Joining (c-NHEJ) and alternative NHEJ (a-NHEJ). While the first two are well defined, our understanding of a-NHEJ is still in its infancy.

1.1 DNA Repair mechanisms

Homologous recombination

The joining of DSBs by HR is error-free since it copies the missing genetic information from an undamaged sister chromatid and as such only functions in late S phase when a sister chromatid is available (3). It is considered the best

repair mechanism to protect against mutagenic outcomes from DNA damage (Fig. 1.1). HR repair is initiated via DSB detection and binding by the MRE11-RAD50-NBS1 (MRN) complex (4). This step promotes resection together with CtBP-interacting protein (CtIP; CtBP = C-terminal binding protein), to generate short 3'-single-stranded-DNA (ssDNA) and then long ssDNA with the help of Exo1 and the BLM proteins (4). ssDNA is immediately coated with replication protein A (RPA) (4,5). Following is displacement of the RPA A by the BRCA2-PALB2 complex to create RAD51 nucleation onto ssDNA (6). The RAD51 nucleoprotein filament then invades into a similar or identical DNA duplex forming a displacement loop (D-Loop). A DNA polymerase then extended the invading strand creating a Holliday junction. The second end of the broken chromosome (non-invading strand) anneals to the complementary strand of the donor DNA molecule, resulting in a second Holliday junction. Once DNA synthesis and ligation of both strands is completed, the HJs are resolved in order to complete repair (7).

Classical Non Homologous End joining

Unlike HR, c-NHEJ is active throughout the cell cycle (though favored in the G1 phase) and, as its name suggests, is not dependent on sequence homology for repair (8). In a simplistic view, c-NHEJ allows direct ligation of two DSBs. Nevertheless, more than often, the breaks are not ligatable, requiring a degree of

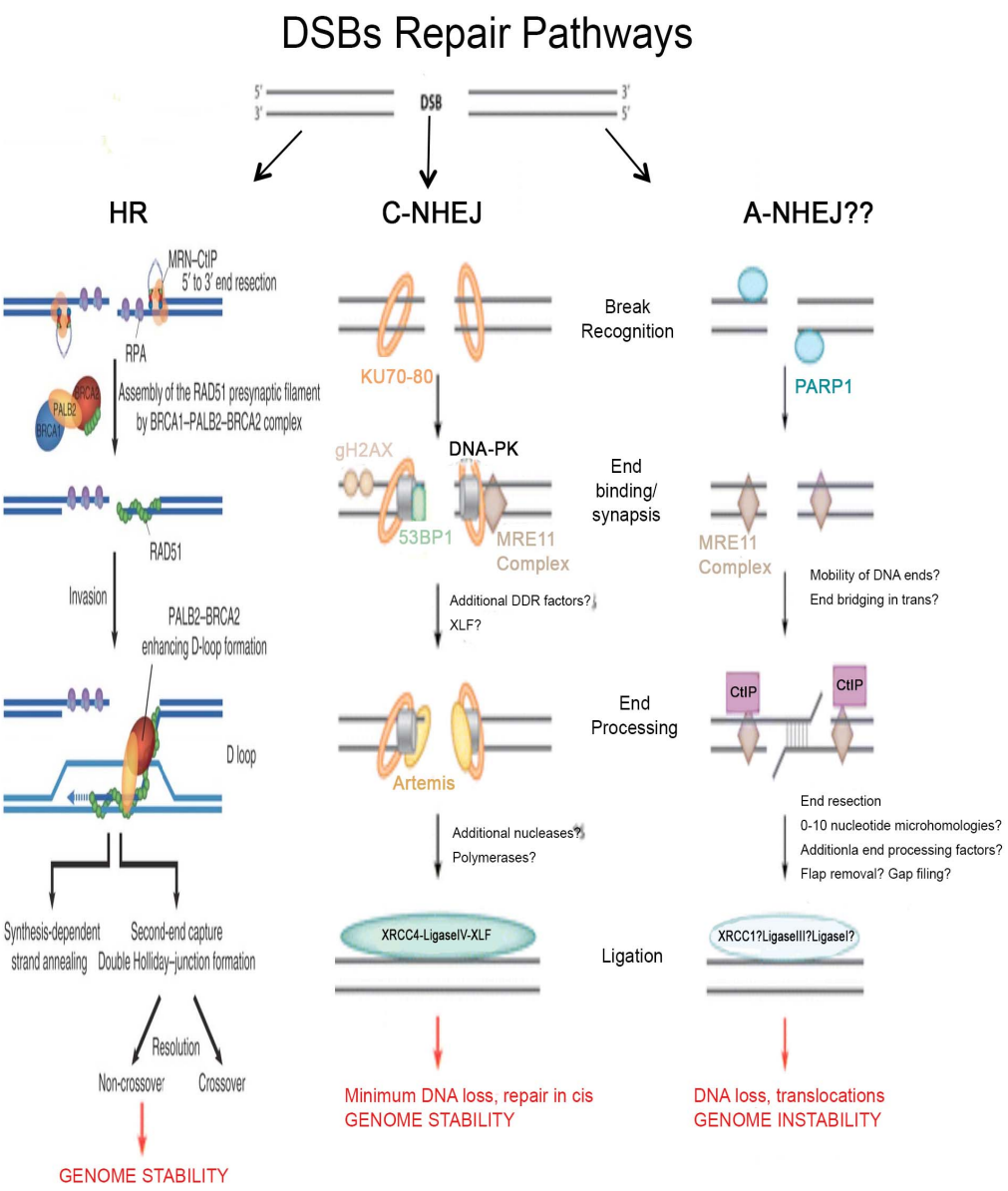


Figure 1.1 DSBs repair pathways. Homologous Recombination (HR) was adapted and modified from Rémi Buisson & Jean-Yves Masson Nature Structural & Molecular Biology 2010. Classical NHEJ (C-NHEJ) and Alternative NHEJ (A-NHEJ) were adapted and modified from Deriano L & Roth DB Annu.Rev.Genet 2013.

processing leading to nucleotide loss and/or addition. Therefore, c-NHEJ is considered ‘error prone’. However, on a genomic level, c-NHEJ appears to limit genomic damage and suppress tumorigenesis (9,10,11,12). c-NHEJ repairs DSBs that are either generated inadvertently (e.g irradiation) or ones that are physiologically important for development like in the process of Variable (V) Diversity (D) Joining (J) (V(D)J) and Class Switch Recombination (CSR) in the immune system. A lot of our understanding the c-NHEJ in the mammalian system comes from various manipulations carried out in these two systems.

c-NHEJ is initiated with the binding of the Ku70/80 heterodimer (KU) and the MRN complex to the broken ends (Fig.1.1). The KU complex is the first to bind to the DNA on the basis of its abundance (estimated at ~400,000 molecules per cell) and its binding capacity ($\sim 10^{-9}$ M)(13,14). The KU:DNA complex on each end serves as a ‘tool belt’ for other proteins to dock onto in preparation for joining (8). Though there is no evidence for a single recruitments cascade of proteins to the KU:DNA complex to facilitate repair (15,16), it is likely that the catalytic subunit of DNA-dependent protein kinase (DNA-PKcs) follows to create the DNA-PK holoenzyme (DNA-PK) (17). This step leads to phosphorylation events mediated by the DNA-PK resulting in diverse DNA end processing by the Artemis nuclease to prepare ends for ligation (18,19). Often but not obligatory there is an involvement of the pol X polymerase family. Poly μ, λ and terminal deoxynucleotidyltransferase (TdT) are capable of binding to the KU:DNA

complex and insert short templated or non-templated neucleotides (15,20). Completion of the repair is catalysed by a complex consisting of X-ray repair cross-complementing protein 4 (XRCC4), DNA ligase IV and XRCC4-like factor (XLF) (21-24). It is important to mention that some factors of c-NHEJ are conserved throughout evolution and are essential (Ku70/80, XRCC4Ligase 4 and XLF) for repair while others can be dispensable depending on the context. For example, Artemis is necessary to generate coding joints as it opens the hairpin but it is dispensable for signal joints (25).

Great flexibility exists in this repair both in terms of the breaks it can handle and in the ligation step (18,19,26,27). This is beautifully demonstrated in V(D)J recombination. First, the c-NHEJ machinery needs to join two very different ends (blunt and hairpin sealed, reviewed later). Second, in the case of the hairpin ends, although they are the same starting substrate, the joining junctions at the end of the repair are entirely different. This c-NHEJ flexibility is beneficial for the organism but complicates our efforts for understanding processes that occur in its absence.

c-NHEJ is considered error prone suggesting its involvement in genomic lesions. However, studies of genomic aberrations in various tumors cannot unequivocally prove or disprove this statement. On the one hand, mice deficient for both c-NHEJ factors and the p53 develop lymphomas very rapidly with recurrent translocations (9,10) suggesting that c-NHEJ suppresses tumorigenesis. On the other hand, p53 deficient or Ataxia telangiectasia mutant (ATM) mice, in which c-

NHEJ is functional, also develop lymphomas with different translocations (28-30). Hence, the contribution of c-NHEJ to such events is still an open question.

Alternative Non Homologous End joining

A third repair mechanism has emerged in the last decade but our understanding of it is still a work in progress (Fig. 1.1). As a testimony are the multiple names given to it: Alternative NHEJ (a-NHEJ; which will be used throughout this body of work), Backup NHEJ (B-NHEJ) or microhomology-mediated end joining (MMEJ). a-NHEJ was discovered in cells deficient for c-NHEJ where transfected, linearized plasmids were able to recircularize and be recovered from cells so that their joining properties could be evaluated (31,32). Aside from clearly showing that c-NHEJ-independent end joining is present it also started shaping our thoughts of what a-NHEJ might look like in terms of joint structure. Unfortunately the analysis of junctions in c-NHEJ-deficient settings emphasized the joints that looked different from those formed by c-NHEJ, **underestimating** the fraction that did resemble cNEHJ joints. Moreover, some results were not consistent across different systems (31,33,34) or different c-NHEJ deficiencies (31). These issues in turn raised the possibility that a-NHEJ might be a collection of different pathways rather than a single one.

A repair pathway can be evaluated through different lenses and in the case of a-NHEJ it is concentrated on three criteria: the quality of the end joining, potential factors and its predisposition to genomic instability.

End joining quality: simply put, any characteristics of joined ends that deviate from typical c-NHEJ are considered a-NHEJ. This is a very problematic criterion because c-NHEJ, as mentioned earlier, shows a diverse spectrum of joining products. Nonetheless, finding of extensive deletions, microhomologies and long insertions became viewed as indications of alternative joining (35,36,31,33). In common practice, as long as one characteristic can be demonstrated it is considered a-NHEJ. The great challenge with this criterion is how we interpret such results. For instance, in a Ku70-deficient cell line the length of microhomology and median length of deletion in translocation events were not different from those in wild-type cells where c-NHEJ is present (37). Hence, we cannot be certain that the junctional features associated with a-NHEJ are really a reliable criterion to distinguish between classical and alternative NHEJ.

Potential Factors: To date no one specific factor associated with a-NHEJ has been described. Moreover, one can find contradictory results regarding a potential factor and its contribution to a-NHEJ (38). Most factors are associated with other DNA repair pathways adding to the complexity of elucidating this repair. The potential factors are almost a “tit for tat” of the c-NHEJ machinery. Poly [ADP-ribose] polymerase 1 (Parp1) is thought to compensate for Ku80 deficiency and bind to the broken ends to initiate the repair process (39-42). This is followed by end resection via the CtBP-interacting protein (CtIP) and the Mre11 nucleases (37,43,44). This resection creates the extensive deletions observed in

order to uncover stretches of microhomology that can anneal. Subsequently, gap filling by a DNA polymerase and ligation via the XRCC1/ligase 3/1 complex completes the repair process (39,45,46).

It is still unclear whether, how and to what degree these factors interact with each other in order to achieve this repair. It is plausible that there are multiple ways for this cascade of events to take place. Additionally, for the most part, these factors were tested in the absence of one of the c-NHEJ proteins. Hence it is also a possibility that these potential factors might interact with part of the c-NHEJ cascade to complete repair (47).

Predisposition to genomic instability: This functional assessment of a-NHEJ was determined due to tumor studies done in c-NHEJ-deficient mice (9,10). Mice deficient for p53 and any either of the following: Ku80-/-, Ligse4-/- or Xrcc4-/-, developed rapid pro-B cell lymphomas with recurrent translocations involving the IgH locus and the c-myc oncogene (9,10). As the c-NHEJ repair is disabled in these mice it was concluded that a-NHEJ repair disrupts the integrity of the genome leading to tumorigenesis. Further studies using various cell lines deficient for c-NHEJ supported that conclusion by showing an increase in translocation frequencies (38,48,49). This was followed by research demonstrating the involvement of CtIP and ligase3, potential a-NHEJ factors, in mediating such chromosomal translocations (37,46). However, contrasting results were obtained from mice and cells lacking Parp1 concluding that this factor is important for genome stability (50,51). Lastly, involvement of a-NHEJ

was also documented in human tumors where a decrease in Ku80 and increase in Parp1 or Ligase 3 were observed (52-54). It is important to note that the quality of junctions seen in translocation events from the different systems could not **consistently** be distinguished from junctions mediated via c-NHEJ repair. This issue nicely illustrates the problem we are facing with the current criteria that define a-NHEJ. Moreover, other genomic aberrations such as deletions, duplications and inversions were not systematically studied in these systems to evaluate the extent of a-NHEJ contribution to genomic instability.

Based on the above criteria a-NHEJ was evaluated also in cultured cells proficient for c-NHEJ repair and was shown to be active (36,46,55). This suggested that mechanisms might exist to limit its usage, presumably to preserve genomic integrity and suppress tumorigenesis. Whether a-NHEJ is another repair pathway, a collection of sub pathways, or just the extreme end of the c-NHEJ flexibility is yet to be determined.

1.2 DNA repair and pathway choice

How cells control choice of a particular pathway (homologous recombination, c-NHEJ, or a-NHEJ) for repair of a given DSB is a question of intense current interest. The prevailing notion is that there is an active decision between the different repair pathways. Early experiments showed that an extra chromosomal

substrate, when transfected into mammalian cells, could be repaired either by HR or NHEJ, indicating that both repair pathways are available and compete with each other (56). Misrepair of a DSB is associated with a variety of pathologies (e.g immunodeficiency) and increased cancer susceptibility. What criteria channels a certain DSB from one repair pathway to another would be beneficial in understanding tumor etiology, evolution and in developing targeted therapy against it.

Different pathways can repair DSBs, which is determined in part by the cell cycle phase in which the breaks occurred. Through the cell cycle phases the cell controls availability of DSBs sensors/repair factors and presence of a homologous template that is critical for HR. However, the stage of the cell cycle is not **sufficient** for correct repair as we know that c-NHEJ factors exist throughout the cell cycle and homologous chromosome, that can serve as a template and HR factors, exist in the G0/G1 stage. Moreover, the recruitment of factors from both pathways to DSBs can happen simultaneously and independently from each other both in G1 and S/G2 phases (57).

It is thought that end resection is one critical step that can commit cells to HR rather than c-NHEJ and this is mediated via the CtIP/MRN complex (58,59). Whether in the G1 or S/G2 phase, recruitment of CtIP would lead to homology-mediated repair; a-NHEJ in the G1 phase a-NHEJ/HR in the S/G2 phase. Hence, different mechanisms exist to regulate CtIP activity. In the S/G2 phase CtIP activity is increased via Cyclin dependent kinase 1 (CDK1) to promote end

resection for HR (4). In the G1 phase, histone H2AX prevents CtIP end resection in lymphocytes (58), and 53BP1 inhibits BRCA1, a CtIP interacting protein, from accumulating on DSBs to suppress end resection favoring NHEJ (59). However, a different study showed that BRCA1 stabilized Ku80 binding to chromosomal DSBs resulting in high fidelity NHEJ repair (60). In addition to end resection, maintaining an overall balance between HR and NHEJ factors and their activity is also important for correct repair. Deficiency in c-NHEJ factors shifts the balance in favor of HR/a-NHEJ (40,48,61,62) while decrease in HR proteins in G1 phase supports NHEJ repair (63). For example, DNAPKcs levels are constant throughout the cell cycle yet its activity is not. Phosphorylation of residue 2056 in G1 is important for proper NHEJ and is diminished in S phase, inhibiting this repair (64). It is important to note that some repair factors are needed for both repair machineries adding another layer of complexity (43,60,64)

Most of our knowledge refers to the choice between HR and NHEJ as a whole. However, within NHEJ we distinguish c-NHEJ vs a-NHEJ. Several proposals exist regarding this pathway choice. Some a-NHEJ are mediated via microhomology, positioning short end resection by the MRN/CtIP complex as a regulator of that pathway (37,43). Similarly, 53BP1 inhibits such resection therefore promoting c-NHEJ (65,66). In the V(D)J recombination process, the RAG complex was shown to shuttle broken ends away from a-NHEJ to the c-NHEJ in vitro (36).

However, most attempts to study this pathway choice were investigated in the absence of one or more critical c-NHEJ repair factors and often using an artificial readout (31,33,35). These genetic approaches are beneficial once a comparison between **defined** systems has been investigated. The best deficnition of a-NHEJ to date is 'repair in the absence of c-NHEJ'. Hence, how can one study a choice between two pathways when one is dependent on the other's absence? This hampers our attempts to understand mechanisms of pathway choice, as the missing factor might itself be involved. Another caveat is that these approaches may bias the type of junctions produced by a-NHEJ. For instance, according to some research, Ku80 is thought to protect the ends from degradation (33). Hence, disabling end protection mechanisms can lead to excessive deletion or production of long single-stranded tails that are not necessarily a true feature of a-NHEJ. Lastly, it has been shown that experimental knockouts can generate secondary mutations and adapt to the deficiency (67). As a result we must find a system where the c-NHEJ factors are present and the stimulus of DSBs is physiological. One such system is the V(D)J recombination process where a-NHEJ repair has been shown to be very active in the presence of c-NHEJ repair machinery (36).

1.3 V(D)J recombination

V(D)J recombination provides a tractable and physiologically relevant system in which to explore end-joining pathways and regulation of pathway choice. This process is responsible for the development and the diversity of our B and T cells that grant us with a functional and robust immune system. In a tightly regulated process, our DNA is purposefully cleaved, 'rearranged' and then rejoined to create the immunoglobulin Ig and T cell receptor (TCR) while maintaining genome integrity (8,68,69). There are three Ig (IgH, IgK, Igλ) and 4 TCR loci (TCR α/β/γ/δ) that are located on different chromosomes. IgH and TCRβ/γ undergo 2 sequential rounds of recombination: D-J followed by V-DJ to generate a functional rearrangement. After a successful attempt, IgK/λ and TCRα/δ carry a single V to J recombination event. A successful rearrangement results in a functional BCR/TCR that then undergoes positive and negative selections that determine its faith (survival vs apoptosis).

The process begins when the V(D)J recombinase, comprised of the protein products of recombination activating genes 1 and 2 (the RAG1/2 proteins, RAG), cleaves the DNA between an antigen receptor coding segment (V/D/J) and a flanking recombination signal sequence (RSS)(69) (Figure 1.2). RSS consist of conserved heptamer (CACAGTG) and less conserved nonamer (ACAAAAACC) sequences separated by a non-conserved spacer of 12 or 23 nucleotides. Normally a synapsis of a 12/23 RSS pair is required to form DSBs. It is thought

that the synaptic complex formation occurs in a sequential manner where the RAG complex binds and perhaps nicks one RSS and then captures the other RSS to form the synaptic complex (70-72). Following is nicking of the partner RSS culminating in coupled cleavage of both RSSs via a transesterification process through the 3′-OH group exposed by nicking. This process produces two covalently sealed (hairpin) coding ends and two 5′-phosphorylated blunt signal ends (69,73). Lastly is a joining step that is facilitated by the RAG post cleavage complex (PCC) and the c-NHEJ repair machinery (36,69,74). The coding ends undergo a degree of processing starting with the opening of the hairpin by the DNAPK-Artemis complex (19). This can be succeeded by a loss and/or insertion of few nucleotides. Insertions are of two types: ‘P’ (Palindromic) and ‘N’ (Non-Templated) nucleotides. The first are the result of an asymmetrical opening of the hairpin by Artemis and are a complementary copy of the termini (75). The second are insertions of non-templated nucleotides mostly by the TdT polymerase (76). A ligation step mediated by the XRCC4/Ligase4 complex follows to generate a coding joint (CJ) representing the rearranged variable regions of antigen receptor genes. In contrast to coding ends, signal ends are slower to repair and are blunt ligated with rare instances of a few nucleotides insertions or deletion (77-80). For the most part they form an excised circular molecules that are lost with cell divisions. Functionally, signal joints were thought to be immunologically inert but some studies have shown that they can be recleaved and reinserted into the genome disturbing its integrity (81,82).

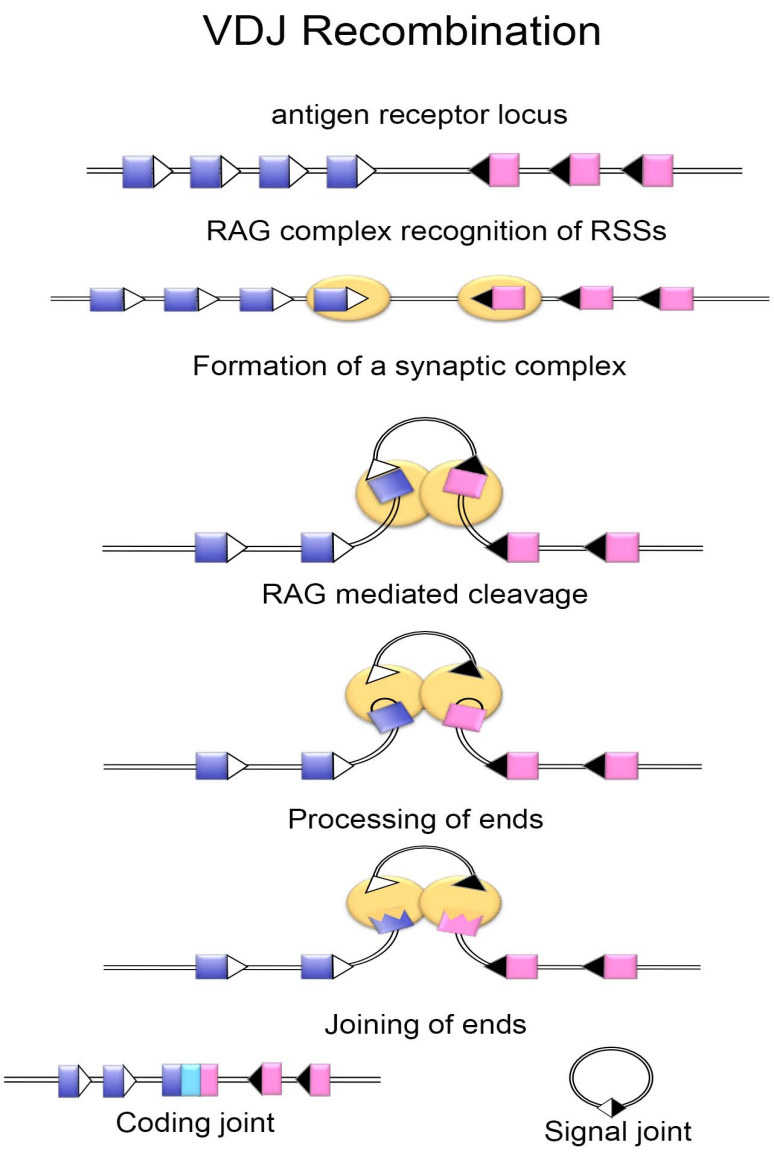


Figure 1.2 A simplified overview of the V(D)J recombination process. The RAG complex (yellow circles) recognizes and binds the RSSs (triangles). Formation of a synaptic complex and culmination of the cleavage reaction. Subsequent is processing of the coding ends and then ligation to generate a coding joint and a signal joint.

Another, atypical type of end joining was observed in the late 80's where a signal end joined a coding end, generating a hybrid joint (HJ)(83). These joints clearly do not result in functional antigen receptor rearrangements but can be used as a marker of improper recombination. For example, in RAG2 Core mice (amino acid 1-352), which are devoid of the C terminus, there is an increase in HJ formation vs wild-type mice implying its importance in regulating proper recombination and maintaining genomic information (84).

There are multiple levels of regulation to ensure proper recombination. First, there is lineage specificity; Ig loci rearrange in B cells and TCR in T cells. This is achieved by actively transcribing only the locus that needs to rearrange (85,86). Second, recombination is carried out in cis rather than in trans to eliminate chromosomal translocation (87). Third is the control of RAG1/2 expression. While RAG1 can be detected throughout the cell cycle, RAG2 is restricted to the G0/G1 phase due to a protein degradation signal (88). RAG1 cannot create DSBs without RAG2 ensuring that they do not occur in the S/G2 phase when replication occurs. Last but not least are epigenetic markers. Though H3K4me3 is not essential for recombination it is an important determinant of RAG2 localization and efficient recombination (89,90). Interfering with any of these regulations and others might decrease V(D)J recombination efficiency and could also lead to illegitimate recombination that can be very deleterious to the genome integrity.

Illegitimate V(D)J recombination

DNA cleavage and joining during the V(D)J process must be tightly regulated to prevent genomic lesions. A role for V(D)J recombination in such events was hypothesized when translocations in certain tumors showed a fusion of the antigen receptor gene to another region of DNA (91,92). Since then, various Illegitimate V(D)J recombination events have been well documented in the literature both in mice and humans (93-97).

These events can be categorized into five groups (Fig. 1.3):

1) Joining of bona fide RSSs in trans rather than cis. In such rearrangements RSSs from loci on different chromosomes recombine and create a translocation.

2) Joining of a bona fide RSS to an RSS-like sequence outside of the antigen receptor loci (termed cryptic RSS, cRSS).

3) Joining of two cRSSs.

4) Joining of a bona fide RSS to a random DSB that was generated in a process not related to V(D)J recombination (referred to as end donation).

5) Transposition. DNA cleavage by the RAG complex creates a 3' hydroxyl group (OH) on a signal end. This reactive group can 'attack' a target DNA and transpose the signal end into that region (81). These events are very rare in vivo because they can only be identified if the signal end retains its sequence (98).

Aside for group one, which is an obligatory trans rearrangement, all the rest can happen either in cis or trans. Genomic structural variants (SVs) such as

translocations, deletions, inversions and amplifications generated via these illegitimate processes have all been reported (91-95,97). The different types of events are also telling in regards to the mechanism that might promote them. In groups 1 and 4 where authentic RSSs are used we can assume that there was no mistake in RSS sequence recognition. This is in striking contrast to groups 2&3 where the RAG complex misrecognizes a sequence to be an RSS and at minimum binds and cleaves it.

Outcomes of V(D)J aberrant events are detrimental to a cell and to an organism as a whole. They can impede the development of the immune system and decrease its diversity. For example, inter-locus rearrangements between the TCR beta on chromosome 6 and TCR gamma on chromosome 14 can lead to a translocation that abolishes functional TCR rearrangements (99,100) resulting in cell death (101).

More hazardous are illegitimate V(D)J recombination events that participate in pathogenesis of lymphomas. A recent study found that almost 40% of SVs detected in ETV6-RUNX1–positive Acute Lymphoblastic Leukemia (ALL) are a consequence of such events, serving as a major mechanism in this tumor progression (97). Also, end donation events (group 4) are implicated in translocations found in 30-40% of follicular and mantle cell lymphomas (100,102). So far there is no standardized scheme to evaluate cRSS. The minimal requirement is to have a ‘CAC’ sequence in the vicinity of the break (96).

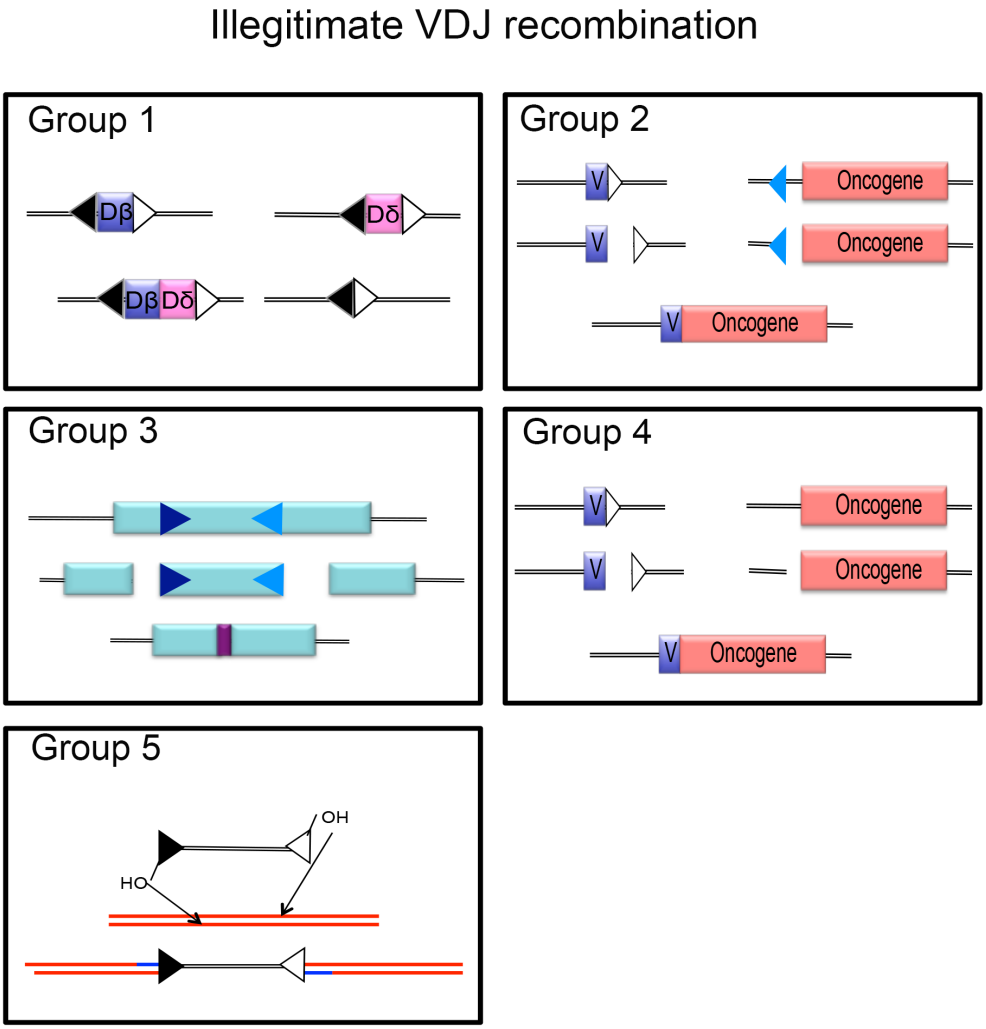


Figure 1.3 illegitimate V(D)J recombination scenarios. Group1; recombination between two bona fide RSSs that are on different chromosomes resulting in translocation. The same scenario can present itself in cis. For example, Db1-Db2. Triangles are 12/23 RSS Group2; recombination between bona fide RSS and a cRSS. Again this can happen in cis or trans. The oncogene is used as a demonstration for the deleterious effect of such event. Light blue triangle represents cRSS. Group 3; recombination between two cRSSs; the cis configuration is presented but it can also happen in trans. Triangles represent cRSSs. Group 4; recombination between a bona fide RSS and a random break, again in cis or trans. Group 5; transposition event mediated by 12/23 signal ends, which can interrupt genetic information. In blue are the direct repeats of the targeted sequence.

Several other features, if present, can be used to strengthen involvement of V(D)J recombination: nonamer-like sequence within 12/23 bp of the heptamer, and presence of 'N' or 'P' nucleotides at the joint. Lewis et al showed that a cryptic heptamer, which is sufficient for recombination, could be found every 600bp in the mouse genome illustrating the great potential of illegitimate V(D)J recombination(96).

1.4 V(D)J recombination and repair pathway choice

Efficient and faithful end joining in the V(D)J process is critical to ensure a diverse immune system and to prevent both genome instability and lymphomagenesis. The process, as outlined earlier, is dependent on c-NHEJ machinery. Mice deficient for c-NHEJ factors, if not embryonic lethal (103,104), exhibit developmental block at the pro-B/pro-T cell stage accompanied by severe lymphopenia because they cannot complete the antigen receptor rearrangements (105-108).

On the face of it, V(D)J recombination has tremendous potential for errors. It introduces DSBs in large numbers of lymphocyte progenitors, and, through end-to-end joining, generates megabase-sized modifications of the genome (69). The system has evolved to minimize aberrant events. For example, the C-terminus of RAG2 is evolutionarily conserved and necessary for **proper** recombination (28,36,74,99,109,110). Moreover, this region is important for the stabilization of

the post cleavage complex, as seen in biochemical studies, possibly to ensure that the ends are shuttled to the c-NHEJ rather than the a-NHEJ repair pathways (74,84).

Our understanding of 'classical' NHEJ has relied heavily on determining requirements for coding and signal joint formation in V(D)J recombination and repair of radiation-induced DNA breaks. Because exposure to a different suite of enzymes and repair 'platforms' is likely to affect the fine-structure of the repair junctions, the available comprehensive analyses of V(D)J recombination outcomes collected over the last thirty years provide an ideal basis by which to explore and define alternative NHEJ. In this system the DSBs are physiological rather than artificial, the breaks are tractable and we can assess the quality of both non-processed and processed end joining represented by signal and coding joints respectively. Previous work in the Roth lab suggests that it might be possible to alter the end-joining environment encountered by broken DNA ends by only mutating RAG1/2 that are extrinsic to the c-NHEJ repair pathway (36,74). Hence, this can provide a very 'clean' system to test the junctional features associated with a-NHEJ without modifying repair factors. Moreover, these mutants can be used to investigate whether a-NHEJ predisposes to tumorigenesis. If the increase in a-NHEJ observed with extrachromosomal substrate holds true in vivo, then these mice should develop lymphomas with genomic lesions mediated via a-NHEJ.

CHAPTER 2

RAG2 Mutants Alter DSB Repair Pathway Choice *In Vivo* and Illuminate The Nature of 'alternative NHEJ'

In order to look at repair pathway choice in a non-perturbed fashion we needed a system that does not modify the end joining factors. Our previous work suggests that it might be possible to alter the end-joining environment encountered by broken DNA ends by only mutating RAG1/2 that are extrinsic to the c-NHEJ repair pathway (36,74). Indeed, a particular C-terminally truncated RAG2 mutant termed FS361, identified in our lab, allows coding ends to abnormally access a-NHEJ (36). This was assessed using RAG expression vectors transfected into fibroblasts along with an extrachromosomal substrate specifically designed to detect joints bearing both excessive deletion and microhomologies that have been considered characteristic of such repair (35,36,68,111). This implies that RAG2's C-terminus is important for control of pathway choice, at least in this artificial system.

Focus on RAG2 is additionally supported by studies in which the consequences of germline mutations in the RAG2 C-terminus have been examined in whole mice (28,99,110). Though not addressed directly in those reports, the

recombinant V(D)J junctions that were observed raised the possibility that each of the C-terminal mutations may have had an impact upon pathway choice. Hence, to seek definitive evidence of functional alternative pathways, specifically a-NHEJ, without changing components of c-NHEJ we generated homozygous knockin mice bearing the FS361 mutation ($RAG2^{FS/FS}$) that exhibited the highest a-NHEJ signal among those tested in transfection studies. In addition to establishing physiological relevance of pathway choice *in vivo,* this minimally-manipulated, physiologically relevant system is ideally suited to investigate the hypothesis that a-NHEJ is error-prone and makes a disproportionate contribution to oncogenic genome rearrangement.

Materials and Methods

Mice

We obtained wild-type C57BL6 (Taconic), Ku80 KO (The Jackson laboratory, 108), RAG2 KO (The Jackson laboratory) and $RAG2^{C/C}$ which we renamed $RAG2^{del352/del352}$ (112). $RAG2^{FS/FS}$ mice were generated by Ingenious Targeting Laboratories as described in figure S1. The nucleotide sequence of the entire RAG2 ORF was verified by sequencing genomic DNA from somatic tissues of the knockin mouse. $Rag2^{FS/FS}$ and $RAG2^{del352/del352}$ were bred with Ku80 KO mice to generate doubly deficient mice. Genotyping of all mice was performed by PCR of tail DNA as described (108,112). The animals' care was approved by UPenn Institutional Animal Care and Use Committee (IACUC) Protocol # 803893

Flow cytometry

Cells from thymus, bone marrow and spleen were obtained from the indicated genotypes and stained for B cell (B220, CD43, IgM) and T cell (CD4, CD8, thy1.2, TCR beta, CD25, CD44) markers. FACS analysis was done using the BD LSR II and FlowJo software.

PCR for CJs, SJs and Interchromosomal rearrangements

Genomic DNA from thymus and BM were prepared from 6-9 week old mice for wild-type, $RAG2^{FS/FS}$ and $RAG2^{del352/del352}$ characterization and 4 week old mice for joints from Ku80 deficient backgrounds. Genomic DNA for coding joints (CJ, 200 ng) or signal joints (SJ)/interchromosomal (500 ng) were amplified by PCR.

CJ/SJ primers were described previously (110) aside from Vb10 SJ, Vh7183 CJ that are described below. For interchromosomal rearrangements we used nested PCR with primers described below. All PCR products were cloned using the TA cloning kit. For SJ, single clones were subjected to PCR with TOPO TA primers and products were then digested with ApaL1 to detect precise joints. Plasmid DNA from resistant SJs clones, all CJs and interchromosomal rearrangement clones were prepared and Sanger sequenced.

Interchromosomal rearrangement primers 5'-3':

Dd2 F1 CAAGCATTAGACAGTAAGTACCCAG

Dd2 F2 GCCAACCACTTTGATAGTCTGTGGCTTG

Db1 R1 GAGTAATCGCTTTGTGTGCATCACA

Db1 R2 CATTCTGGATCTAAACACATCTAGGCTTGC

Vb10 SJ CTCAGTGAGACTCATCGGTGC

Db1.1 SJ CATTAGCTCGCATCTTACCAC

VhE GTGGAGTCTGGGGGAGGCTTA

Vh7183 CCAAGAAGACCCTGTACCTGCAAATGA

Jhintronic CTCCACCAGACCTCTCTAGACAGC

Jh4 TCAAATGAGCCTTCCAAAGTCC

Spectral karyotyping (SKY)

For metaphase preparations, primary tumor cells were grown in RPMI supplemented with 20% fetal bovine serum, L-Glutamine and beta

mercaptoethanol for 2 h and exposed to colcemid (0.025 µg/ml) and BrdU (28 ug/ml) for an additional 2 hours at 37 °C. Then, cells were incubated in 40 mM KCl for 25 min at 37 °C, fixed in fixative solution (75% methanol/25% acetic acid) and washed twice in the fixative. Cell suspensions were dropped onto pre-chilled glass slides and air-dried. Spectral karyotyping was performed using the Applied Spectral Imaging mouse SKY probe according to the manufacturer's instructions to determine chromosomal rearrangements. The slides were analysed using a Nikon Eclipse 80i microscope. SKY images were captured and karyotyped using an Applied Spectral Imaging system.

Adjacent Direct Repeat (ADR) Quantification

ADRs were scored in a fashion such that repeats, which were generated randomly, were removed. First we calculated the insertion probabilities of A,T,G,C nucleotides. For that purpose, we omitted any insertions that could have been generated as P nucleotides and calculated nucleotides probabilities for each TCR beta locus separately. Second, we calculated the probability of a certain ADR to arise by multiplying the relevant nucleotide probability times the number of total junctions (For example: $P_{(T)}*P_{(A)}*P_{(A)}*N$) for each locus. Any multiplication above 1 meant that, by chance, there is a likelihood of 1 junction to have a certain ADR. Therefore, to be stringent, we scored as statistically significant multiplication less than 0.5.

Whole genome sequencing

Detailed procedures can be found in ref 113. Genomic DNA from $Rag2^{FS/FS};p53^{-/-}$ tumor and liver (control) tissues was purified. Paired-end libraries were generated according to manufacturer's recommendations. Libraries were then analyzed for size distribution and sequenced on an Illumina HiSeq 2000. Sequence coverage was calculated by (# Reads adjusted to duplication * average insert size in bp)/bp in the mouse genome). We obtained 35 and 37 fold coverage for tumor and liver respectively. Filtering pipeline is described in ref 30. Potential genomic candidates were validated by PCR using custom designed primers against tail DNA. Lesions that were validated as tumor specific were cloned (TOPO TA) and Sanger sequenced.

Cryptic RSS (cRSS) Definition: To score for aberrant rearrangements involving cRSSs we used criteria that were generated in our lab (Mijuskovic M, in preparation). Basically, we identified the first 'CAC/GTG' from the break point and reconstructed both a 12/23 RSS. V(D)J recombination requires synapsis of 12/23 RSSs. Hence, we assessed our sequences from both sides of the break point in pairs. For example, if only **one** side of the break showed good identity to a consensus RSS we did not score this rearrangement as cRSS mediated. In addition we did not consider 'CAC' that where further than ~25bp from the break point.

V(D)J recombination assay

The 293T cell line was grown in DMEM supplemented with fetal bovine serum (10%), non-essential amino acids and penicillin-streptomycin. Cells were grown at 37 °C in the presence of 5% CO2. To assess V(D)J recombination, 0.5 µg of the indicated murine Rag1, murine Rag2 and recombination substrate were transfected into cells using a FuGENE 6:DNA ratio of 3:1. Forty-eight hours after transfection cells were harvested and fluorescent intensity was measured using BD LSR II for FACS readout.

Spectratyping

Immunoglobulin heavy chain repertoire analysis was performed using CDR3 spectratyping. Briefly, genomic DNA from splenocytes was purified using PureGene and amplified using the J606.1 and the J558.85 VH primers and a fluorescent labeled JH2 reverse primer. 2 µL of PCR product per reaction were resolved by capillary electrophoresis on an ABI 3100 analyzer (Applied Biosciences Inc). Peak scanner v. 1.0 was used to generate and analyze the spectratypes (Applied Biosciences Inc).

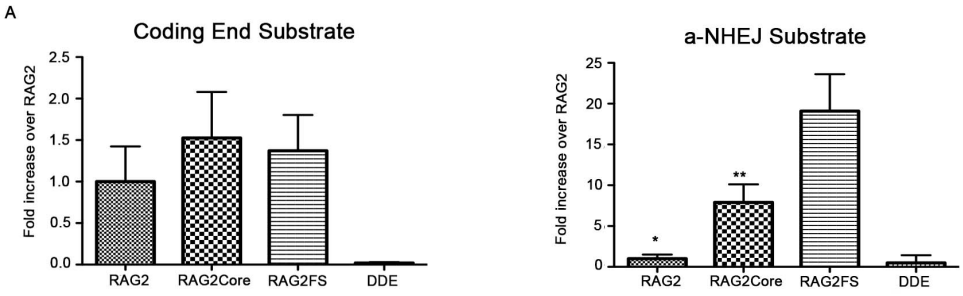
Statistical Analysis

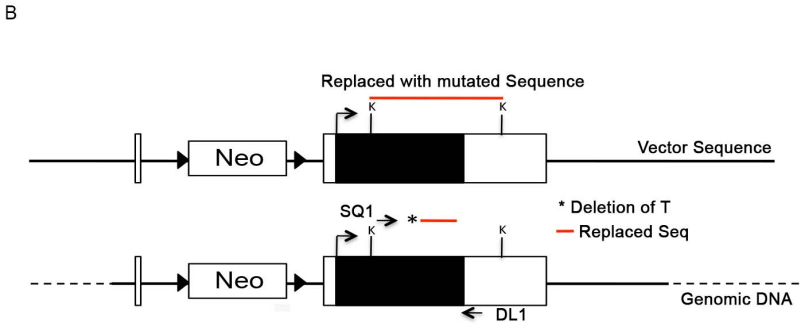
A two-tailed unpaired t-test was applied for extrachromosomal recombination efficiency and thymus cellularity. For all other statistical analysis we used the Chi-Square test of independence.

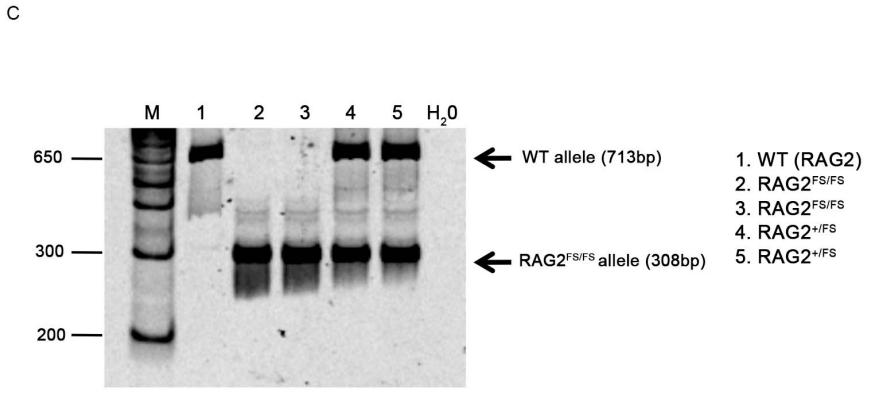
RESULTS

Signal joints from RAG2$^{FS/FS}$ mice show reported features of a-NHEJ

To explore pathway choice control at the chromosomal level in repair-proficient animals we generated FS361RAG2 knock-in mice (Fig. 2.1B-D). We chose this mutant because it yields the highest levels of a-NHEJ of any RAG mutant tested *in vitro* (Fig. 2.1A and 36,74). Specifically, this allele yields 2.5 to 5 fold higher levels of a-NHEJ than "core" RAG2 (truncated at amino acid 383; 114,115) (Fig.2.1A, 36). RAG2$^{FS/FS}$ knockin mice exhibit a mild block in early lymphocyte differentiation (Fig.2.2) closely resembling that described for core RAG2 mice (RAG2$^{C/C}$), and, as expected, remain capable of generating mature lymphocytes (112,116).


A
Coding End Substrate
Fold increase over RAG2
2.5
2.0
1.5
1.0
0.5
0.0
RAG2
RAG2Core
RAG2FS
DDE
a-NHEJ Substrate
Fold increase over RAG2
25
20
15
10
5
0
*
**
RAG2
RAG2Core
RAG2FS
DDE


B
Replaced with mutated Sequence
K
K
Neo
Vector Sequence
SQ1
*
* Deletion of T
Replaced Seq
K
K
Neo
Genomic DNA
DL1


C
M
1
2
3
4
5
H2O
650
300
200
WT allele (713bp)
RAG2FS/FS allele (308bp)
1. WT (RAG2)
2. RAG2FS/FS
3. RAG2FS/FS
4. RAG2+/FS
5. RAG2+/FS

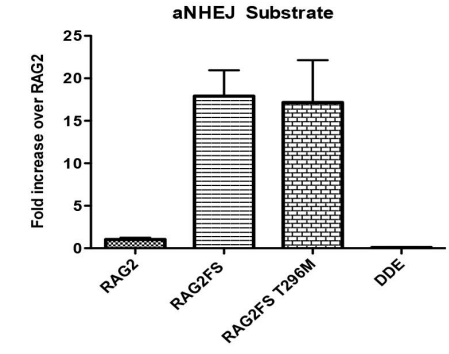
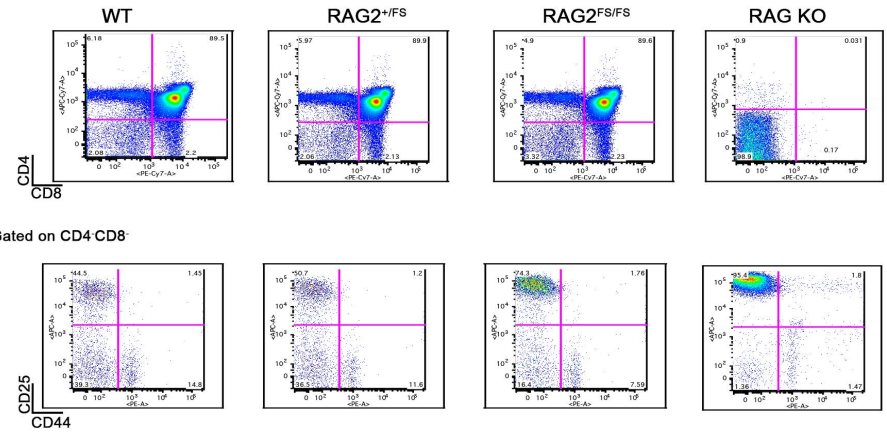

D
aNHEJ Substrate
Fold increase over RAG2
25
20
15
10
5
0
RAG2
RAG2FS
RAG2FS T296M
DDE

Figure 2.1 Generation of RAG2$^{FS/FS}$ knock-in mice. A. RAG2FS has the highest a-NHEJ activity in-vitro. 293T cells were transiently transfected with full length Rag1, the different Rag2 mutants and the indicated substrate. Recombination was measured 48h post transfection by FACS analysis. Calculations are fold increase over full length Rag2 (RAG2). A. Recombination with RSSs in deletion configuration; RAG2, RAG2Core (a.a1-383) RAG2FS - Frame-shift mutation at amino acid 361, DDE- inactive Rag1. Absolute recombination for RAG2 28%±12. B. Recombination with a-NHEJ substrate (15). * $p<0.0001$ vs RAG2, ** $p<0.001$ vs RAG2Core. RAG2 absolute recombination is 0.5%±0.25. B. Mice were generated in inGenious Targeting Laboratory Inc. A~11.5kb region used to generate the targeting vector was first sub cloned from a positively identified C57/Bl6 BAC clone. Two types of mutations were generated in exon 3 utilizing overlap extension PCR. The first mutation comprised deletion of base 1082 (T) to generate a frameshift. The second mutation is located 67bp 3' of the T deletion and comprised replacement of the last 435bp of coding sequence with the sequence- AAGCGGCCGCGACTCTAG followed by the 3' UTR sequence. First, primers located 5' and 3' to two unique Kpn1 (K) sites that flank the location of the mutations were used to amplify a 1.6kb fragment. The mutations were introduced into this fragment, which was then reintroduced back into the construct via ligation into the Kpn1 sites, thus replacing the wild-type Kpn1 fragment with the mutated Kpn1 fragment. The Neo cassette is inserted 228bps 5 to the ATG in exon 3 using Red/ET recombineering technology with a short homology arm that extends ~1.5kb 5' to the Neo cassette. The total size of the targeting construct (including the backbone vector) is 15.3kb. C. Identified F1 heterozygous mice were crossed with Ella-Cre transgenic mice to delete the floxed Neo cassette in vivo. An additional cross between the +/F mice and wild-type mice eliminated the transgene. Subsequent mating of Cre-F2 (RAG2+/FS) generated littermates with the wild-type (+/+), heterozygous (RAG2$^{+/FS}$) and homozygous (RAG2$^{FS/FS}$) genotypes for analysis. Genotyping analysis was done using the SQ1-GGAGACTCCTGACTGGACCCTCAG and DL1-GATTCAGAGAGCAATATACCT primers. D. Sequence analysis of tail and liver DNA showed an additional amino acid change at T296M that occurred during the targeting. Nevertheless, this change did not affect the FS mutant functionality in our cell system assay allowing us to still use it as a mouse model to investigate our questions.

A. Thymus



B.

Thymus	% CD4⁺CD8⁺	%CD4⁻CD8⁻	% CD44⁻CD25⁺	% CD44⁻CD25⁻
WT	88.37±1.59	1.85±0.46	38.92±11.81	46.10±11.73
RAG2$^{+/FS}$	87.74±2.96	2.44±0.45	42.06±12.31	42.63±14.48
RAG2$^{FS/FS}$	**88.20±2.25**	**3.02±0.63***	**65.09±6.23***	**22.42±6.67***
RAG$^{-/-}$	0.06±0.00	99.10±0.28	95.5±1.13	1.71±1.06

BM	% IgM⁺	% IgM⁻	% B220⁺CD43⁺	% B220⁺CD43⁻
WT	42.2±13.7	52.1±17.1	16.3±3.5	51.6±15.5
RAG2$^{+/FS}$	43.0±5.1	52.3±9.9	14.4±4.5	49.6±15.6
RAG2$^{FS/FS}$	**31.6±7.4**	**63.9±13.2**	**24.0±2.7***	**39.1±11.4**
RAG$^{-/-}$	1±0	99±0	16.8±1.9	1.7±0.25

Spleen	% CD4⁺	%CD8⁺	% B220⁺IgM⁺
WT	17.2±1.5	7.7±1.4	35±3.1
RAG2$^{FS/FS}$	**14.7±1.9**	**8.2±1.2**	**29.3±4.4**
RAG$^{-/-}$	2.5±1.7	0.4±0.4	0.06±0.09

Figure 2.2 T and B cell development in RAG2$^{FS/FS}$ knock-in mice. A. Thymocytes from mice of the indicated genotype were stained with antiCD8-PE-CY7, antiCD4-APC-CY7, antiCD25-APC and antiCD44-PE. B. Summary table of the different B and T cell populations from 6-8 week old mice n=7. RAG KO (RAG$^{-/-}$) are combined analysis of Rag1 and Rag2 KO n=4. B220+ cells in the BM are gated on IgM-. *p<0.05 vs wild-type.

Initially we examined signal joints, which are informative for two reasons. First, the sequence of these junctions is noncoding, and thus not subjected to selective pressures during lymphocyte differentiation. Second, signal joints have a well-defined structure, being formed by blunt ligation of the signal ends. In a signal joint the two RSSs abut and although the occasional insertion of nucleotides is seen, deletions in either RSS is rare. Hence, nucleotide sequence features considered characteristic of a-NHEJ (excessive deletions, long insertions, and junctional microhomologies) are readily identified.

Signal joints arising from recombination at the T cell receptor (TCR) beta locus in $RAG2^{FS/FS}$ mice exhibited a significant increase in imprecise joints (63/140, 45%) compared with age-matched wild-type controls (34/159, 21.4%) ($p<0.0001$, Table 1). $RAG2^{FS/FS}$ junction sequences showed a significant increase in deletions (32/63, 50.7%, vs 6/34, 17.6%, seen in junctions from wild-type mice, $p<0.001$; Table 1). Strikingly, at least 50% of deleted signal ends from $RAG2^{FS/FS}$ mice had deletions greater than 5 nucleotides, a feature not observed in over 150 signal joints from wild-type mice (Table 1, Fig.S1). Other sequence features considered characteristic of a-NHEJ not observed in junctions from wild-type mice included occasional microhomologies (5/140 junctions, ranging from 2-9bp), and large insertions (350bp and 26bp; Fig.S1). A similar trend was previously observed in mice bearing the less severely truncated core RAG2 allele (110) and also the RAG2 T490A allele (in which the protein degradation signal is ablated; 28)

signifying the importance of an intact RAG2 C terminus. In aggregate, these features indicate that signal ends are abnormally available to a-NHEJ in RAG2$^{FS/FS}$ mice, suggesting that repair pathway choice is disabled.

Table 1.

	Vβ14-Dβ1 Inv	Vδ5-dδ2 Inv	Vβ8.3-Dβ1.1 Del	Vβ10-Dβ1.1 Del
WT				
Precise joints	68/79 (86%)	19/28 (68%)	19/27 (70%)	19/25 (76%)
Imprecise joints	11/79 (14%)	9/28 (31%)	8/27 (30%)	6/25 (24%)
N addition	10/11 (91%)	9/9 (100%)	8/8 (100%)	6/6 (100%)
Deletion	0/11 (0%)	3/9 (30%)	3/8 (37.5%)	0/8 (0%)
%Deletions >5bp	0%	0%	0%	0%
Miscleavage	1/11 (9%)	0/9 (0%)	0/8 (0%)	0/6 (0%)
Microhomology≥2bp	0/11 (0%)	0/9 (0%)	0/8 (0%)	0/6 (0%)
RAG2$^{FS/FS}$				
Precise joints	47/70 (67%)	6/26 (22%)	15/28 (54%)	9/16 (56%)
Imprecise joints	23/70 (33%)§	20/26 (78%)§	13/28 (46%)	7/16 (44%)
N addition	16/23 (69%)*	14/20 (70%)	13/13 (100%)	6/7 (86%)
Deletion	9/23 (39%)	17/20 (85%)	4/13 (30%)	2/7 (29%)
%Deletions >5bp	10/14 (71%)	16/30 (53%)	4/4 (100%)	0%
Miscleavage	3/23 (13%)	4/20 (20%)**	5/13 (38%)	2/7 (29%)
Microhomology≥ 2bp	2/23 (8.5%)	2/20 (10%)	0/13 (0%)	1/7 (14%)

Table 1. RAG2$^{FS/FS}$ mice exhibit a-NHEJ at the signal joints. Sequences analyses from 2-3 independent mice were combined. Frequencies of N nt additions, deletions, miscleavage and microhomology were calculated out of imprecise junctions. Deletions>5bp is calculated out of total deleted ends (5bp deletion was the longest seen in wild-type and thus chosen as a cutoff). Miscleavage represnts coding end sequences at the ends, § $p<0.01$ vs wild-type; * one insertion was 350bp from Vb3 region (Fig. S1); ** 2 junctions underwent open-shut intermediate (indicated in blue, Fig.S1).

Coding joints from RAG2$^{FS/FS}$ mice fail to exhibit reported features of a-NHEJ

We next examined coding joints formed at immunoglobulin (Ig) and TCR loci. Unlike signal joints, nucleotide sequences of coding joints from the two models (RAG2$^{FS/FS}$, n=136; wild-type, n=94) were qualitatively similar (Table 2, Fig.S2). Because of the potential for bias imposed by biological selection for productive rearrangements (101), we also analyzed coding joints from (non-coding) D-J rearrangements and from sorted CD4/CD8 double negative thymocytes (which are not subject to selection for productive rearrangements; Table 2) Again, there was no qualitative difference between RAG2$^{FS/FS}$ and wild-type mice. Finally, we looked at the third complementarity-determining region (CDR3) sequence of antibody heavy chain gene rearrangements in genomic DNA of splenocytes from wild-type and RAG2$^{FS/FS}$ mice. The CDR3 is generated by V(D)J rearrangement and is influenced by non-templated additions and deletions. Significant shifts in CDR3 length, therefore, can serve as indirect evidence of a-NHEJ repair. CDR3 spectratyping of VH606 and VH558 rearrangements to JH2 from splenocytes revealed no significant differences (Fig.S2F).

We considered three reasons for the lack of distinctive sequence features at coding joints formed in RAG2$^{FS/FS}$ mice. One, the RAG2 FS allele might selectively enforce pathway choice for signal ends, but not for coding ends. Two, competition from c-NHEJ could render the "background" of normal coding joints too high to allow us to detect rare joints formed by a-NHEJ. Three, coding joints

formed by a-NHEJ may not be structurally distinctive. To explore these possibilities further, we used more sensitive assays to detect coding joints formed by a-NHEJ.

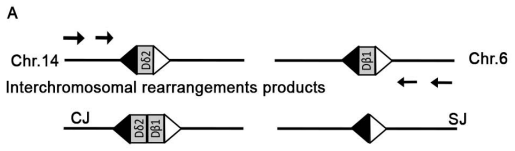
Table 2.

	Vβ6/7/8-Jβ2	Vβ10-Jβ2.1	Vβ14-Jβ1.1	Dβ2-Jβ2.6	Vh7183-Jh4
WT	**n=23**	**n=16**	**n=11**	**n=15**	**n=17**
Deletion in V region	80%/-3.6	100%/-4.2	36%/-3.5	69%/-2.2	82%/-2.92
>4 bp deletion	31%	37.5%	25%	20%	21%
Deletion in J region	95%/-4.4	94%/-3	63%/-5.4	87%/-5.38	94%/-5.93
>4 bp deletion	41%	27%	57%	46%	57%
RAG2$^{FS/FS}$	**n=31**	**n=32**	**n=14**	**n=17**	**n=21**
Deletion in V region	77.5%/-4.63	97%/-4.45	43%/-3.3	53%/-2.9	48%/-2.7
>4 bp deletion	38%	48%	33%	22%	20%
Deletion in J region	90%/-5.29	81%/-3.1	64%/-5	82%/-4.78	76%/-6.12
>4 bp deletion	57%	23%	56%	43%	56%

Table 2. No detectable a-NHEJ repair at antigen receptor coding joints in RAG2$^{FS/FS}$ mice. Sequence data from 2-3 independent thymic or BM genomic DNA samples were combined. The frequency of deleted ends at the V or J regions was from total events. Average of base pair (bp) deletion and frequency of >4bp deletion was calculated out of deleted events only (4bp was the average deleted length in wild-type junctions and thus chosen as a cutoff).

RAG2$^{FS/FS}$ mice show increased inter-chromosomal rearrangements between antigen receptors together with excessive deletions

a-NHEJ has been firmly implicated in chromosome translocations in various end joining-deficient backgrounds, and the translocation junctions show characteristic sequence features such as microhomologies and excessive deletions (9,46,37). We therefore investigated our RAG2$^{FS/FS}$ mice for elevated levels of translocations. In particular, abnormal translocations between the TCRβ and TCRδ loci, located on chromosomes 6 and 14, respectively, has previously been observed in mice bearing another RAG2 mutation (99). The D regions involved are flanked by a 12 and a 23 RSS which can lead to signal joints or coding joints products in a translocation junction (Fig.2.3A). We detected translocations in 4 out of 5 RAG2$^{FS/FS}$ mice, but not in thymocytes from wild-type controls (n=4, $p<0.02$). We observed excessive deletions in 6/10 junctions with evidence in one case for a 2 bp junctional microhomology (Fig.2.2B). Unexpectedly, none of the junctions retained the D region coding end sequences. Because of this we could not establish whether the extensively deleted translocation junctions were abnormal coding joints or signal joints. Nonetheless, the observation of chromosomal translocations suggests that pathway choice is defective in the RAG2 mutant mice, in agreement with our analysis of signal joints.



B

Dδ2		Dβ1
GGTTTTTGCAAAGCTCTGTAGCACCGTG		CACGGTGATTCAATTCTATGGGAAGCCTTTACAAAAACC
GGTTTTTGCAAAGCTCTGTAGCACCGTG		CACGGTGATTCAATTCTATGGGAAGCCTTTACAAAAACC
GGTTTTTGCAAAGCTCTGTAGCACCGTG	TCC	CACGGTGATTCAATTCTATGGGAAGCCTTTACAAAAACC
GGTTTTTGCAAAGCTCTGTAGCACCG(-2)	CGTCC	CACGGTGATTCAATTCTATGGGAAGCCTTTACAAAAACC
GGTTTTTGCAAAGCTCTGTAGCACCG(-2)	CG	CACGGTGATTCAATTCTATGGGAAGCCTTTACAAAAACC
GGTTTTTGCAAAG(-15)	G	GATTCAATTCTATGGGAAGCCTTTACAAAAACC(-6)
GGTTTTTGCAA(-17)	TAGG	GATTCAATTCTATGGGAAGCCTTTACAAAAACC(-6)
GG***TT***(-24)		CTGT(-42)
CCCA(-33)		CAATTCTATGGGAAGCCTTTACAAAAACC(-10)
AAAG(-37)	GCTG	TGGGAAGCCTTTACAAAAACC(-18)
attccaaggttt		GGAAGCCTTTACAAAAACC(-19)

Figure 2.3 RAG2$^{FS/FS}$ interchromosomal rearrangements within the antigen receptor loci.

A. Scheme of the germ-line configuration of TCR beta and delta and the predicted products upon interchromosomal rearrangements. Rectangles represent coding gene segments; triangles represent RSS (Filled -12RSS, open - 23 RSS). The arrows indicate nested PCR primers. B. Sequence analysis of purified PCR products (n=4 RAG2$^{FS/FS}$ mice). Capital letters at the middle of the junction represents N nt, Bold italic are microhomology, deletions are indicated in parentheses and small letters are Sanger sequence that did not align to mouse mm9 database.

Coding joint formation bypasses Ku80 deficiency in RAG2$^{FS/FS}$ mice

Ku80$^{-/-}$ mice lack a critical component of c-NHEJ, and are thus severely deficient for both coding and signal joints, leading to a complete block in lymphocyte differentiation at the proB/proT stage (108). We reasoned that if pathway choice were no longer imposed by C-terminally truncated RAG2, the joining of V(D)J recombination intermediates, no longer restricted to c-NHEJ only, might become possible. In other words, lack of joining in Ku80-null animals might be rescued by

an alternative pathway in Ku80/RAG2FS double mutants. Furthermore, nucleotide sequence analysis of the V(D)J junctions formed in these mice (in the absence of c-NHEJ) should help interpreting the coding joints formed in $RAG2^{FS/FS}$ mice.

$RAG2^{FS/FS}$;$Ku80^{-/-}$ double mutants demonstrated a significant (~ 4 fold, $p<0.05$) increase in thymus cellularity compared with $Ku80^{-/-}$ mice, suggesting a partial bypass of the developmental block (Fig.2.4A). FACS analysis showed that T cells progressed into the CD4/CD8 double positive (DP) stage in over half of the animals (10/17 $p<0.05$, Fig.2.4B-C). As expected, $Ku80^{-/-}$ mice lacked CD4/CD8 positive cells except for one mouse that had a very small DP population (Fig.2.4C, 108). Because developmental progression is linked to successful V(D)J recombination, these indications invited further examination of the possibility that the RAG2FS allele rescued TCR beta rearrangement.

We detected cell surface TCR beta expression in one double mutant mouse (Fig.2.4B), and demonstrated V(D)J joining via PCR analyses in several other $RAG2^{FS/FS}$;$Ku80^{-/-}$animals. We detected no TCR rearrangements in $Ku80^{-/-}$ mice, in accordance with previous work (108). In the double mutants we observed TCR beta and alpha rearrangements (Vβ14, Vβ10, and Vα8) (Fig.2.5). DNA sequence analysis revealed complete (V to D to J) and diverse coding joints. The FS allele can, therefore, substantially rescue joining of coding ends and bypass

the joining defect that is otherwise observed when Ku80 is not present. The joints we observed were formed in the absence of c-NHEJ, and are therefore generated, by definition, by a-NHEJ. We conclude that the C-terminal RAG2 truncation ablates pathway choice control, allowing coding ends to be joined via a-NHEJ.

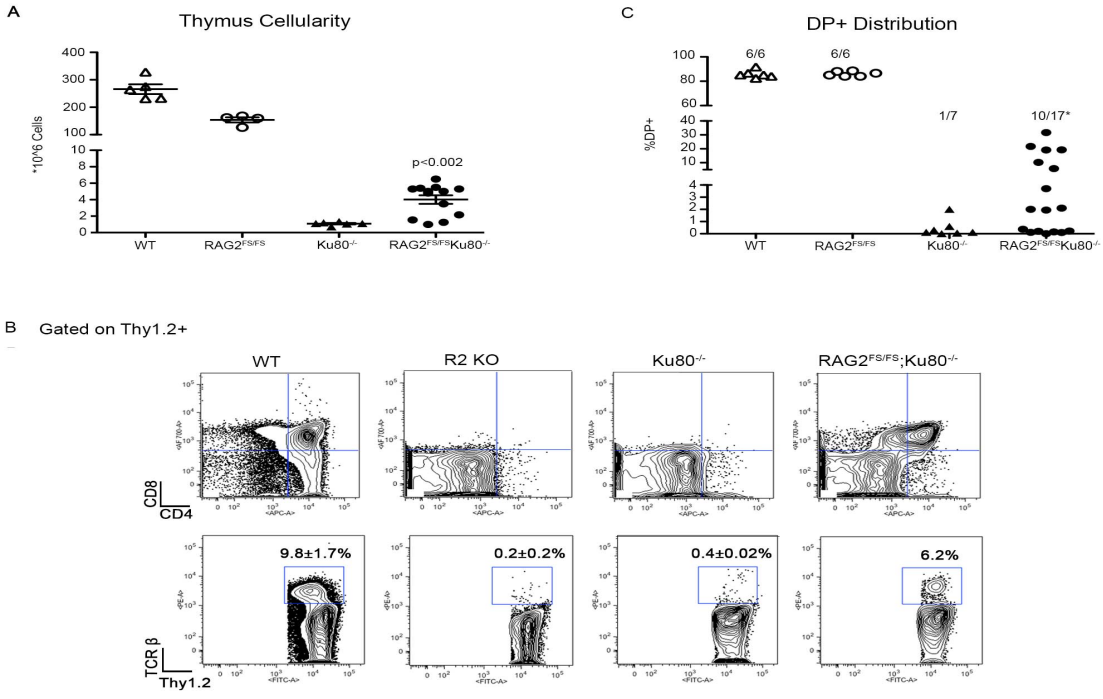
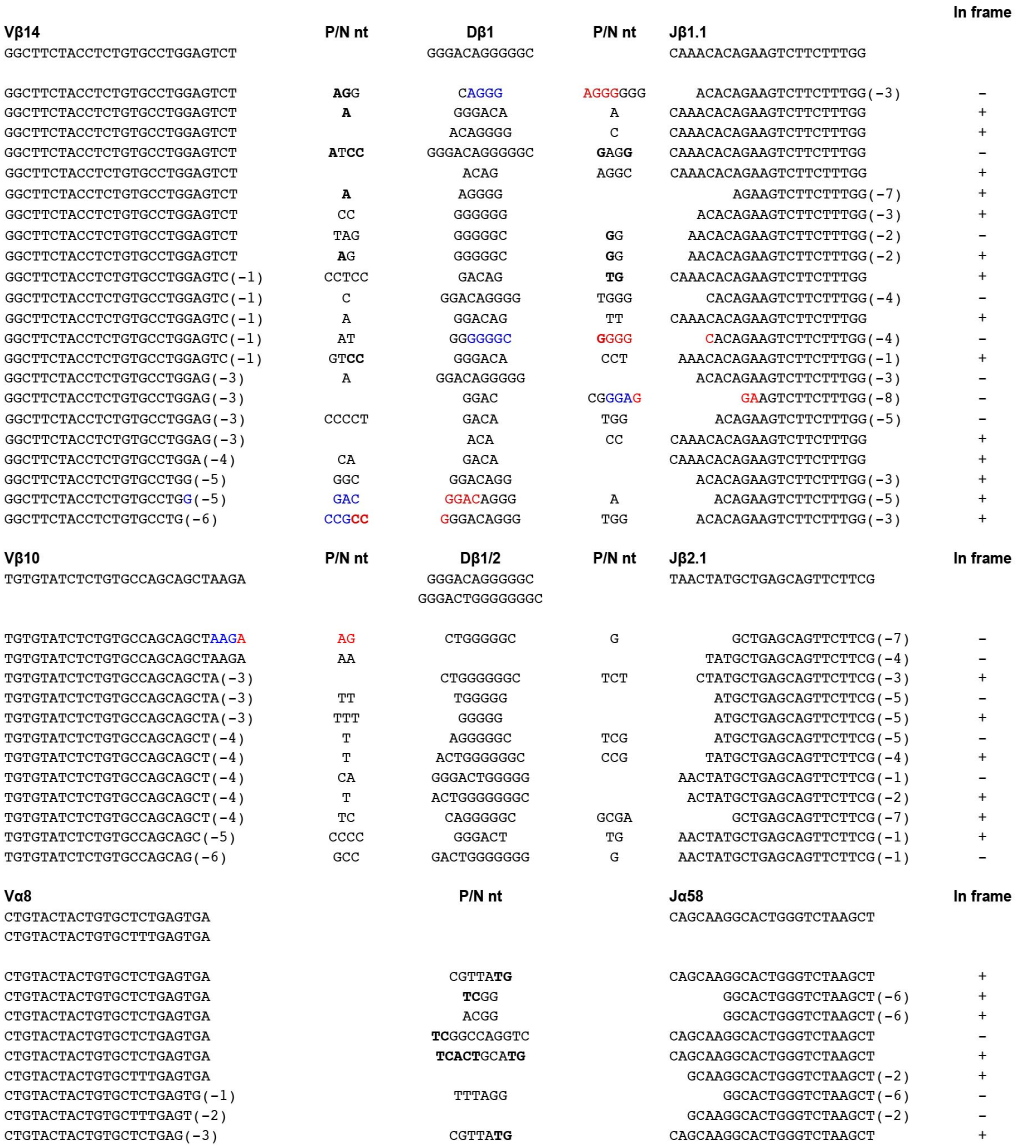


Figure 2.4 RAG2$^{FS/FS}$ mutant can by-pass Ku80 deficiency. Mice from the indicated genotypes were sacrified at 4 weeks for analysis. A. Thymus cellularity by trypan blue. p vs Ku80$^{-/-}$ B. Thymocytes were stained with antiThy1.2+FITC, antiCD4-APC, antiCD8-AF700 and antiTCR beta-PE. Double positive (DP) populations and TCR beta were calculated by gating on Thy1.2+ cells. C. Distribution of DP cells; only DP >1% was considered as positive. Number of mice analyzed is indicated. *p <0.05 vs Ku80$^{-/-}$



Vβ14	P/N nt	Dβ1	P/N nt	Jβ1.1	In frame
GGCTTCTACCTCTGTGCCTGGAGTCT		GGGACAGGGGGC		CAAACACAGAAGTCTTCTTTGG	
GGCTTCTACCTCTGTGCCTGGAGTCT	**AG**G	C**A**GGG	AGGGGGG	ACACAGAAGTCTTCTTTGG(-3)	–
GGCTTCTACCTCTGTGCCTGGAGTCT	**A**	GGGACA	A	CAAACACAGAAGTCTTCTTTGG	+
GGCTTCTACCTCTGTGCCTGGAGTCT		ACAGGGG	C	CAAACACAGAAGTCTTCTTTGG	+
GGCTTCTACCTCTGTGCCTGGAGTCT	**ATCC**	GGGACAGGGGGC	**GAGG**	CAAACACAGAAGTCTTCTTTGG	–
GGCTTCTACCTCTGTGCCTGGAGTCT		ACAG	AGGC	CAAACACAGAAGTCTTCTTTGG	+
GGCTTCTACCTCTGTGCCTGGAGTCT	**A**	AGGGG		AGAAGTCTTCTTTGG(-7)	+
GGCTTCTACCTCTGTGCCTGGAGTCT	CC	GGGGGG		ACACAGAAGTCTTCTTTGG(-3)	+
GGCTTCTACCTCTGTGCCTGGAGTCT	TAG	GGGGGC	**GG**	AACACAGAAGTCTTCTTTGG(-2)	–
GGCTTCTACCTCTGTGCCTGGAGTCT	**AG**	GGGGGC	**GG**	AACACAGAAGTCTTCTTTGG(-2)	+
GGCTTCTACCTCTGTGCCTGGAGTC(-1)	CCTCC	GACAG	**TG**	CAAACACAGAAGTCTTCTTTGG	+
GGCTTCTACCTCTGTGCCTGGAGTC(-1)	C	GGACAGGGG	TGGG	CACAGAAGTCTTCTTTGG(-4)	–
GGCTTCTACCTCTGTGCCTGGAGTC(-1)	A	GGACAG	TT	CAAACACAGAAGTCTTCTTTGG	+
GGCTTCTACCTCTGTGCCTGGAGTC(-1)	AT	GG**GGGGC**	**GGGG**	CACAGAAGTCTTCTTTGG(-4)	–
GGCTTCTACCTCTGTGCCTGGAGTC(-1)	GT**CC**	GGGACA	CCT	AAACACAGAAGTCTTCTTTGG(-1)	+
GGCTTCTACCTCTGTGCCTGGAG(-3)	A	GGACAGGGGG		ACACAGAAGTCTTCTTTGG(-3)	–
GGCTTCTACCTCTGTGCCTGGAG(-3)		GGAC	CGGGAG	GAAGTCTTCTTTGG(-8)	–
GGCTTCTACCTCTGTGCCTGGAG(-3)	CCCCT	GACA	TGG	ACAGAAGTCTTCTTTGG(-5)	–
GGCTTCTACCTCTGTGCCTGGAG(-3)		ACA	CC	CAAACACAGAAGTCTTCTTTGG	+
GGCTTCTACCTCTGTGCCTGGA(-4)	CA	GACA		CAAACACAGAAGTCTTCTTTGG	+
GGCTTCTACCTCTGTGCCTGG(-5)	GGC	GGACAGG		ACACAGAAGTCTTCTTTGG(-3)	+
GGCTTCTACCTCTGTGCCTGG(-5)	GAC	GGACAGGG	A	ACAGAAGTCTTCTTTGG(-5)	+
GGCTTCTACCTCTGTGCCTG(-6)	CCG**CC**	GGGACAGGG	TGG	ACACAGAAGTCTTCTTTGG(-3)	+

Vβ10	P/N nt	Dβ1/2	P/N nt	Jβ2.1	In frame
TGTGTATCTCTGTGCCAGCAGCTAAGA		GGGACAGGGGGC GGGACTGGGGGGGC		TAACTATGCTGAGCAGTTCTTCG	
TGTGTATCTCTGTGCCAGCAGCTAAGA	AG	CTGGGGGC	G	GCTGAGCAGTTCTTCG(-7)	–
TGTGTATCTCTGTGCCAGCAGCTAAGA	AA			TATGCTGAGCAGTTCTTCG(-4)	–
TGTGTATCTCTGTGCCAGCAGCTA(-3)		CTGGGGGGC	TCT	CTATGCTGAGCAGTTCTTCG(-3)	+
TGTGTATCTCTGTGCCAGCAGCTA(-3)	TT	TGGGGG		ATGCTGAGCAGTTCTTCG(-5)	–
TGTGTATCTCTGTGCCAGCAGCTA(-3)	TTT	GGGGG		ATGCTGAGCAGTTCTTCG(-5)	+
TGTGTATCTCTGTGCCAGCAGCT(-4)	T	AGGGGGC	TCG	ATGCTGAGCAGTTCTTCG(-5)	–
TGTGTATCTCTGTGCCAGCAGCT(-4)	T	ACTGGGGGGC	CCG	TATGCTGAGCAGTTCTTCG(-4)	+
TGTGTATCTCTGTGCCAGCAGCT(-4)	CA	GGGACTGGGGG		AACTATGCTGAGCAGTTCTTCG(-1)	–
TGTGTATCTCTGTGCCAGCAGCT(-4)	T	ACTGGGGGGGC		ACTATGCTGAGCAGTTCTTCG(-2)	+
TGTGTATCTCTGTGCCAGCAGCT(-4)	TC	CAGGGGGC	GCGA	GCTGAGCAGTTCTTCG(-7)	+
TGTGTATCTCTGTGCCAGCAGC(-5)	CCCC	GGGACT	TG	AACTATGCTGAGCAGTTCTTCG(-1)	+
TGTGTATCTCTGTGCCAGCAG(-6)	GCC	GACTGGGGGGG	G	AACTATGCTGAGCAGTTCTTCG(-1)	–

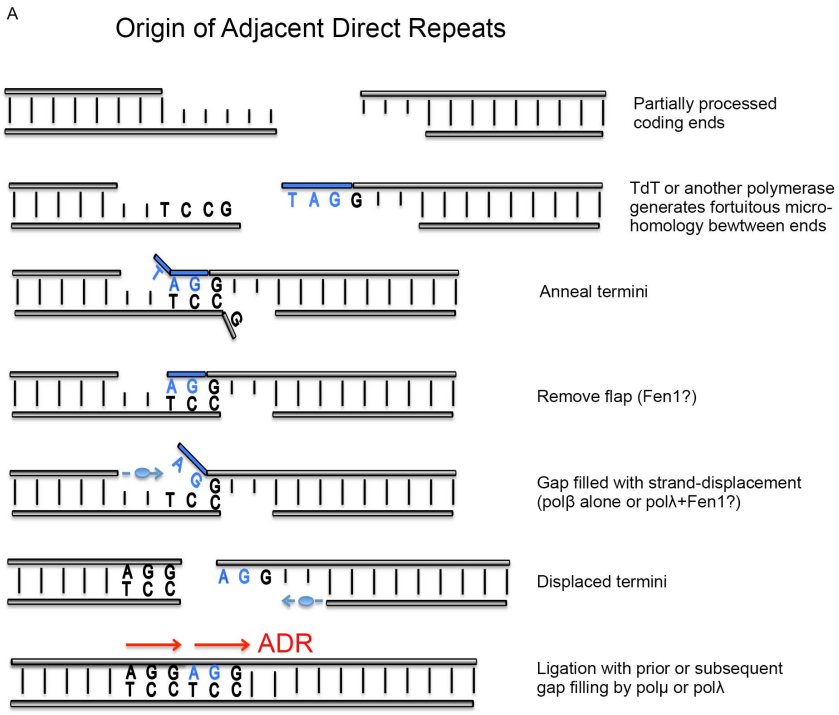
Vα8	P/N nt	Jα58	In frame
CTGTACTACTGTGCTCTGAGTGA CTGTACTACTGTGCTTTGAGTGA		CAGCAAGGCACTGGGTCTAAGCT	
CTGTACTACTGTGCTCTGAGTGA	CGTTA**TG**	CAGCAAGGCACTGGGTCTAAGCT	+
CTGTACTACTGTGCTCTGAGTGA	**TC**GG	GGCACTGGGTCTAAGCT(-6)	+
CTGTACTACTGTGCTCTGAGTGA	ACGG	GGCACTGGGTCTAAGCT(-6)	+
CTGTACTACTGTGCTCTGAGTGA	**TC**GGCCAGGTC	CAGCAAGGCACTGGGTCTAAGCT	–
CTGTACTACTGTGCTCTGAGTGA	**TCACT**GCA**TG**	CAGCAAGGCACTGGGTCTAAGCT	+
CTGTACTACTGTGCTTTGAGTGA		GCAAGGCACTGGGTCTAAGCT(-2)	+
CTGTACTACTGTGCTCTGAGTG(-1)	TTTAGG	GGCACTGGGTCTAAGCT(-6)	–
CTGTACTACTGTGCTTTGAGT(-2)		GCAAGGCACTGGGTCTAAGCT(-2)	–
CTGTACTACTGTGCTCTGAG(-3)	CGTTA**TG**	CAGCAAGGCACTGGGTCTAAGCT	+

Figure 2.5 Sequence analysis of TCR beta and alpha junctions from RAG2$^{FS/FS}$ Ku80$^{-/-}$ mice. Genomic DNA was prepared from RAG2$^{FS/FS}$;Ku80$^{-/-}$ thymocytes (n=2-4). TCR Vβ14-Jβ1.1, TCR Vβ10-Jβ2.1 and TCR Vα8-Jα58. Germline sequences are indicated at the top. Capital letters in the middle of the junction indicate N nt, capital bold - P nt, deletions are in parentheses and blue/red represents the Adjacent Direct Repeats (ADRs). +/- indicates in/out of frame rearrangements respectively.

Unusual characteristics of coding joints in RAG2FS/Ku80 double mutants

The rescued, Ku80-independent, coding joints lacked features of a-NHEJ that have been reported in other systems (49,117). As in the RAG2$^{FS/FS}$ mice, we saw neither a dependence upon microhomology nor unusually large deletions (Fig.2.5). These data suggest that, in the context of coding joint formation, a-NHEJ does not display attributes commonly observed in other systems. We did, however, observe an unusual feature: short, 3-5 nucleotide repeats were evident in many of the junctions formed at the Vβ14 locus (Fig.2.5) in double mutant mice. By a conservative analysis designed to minimize counting repeats generated by chance (see Methods), 5/22 (23%) Vβ14 junctions contained "Adjacent Direct Repeats" (ADRs). The sequence features of ADRs (being in direct orientation and immediately adjacent) imply a mechanism in which processed ends with complementary extensions (acquired through addition of P or N nucleotides) are annealed and then displaced by a gap-filling polymerase before ligation (Fig.2.6A). Efficient strand displacement is not a known characteristic of either polλ or polμ (118,119) which, along with TdT, are the polymerases thought to be associated with c-NHEJ (8). Pol β, another member of the polX polymerase family, has gap-filling and strong strand-displacement properties (118) and while active in base excision repair (120), has not been described as a factor in c-NHEJ.

A



B

	Genotype	Vβ14-Jβ1.1	Vβ10-Jβ2.1	Vβ6/7/8-Jβ2	Dβ2-Jβ2.6
Talukder SR et al (25)	WT	3/40	1/30	nd	nd
Zhang L et al (Supp Info, 23)	WT	3/47	nd	nd	nd
Gigi V et al (This study)	WT	1/11	0/16	0/23	0/15
Gigi V et al (This study)	$RAG2^{FS/FS}$	1/14	0/32	2/31	2/17
Gigi V et al (This study)	$RAG2^{FS/FS}$; $Ku80^{-/-}$	5/22	1/12	nd	nd

Figure 2.6 Generation of Adjacent Direct Repeats -ADRs

A. Scheme showing how small ADRs (adjacent direct repeats) might be generated. The first step may or may not be a mechanistic constraint however it is necessary in the identification of ADRs. We show homologies between ends having been generated by TdT or perhaps another polymerase (121). Terminal homologies can also be revealed between two ends by resection but were excluded in our analysis because prior existing sequences between ends that can generate ADRs are indistinguishable from simple direct joining products and difficult to score. The next step is stabilizing end-to-end interaction via complementary bases in the two single strand extensions compensating for the lack of Ku80. We suggest that annealing occurs near or at the termini. For the sake of parsimony, we have depicted Fen-1 as the flap removal factor prior to the gap-filling and strand displacement steps. ADRs are then generated when a gap-filling polymerase with strand displacement activity is present. This could be supplied by pol lamda (119), which has sufficient strand displacement activity in vitro with the cooperation of Fen-1, to invade three or four bp into a duplex, or by pol beta, which display stronger strand displacement on its own (118). The latter is not a known participant in c-NHEJ, but is also stimulated by Fen-1 *in vitro* (119). Following polymerization/strand displacement of one end the second detached end could be either filled in by 'conventional' extension and then ligated or alternatively, ligated and then filled in.

B. Summary of the number of junctions with ADRs detected in our study and by re-examination of published data (28,110). The ADRs are indicated in Blue-Red in Fig S2. nd stands for Not Done.

The high prevalence of ADRs at the Vβ14 locus prompted us to re-examine coding joints from this locus in wild-type RAG2 and RAG2$^{FS/FS}$ mice. We found that 1/11 TCR rearrangements from wild-type RAG2 and 1/14 from RAG$^{FS/FS}$ mice exhibit this feature (Fig.S2C). We also detected similar levels of ADRs upon review of published coding joint sequences at Vβ14 collected from wild-type mice (3/47 and 3/40; refs 28 and 110, respectively, Fig.2.6B). Although initially observed at Vβ14, ADRs can also be detected elsewhere (Vβ10, Vβ6-8, Dβ2; Fig.2.6B). The frequency of ADRs is increased in the absence of Ku80, implying an association with a-NHEJ. However, like other reported a-NHEJ sequence features (35), they are also observed in junctions from c-NHEJ-proficient animals.

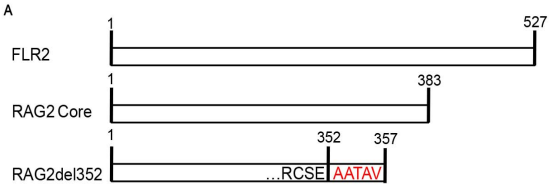
Together, these data allow us to draw the following conclusions. Ends that are formed in the absence of RAG2's C-terminus in c-NHEJ-proficient mice are accessible to a-NHEJ, as shown by signal joints bearing features characteristic of a-NHEJ, and also by the coding joints that are rescued in Ku80-deficient mice. Thus, the C-terminus of RAG2 is important for restricting the end-joining pathways that can repair RAG-generated DSBs *in vivo*. Additionally, our results suggest that Ku80-independent repair is not a disordered, unregulated alternative to c-NHEJ, because rescued coding joints exhibit uniform structural features, resembling junctions formed in RAG2 wild-type and RAG2$^{FS/FS}$ mice.

A more extensive RAG2 C-terminal truncation also generates aberrant V(D)J joints and rescues coding joint formation in Ku80-deficient mice

In the FS allele, the normal RAG2 sequence truncates at amino acid 361, but is followed by a novel stretch of 28 amino acids at its C terminus before the protein sequence terminates at a fortuitous stop codon (36). Through whole genome sequencing, we discovered that the core RAG2 allele in the $RAG2^{C/C}$ mice (derived in the laboratory of M. Schlissel) is not as reported (112) and is a more extensive truncation that terminates at amino acid 352 (instead of the reported 383), with 5 extra C-terminal amino acids encoded by the targeting vector (Fig.2.7A). We used mice homozygous for this allele, which we rename RAG2del352, to verify the results obtained with the FS allele.

We analyzed signal joints and interchromosomal rearrangements from homozygous $RAG2^{del352/del352}$ mice and TCR rearrangements from $RAG2^{del352/del352}$;$Ku80^{-/-}$ homozygotes and found results similar to those obtained with RAG2 FS allele (Fig.2.7B-D). These data extend the results obtained with $RAG2^{FS/FS}$ mice, indicating that the phenotype is not an artifact of the C-terminal extension encoded by the FS allele, and underscore the importance of the RAG2 C terminus in repair pathway choice *in vivo*.

A



B

	Vβ14-Dβ1 inversion	Vδ5-dδ2 inversion	Vβ8.3-Dβ1.1 Deletion
Precise joints	60/85 (70%)	5/13 (38%)	7/30 (23%)
imprecise joints	25/85 (30%)	8/13 (58%)	23/30 (77%)
N addition	20/25 (80%)	6/8 (75%)	16/23 (69%)
Deletion	5/25 (20%)	8/8 (100%)	21/23 (91%)
%Deletions >5bp	4/8 (50%)	6/12 (50%)	20/22 (91%)
Miscleavage	3/25 (12%)	1/8 (12.5%)	19/23 (83%)
Microhomology≥2bp	0/25 (0%)	2/8 (25%)	0/23 (0%)

C

Dδ2 **Dβ1**

```
GGTTTTTGCAAAGCTCTGTAGCACCGTG                CACGGTGATTCAATTCTATGGGAAGCCTTTACAAAAACC
GGTTTTTGCAAAGCTCTGTAGCACCGTG                CACGGTGATTCAATTCTATGGGAAGCCTTTACAAAAACC
GGTTTTTGCAAAGCTCTGTAGCA(-5)       ACA                                  CACAGTCCTC(-59)
GGTTTTTGCAAAGCTCTGTAGC(-6)        CCATCGGC     CGGTGATTCAATTCTATGGGAAGCCTTTACAAAAACC(-2)
GGTTTTTGCAAAGCTCTGTAG(-7)         CCTGC        CGGTGATTCAATTCTATGGGAAGCCTTTACAAAAACC(-2)
GGTTTTTGCAAGCTCTGTA(-8)                                          cgccatcngccc
GGTTT(-23)                                                       ctgacattaggt
tgccgattattc                                          CAATTCTATGGGAAGCCTTTACAAAAACC(-10)
```

D

Vβ14	P/N nt	Dβ1	P/N nt	Jβ1.1	In frame
GGCTTCTACCTCTGTGCCTGGAGTCT		GGGACAGGGGGC		CAAACACAGAAGTCTTCTTTGG	
GGCTTCTACCTCTGTGCCTGGAGTCT	**A**	ACAGGGGGC	C	CAGAAGTCTTCTTTGG(-6)	+
GGCTTCTACCTCTGTGCCTGGAGTCT		ACAGGGG	AGGG	AACACAGAAGTCTTCTTTGG(-2)	-
GGCTTCTACCTCTGTGCCTGGAGTCT	**AG**TC	CAGGGGG		CAAACACAGAAGTCTTCTTTGG	+
GGCTTCTACCTCTGTGCCTGGAGTCT	**A**	GGACAG		AACACAGAAGTCTTCTTTGG(-2)	+
GGCTTCTACCTCTGTGCCTGGAG(-3)	GATCCT**CC**	GGGACAG		ACACAGAAGTCTTCTTTGG(-4)	-
GGCTTCTACCTCTGTGCCT(-4)		***GGA***CAGGGGGC	T	CAAACACAGAAGTCTTCTTTGG	+
GGCTTCTACCTCTGTGCCTGGA(-4)	C**CCC**	GGGACA	A	CAAACACAGAAGTCTTCTTTGG	-
GGCTTCTACCTCTG(-12)				ACAGAAGTCTTCTTTGG(-5)	-

Vβ10	P/N nt	Dβ1/2	P/N nt	Jβ2.1	In frame
TGTGTATCTCTGTGCCAGCAGCTAAGA		GGGACAGGGGGC		TAACTATGCTGAGCAGTTCTTCG	
		GGGACTGGGGGGGC			
TGCTGTGTATCTCTGTGCCAGCAGC(-5)	CCT	GGGACTGGGG		TAACTATGCTGAGCAGTTCTTCG	-
TGCTGTGTATCTCTGTGCCAGCAG(-6)	GT**CC**	GGGACAGGGGG	GG	ATGCTGAGCAGTTCTTCG(-5)	+
TGCTGTGTATCTCTGTGCCAGCAG(-6)	GCC	GACTGGGGGGG	GG	AACTATGCTGAGCAGTTCTTCG(-1)	+
TGCTGTGTATCTCTGTGCCAGCAG(-6)		ACTGGGGGGG		CTATGCTGAGCAGTTCTTCG(-3)	-
TGCTGTGTATCTCTGTGCCAGCAG(-6)	AG	GGGGGGGC	**GC**	CTATGCTGAGCAGTTCTTCG(-3)	+

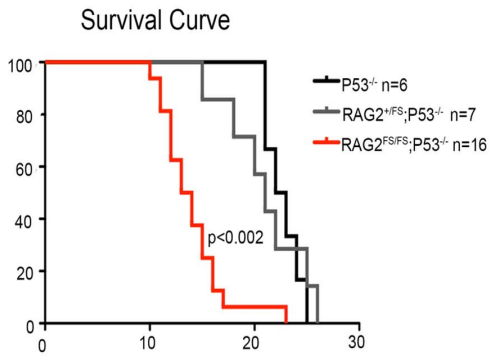
Vα8	P/N nt	Jα58	
CCCTGTACTACTGTGCTCTGAGTGA		CAGCAAGGCACTGGGTCTAAGCT	In frame
CTGTACTACTGTGCTTTGAGTGA			
CTGTACTACTGTGCTCTGAGTGA	**TCA**GGGT	GGCACTGGGTCTAAGCT(-6)	+
CTGTACTACTGTGCTTTGAGTGA	**T**GCCTTC**G**	CAGCAAGGCACTGGGTCTAAGCT	-
CTGTACTACTGTGCTTTGAGTGA	GGG	AAGGCACTGGGTCTAAGCT(-4)	-
CTGTACTACTGTGCTCTGAGTG(-1)	TCCTC**G**	CAGCAAGGCACTGGGTCTAAGCT	-
CTGTACTACTGTGCTCTGAGTG(-1)	TAC**G**	CAGCAAGGCACTGGGTCTAAGCT	-
CTGTACTACTGTGCTTTGAGTG(-1)	CGA**G**	CAGCAAGGCACTGGGTCTAAGCT	-
CTGTACTACTGTGCTTTGAGT(-2)		AGCAAGGCACTGGGTCTAAGCT(-1)	-
CTGTACTACTGTGCTCTGAG(-3)	CGTTA	CAGCAAGGCACTGGGTCTAAGCT	-
CTGTACTTCTGTGCTTTG(-5)	GATACT	CAGCAAGGCACTGGGTCTAAGCT	+

Figure 2.7 RAG2del352 mice junctions' analysis. A. Diagram of the RAG2del352 allele. A change in RAG2 sequence after amino acid 352 originating from the targeting vector is indicated in red. B. Signal joint analysis from different TCR loci n=2-6. (Analysis was preformed as specified in Table 1). C. Dδ2-Dβ1 interchromosomal rearrangements from healthy thymocytes n=3 D. RAG2$^{del352/del352}$;Ku80$^{-/-}$ antigen receptor rearrangements from the indicated loci n=1-3. (Annotations are as given in Fig.2.5).

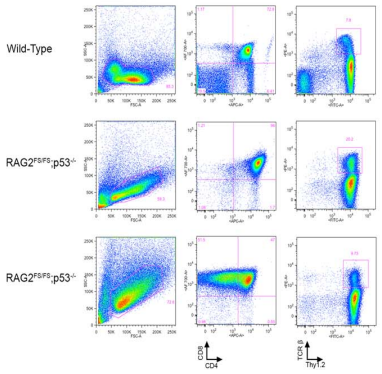
a-NHEJ in aberrant genomic rearrangements from a RAG2$^{FS/FS}$;p53$^{-/-}$ lymphoma

Several laboratories have reported a connection between a-NHEJ, chromosomal translocations, and accelerated lymphomagenesis (9,10,46,37). Indeed, c-NHEJ knock-out mice crossed to a p53-deficient background develop lymphomas with chromosome translocations mediated by a-NHEJ (9,10). To test the possibility that structural features reported for a-NHEJ might exist in genomic lesions from lymphomas, we generated RAG2$^{FS/FS}$;p53$^{-/-}$ mice. Thymic lymphomas emerged rapidly, with median survival of 13.5 weeks (vs. 22.5 weeks in p53$^{-/-}$ mice) ($p<0.002$, Fig.2.8A). Tumor cells expressed surface CD4 and CD8 with variable amounts of surface TCR beta, implying that these lymphomas originated from immature thymocytes (Fig.2.8B). Spectral karyotyping (SKY) analysis of these thymomas showed a wide spectrum of chromosome translocations, including but not limited to chromosomes bearing antigen receptor loci (Fig.2.8C).

A



B



C

RAG2$^{FS/FS}$ P53$^{-/-}$	Translocations	Frequency
4228	t(7;8) t(5;5) t(19;10)	5/5 3/5 2/5
4222	t(6;7)* t(12;16)* t(14;8)* t(1;1)	6/6 4/6 1/6 1/6
4208	t(9;17) t(1;1) t(4;19) t(19;18)	12/13 2/13 1/13 1/13
13422	t(12;9)* t(12;1)*	11/13 2/13
13508§	t(14;1)* t(12;14)* t(10;16) reciprocal* t(16;1)* t(14;16)* t(9;12)* t(5;4) t(8;6)* t(2;4)	16/20 16/20 1/20 1/20 1/20 1/20 1/20 1/20 1/20

Figure 2.8 RAG2$^{FS/FS}$;P53$^{-/-}$ mice develop rapid lymphoma with genomic instability. A. Kaplan-Meier survival curve of the indicated genotyped. p vs P53$^{-/-}$ mice. B. FACS analysis of representative wild-type thymus and RAG2$^{FS/FS}$;p53$^{-/-}$ thymic lymphomas. C. . SKY analysis of RAG2$^{FS/FS}$;P53$^{-/-}$ T cell lymphomas. * indicates antigen receptor chromosome

To determine whether sequence features of a-NHEJ might be apparent in genomic lesions, we employed our previously established analysis pipeline to detect structural variants (translocations, deletions, inversions, and duplications) in paired-end whole genome sequence data (113). We analyzed two $RAG2^{FS/FS};P53^{-/-}$ tumors and identified 47 genomic rearrangements including translocations and intra-chromosomal rearrangements (deletions, inversions and duplications, Table 3). 14 rearrangements were the result of normal antigen receptor junctions. The remaining 33 rearrangements were aberrant junctions: 20 deletions, 9 inversions, 2 duplications and 2 translocations. Both of the detected translocations were between chromosome pairs previously identified as having undergone an exchange by SKY analysis of this same tumor (Fig.2.8C)

16 aberrant junctions showed features associated with a-NHEJ, mainly microhomologies. (Fig.2.9). One junction had both a 3bp microhomology and a large insertion (103bp) comprised of a duplicated sequence centromeric to the break point and 5 non-templated nucleotides, a structure highly suggestive of a-NHEJ. The rest of junctions had microhonology between 1-5bp with a majority of 1-2bp. Interestingly, most of the inversion junctions exhibited microhomology suggesting that the complexity of the repair (joining four ends vs two ends) might effect end joining. A third feature associated with a-NHEJ, excessive deletion, can be measured only if an initial break site is known, which is rarely the case for random rearrangement. However, the immature T cell phenotype of these tumors

(Fig.2.8B), together with evidence of V(D)J recombination in our example suggested it might be worthwhile to examine junctions for evidence of having been RAG-generated. In such cases, we could assume cleavage at sequences fortuitously resembling RSSs (cryptic RSS, or cRSS), and take advantage of this to score deletion. Indeed, we were able to identify 10 junctions associated with credible cRSS bearing identifiable heptamer and nonamer sequences at both ends (Table 3). Two of the junctions exhibited a typical limited amount of deletion from the cRSSs, but one exhibited more extreme deletion from each breakpoint (15bp on one side, and 16bp on the other), consistent with a-NHEJ. This junction was created by a translocation. In summary, if we assume that c-NHEJ cannot tolerate microhomology to any degree, half of the aberrant rearrangements in the tumors exhibited characteristics reported for a-NHEJ. However, the remaining junctions cannot be unequivocally assigned to either a-NHEJ or c-NHEJ because our findings show that a-NHEJ junctions are not necessarily distinct from those formed by c-NHEJ.

Spectrum of SVs

RSS-RSS

	coding joints	-bpL	Lheptamer	spacer	Lnonamer	insert	Rnonamer	space	Rheptamer	-bpR	m-h	genes
	canonical coding joint		CACAGTG		ACAAAAACC		GGTTTTTGT		CACTGTG			
1	422-13-4 (DEL)	7	CACAGTG	12	ACAAAAACT	CAACCCCT	GGTTCTTAT	23	TGTTGTG	3		TCRgv2-J2
2	422-12-11(DEL)	8	CACACTA	23	ACAAAAACC	CCT	GGTTTTTGT	12	CACCGTG	1		IgHD4.1-IgHJ2.1
3	422-6-34 (DEL)	1	CACGGTG	23	ACAAAAACC	AAG	ATTTTTCTC	12	CACTGTG	2		Db1-Jb1.1
4	422-6-14 (DEL)	0	CACGGTG	23	ACAAAAACC	GA	GGTTTGTGT	12	CTCTGTG	7		Db1-Jb2.6
5	422-14-26(DEL)	0	CACAATA	23	ACAAAAACC	CCTT	GGTTTTTGT	12	CACTGTG	2		TCRDv3-Ja34
6	422-6-17 (DEL)	0	CACAGTG	23	ACAGAAAGG	GGGACAGGGGGAAG	ATTTTTCTC	12	CACTGTG	2		TCRbV13.2-Db1-J1.1
7	333-13-5 (DEL)	7	CACAGTG	12	ACAAAAATC	CCCTCC	GGTTCTTAT	23	TGTTGTG	4		TCRgv2-TCRgJ2.1
8	333-13-8 (DEL)	0	CACAGTG	12	ACAAAAATC	A	GGTTCTTGT	23	TGCTGTG	3		TCRg V4-J1-C1
9	333-6-11 (DEL)	3	CACAGCA	12	CAAGAATTC	CATAT CCCCTGT AG	GGTTTGTGC	23	AACTGTG	5		TCRb Vb4-D1Jb2.1
10	333-12-16 (DEL)	12	CACAGTA	12	GTCATCAGT	GC	AGTTTTCTG	23	CACTGTG	0		IgHv2.9-IgHV1.2
11	333-6-17 (DEL)	1	CACAGTC	12	GCACAAACC	A TGTC T	GTTTTATGC	23	CGCTGTG	2		Vb24.2Db1Jb2
	inversion joint	'-or+ bp(RSS)	Lnonamer	space	Lheptamer	insert	Rheptamer	spacer	Rnonamer	'-or+ bp(RSS)	m-h	genes
12	333-6-19 (INV)	2	GGTTTTTGT	23	CACTGTG		GATTTATGT	12	CACTGTG	2	A	IgKJ1-5 -IgKVk12-89
13	333-14-30 (INV)	4	GGTTTGGGT	23	CACagtg	GCCAT	TGTTTTTGT	12	CACTGTG	Plus 5		TCRDV5-TCRdD1
14	333-12-34 (INV)	13	GGTTTTTGT	12	CACCGTG	CCCC	AGTTTTCTG	23	CACTGTG	0		DQ52-IgHV1-2

RSS-cRSS

15	422-13-7 (DEL)	6	CACAGCC	23	GATGGAAAC	GCAT	GATTTTTGT	12	CACTGTG	0		5UTR TVRgV6-TCRgJ1
16	333-14-4 (DEL)	1	CACTGTG	23	TATATTAAA	CCAG	GGTTTTTGG	12	AGCTGTG	0		5UTR TCRdD1-TCRdJ1
17	333-14-18 (DUP)	1	CACAGTG	23	GCATAAACC	C	CTGGGGGAT	12	TTCTGTG	16		TCRdV2.2-5'UTR TCRdV5

cRSS-cRSS

18	422-16-21 (DEL)	4	CACAGCG	12	ACAAAAACT		GATTATCCC	23	CACTGTG	4	GG	Trmt2a
19	422-4-28 (DEL)	3	CACAGTA	23	ATGCTAAGT		AGTTTTTAC	12	CACTGTG	1		Tnfrsf1b, Tnfrsf8
20	422-12;9-24 (Tx)	15	CACCCTC	23	GGAGGGAGG		AGTTTTTGA	12	CACTGTG	16		IGR/Filip1
21	333-16-6 (DEL)	0	CACAGTG	23	ACTGAATCC	AGC	TTTAGAAAA	12	GTGTGTG	1		Plxna4
22	333-2-19/1 (DEL)	1	CACTGTG	12	TAAGTATCA	TTT	GCACTGTGT	23	CACTGTG	2		Notch1
22	333-2-19/2 (DEL)	1	CACTGTG	12	TAAGTATCA	TTTA	GCACTGTGT	23	CACTGTG	3		Notch1
23	333-1-22 (DEL)	3	CACATTG	12	TCATTTAGT	C	GGATTCTGA	23	GGTGGTG	0		Zranb3

Other

24	422-10-5 (DEL)	16	CACAGCA	12	AGGGGGCTG		TGAAGAGTC	23	GCAGGTG	2	AG	fig4
25	422-1-18 (DEL)	35	CACTCGG	23	AGCTTTTAT		CACATGAGG	12	TAAAGTG	47		Cntnap5b
26	422-3-53 (DEL)	35	CACCTAA	12	ACACAGAGA		TTAGCAAGT	23	ATCCGTG	48	TT	Gria2
27	422-11-43 (DEL)	27	CACTGGG	12	TTCATATTA	T	TCTTTTGTT	23	CATGGTG	48		Accn1
28	422-12-2 (DEL)	7	CACTTAA	12	CAAAATAAA	A	CCCTCACG	23	TGAGGTG	3		Psmc1
29	422-12;1-40(Tx)	44	CACTCTT	12	ATACTAAAA		AAAAAATAG	23	TTTTGTG	38		IGR
30	333-14-8 (DEL)	32	CACAAAA	12	GCAAAAGTC		ACAACTCCA	23	CACTGTG	7		Wdhd1-IGR
31	333-14-2 (DEL)	83	CACAGGG	23	GATCCTCTG		GGACTCTAA	12	AACTGTG	507		Entpd4
32	333-8-12 (DEL)	18	CACCCCT	23	TCCCTGCAT		AATTCATGT	12	GGGTGTG	78	A	Herpud1-IGR
33	333-12-13 (DEL)	32	CACACAA	23	GGTATCATA		TATATTCTA	12	TCAAGTG	63	AG	IGR
34	333-19-14 (DEL)	276	CACAGGT	12	CCAGCCATA		AGCCATTCT	23	TTTAGTG	30		Papss2-Rnls1
35	333-6-23 (DEL)	76	CACAGAG	12	GCCAAACAA		TTATATCAT	23	CATTGTG	50	T	Igk UTR not VJ region
36	333-6-27 (DEL)	25	CACATAA	23	AAAAACACA		ATGCCACCC	12	GTGCGTG	27	TG	IgK Not VJ regions
37	333-3-28 (DEL)	67	CACTATT	23	TTCCATCAA		TTTGTTTGT	12	TGTTGTG	334	CAAT	IGR
38	333-14-4 (DUP)	46	CACAAAG	12	TATGTACTC	TTAAGTGATTA	GTCACCGAC	23	GATGGTG	16		Peli2
	inversion joint	'-or+ bp(RSS)	Lnonamer	space	Lheptamer	insert	Rheptamer	spacer	Rnonamer	'-or+ bp(RSS)	m-h	genes
39	333-14-2 (INV)	0	GGGAAGAGA	12	CCCAGTG	ATAT	ACCAGCCAT	23	GTAAGTG	3		IGR
40	422-9-10 (INV)	5	ACAGGAGGT	24	CAGGGTG		AGAGATATC	12	TAATGTG	48	A	Cntn5
41	422-12-12 (INV)	55	GGATATGCA	23	GGCAGTG	AATCA&98bp(dup)	GCAACTGGT	11	AGGAGTG	61	CAG	Atg2b
42	422-12-5 (INV)	268	AGAAGAGCA	12	CCCAGTG		GGGCGGCGC	23	CGGGGTG	166	CCGCT	Psmc1/AKT1
43	333-14-4 (INV)	30	CCTCCTTAG	23	CACTGTG		AAAGGTGAG	12	CTTTGTG	141	GA	Wdhd1
44	333-14-5(INV)	91	GAATCCTCC	23	GAATGTG		TAGCATGGT	12	GTGTGTG	1	GT	Long non-condong RNA-IGR
45	333-14-9(INV)	46	GGAGGTTGT	12	ACCAGTG		AGAAGCAGG	23	TAGAGTG	37	TCTTG	Long non condong RNA
46	333-14-18(INV)	2	AGATATGCT	23	GTTCGTG		GATCCCGAG	12	TCCTGTG	73	GT	Nufip-IGR
47	333-14-28(INV)	44	TCTTTCTGT	23	CAATGTG		AATTCTGAG	12	TCGGGTG	8	AGGC	Nrg3

Table 3. Structural variants from RAG2f$^{FS/FS}$;p53$^{-/-}$ tumors. SVs from WGS of two tumors (13422, 12333) were grouped into 4 categories: 'RSS-RSS' in gray are antigen receptor rearrangements using bona fide RSSs. 'RSS- cRSS' in yellow are rearrangements that only one side had a bona fide RSS with the other being cRSS. 'cRSS-cRSS' in blue are rearrangements outside of the antigen receptor loci potentially mediated via cRSSs. 'Other' represents all the rest of the genomic lesions that could not be assigned to any of the other groups. The name of the SV is #tumor-#ch-variant # (type of lesion; DEL=deletion, INV= inversion, Dup=duplication, Tx=translocation). Red indicates nucleotides similarity to consensus heptamer and nonamer. In 'insert' column, bold letters are P nucleotides. IGR stands for intergenic region.

A	RSS-RSS	
1	DEL4	TGAAGCCACCTACTACTGTGCAGTCTGGATgagcacaacattagagcctc
		TGAAGCCACCTACTACTGTGCAGTCTGGAT**AGGGGTTG**GGGCTTTCACAAGGTATTTGCA
		tgagcactgtgatagctcGGGCTTTCACAAGGTATTTGCA
2	DEL11	GAGTGGTGCCTTGGCCCCAGTAGTCaaagtagtcacactatcataga
		GAGTGGTGCCTTGGCCCCAGTAGTC**CCT**TCCCAGTTAGCACTGTGGTGCTCC
		tgagccacgtgtcaccgtggTCCCAGTTAGCACTGTGGTGCTCC
3	DEL34	AAAGAAGACTTCTGTGTTtgcacagtgccataggatga
		AAAGAAGACTTCTGTGTT**CTT**CCCCCTGTCCCCACAATGTTACAGCTTT
		catagaattgaatcaccgtggCCCCCTGTCCCCACAATGTTACAGCTTT
4	DEL14	TGTAACATTGTGGGGACAGGGGGCcacggtgattcaattctatgggaagc
		TGTAACATTGTGGGGACAGGGGGC**GA**GAACAGTACTTCGGTCCCGGCACCAG
		gagcctctgtgctcctatGAACAGTACTTCGGTCCCGGCACCAG
5	DEL26	ATGTACCACTGCCTTCGGGGAGAcacaataccaaggacac
		ATGTACCACTGCCTTCGGGGAGA**CCTT**TTCCAATACCAACAAAGTCGTCTTTGGAAC
		gaacatgatatcactgtgtcTTCCAATACCAACAAAGTCGTCTTTGGAAC
6	DEL17	GTACTTCTGTGCCAGCGGTGATGcacagtgatgtggggtt
		GTACTTCTGTGCCAGCGGTGATG**GGGACAGGGGGAAG**AACACAGAAGTCTTCTTTGGTAAAGGA
		cctatggcactgtgcaAACACAGAAGTCTTCTTTGGTAAAGGA
7	DEL5	TTGTTCCTTCTGCaAATACCTTGTGAAAGCCCgagctatcacagtgctca
		TTGTTCCTTCTGCaAATACCTTGTGAAAGCCC**CCCTCC**TCCAGACTGCACAGTAGTAGGtGGC
		gaggctctaatgttgtgctcaTCCAGACTGCACAGTAGTAGGtGGC
8	DEL8	TGCAAATACCTTGTGAAAACCTGAGCTATcacagtgctcacagcttctac
		TGCAAATACCTTGTGAAAACCTGAGCTAT**A**TAGCCGTAGGAACAGTAGTAGATAGCTTCGTC
		tctggaagctcagatgctgtgcttTAGCCGTAGGAACAGTAGTAGATAGCTTCGTC
9	DEL11	TGGTCCGAAGAACTGCTCAGCATAGttacacagcagaaaagggctaccaa
		TGGTCCGAAGAACTGCTCAGCATAG**CATATCCCCTGTAG**GCTGCTGGCACAGAGATACACAGCAG
		ctctgcacaactgtgtcttaGCTGCTGGCACAGAGATACACAGCAG
10	DEL16	TTTGTGCAAGTTCACACTGGACTTCCCTCActgtgtgtttggcacagtag
		TTTGTGCAAGTTCACACTGGACTTCCCTCA**GC**TCATGCACAGTAATAGACCGCAGAGTCCTC
		gatgtggtcacaacactgtgTCATGCACAGTAATAGACCGCAGAGTCCTC
11	DEL17	TTCACCAAAGTAGAGCTGCCCGGTGTTTgcacagtcctggaaatgctggc
		TTCACCAAAGTAGAGCTGCCCGGTGTTT**ATGTCT**CAGACTGCTGGCACAGAGCTACAGTGCTGA
		gtacatttcagcgctgtgtaCAGACTGCTGGCACAGAGCTACAGTGCTGA
12	INV19	CAGCTTGGTGCCTCCACCGAACGTCCAccacagtggtagtactccactg
		CAGCTTGGTGCCTCCACCGAACGTCC**A**GGAGTACTTAACACATTTTGACAGTAATA
		atggcttgaatcactgtgggAGGAGTACTTAACACATTTTGACAGTAATA
13	INV30	GGTACAGGCTCCCTGGGCACCTGCACCACagtgataccccgaggcacagt
		GGTACAGGCTCCCTGGGCACCTGCACCAC GCCATCACAGTGAAACACAGCCGTACAAAA
		tacttcaacctgtgtgatatGCCATCACAGTGAAACACAGCCGTACAAAA
14	INV34	TGGAGAGCTCCaAACAGAAGGTTTTTGTTGAGCCacgtgtcaccgtggtcccagttagca
		TGGAGAGCTCCaAACAGAAGGTTTTTGTTGAGCC**CCCC**CACAGTGTTGTGACCACATCCTGAGTA
		cggtctattactgtgcatgaCACAGTGTTGTGACCACATCCTGAGTA

B

	RSS-cRSS	
15	DEL7	AAATACCTTGTGAAAACCTGAGCTATcacagtgctcacagcttctacaaa
		AAATACCTTGTGAAAACCTGAGCTAT**CGTA**TGTACTGTTCTCTTGAGAATCGAG
		tcaggctgtggcacttTGTACTGTTCTCTTGAGAATCGAG
16	DEL4	AACAAAGGGTGTTTTTGTACGGCTGTGTttcactgtgatggcatatcaca
		AACAAAGGGTGTTTTTGTACGGCTGTGT**CCAG**CTACCGACAAACTCGTCTTTGGACAAGGAACCC
		tggcctcagtagctgtgCTACCGACAAACTCGTCTTTGGACAAGGAACCC
17	DUP18	GAGTAGCTTCCATCCTTCTCTGTTATGAGCacacagtggtgcacaagcac
		GAGTAGCTTCCATCCTTCTCTGTTATGAGC**C**CGGTGACACAGGCAACAGAGAGGGTCTAAC
		tgtgctctcatggagcgccaCGGTGACACAGGCAACAGAGAGGGTCTAAC
	cRSS-cRSS	
18	DEL21	GTGTGTGTGTGTGTGTATCCCAAAGGAGATGGtatcacagcgatacgtag
		GTGTGTGTGTGTGTGTATCCCAAAGGAGAT***GG***GCTGAGGATCTGGTGCCAGGTTT
		agttccactgtggcaGGGCTGAGGATCTGGTGCCAGGTTT
19	DEL28	ATGTGCAGGCTCATGTGGGCATGCCTAAGAGtggcacagtaattgtatgctggg
		ATGTGCAGGCTCATGTGGGCATGCCTAAGAGCGGCCTTCATGATGTAGATTTTCTCTG
		tacagagcccacgatcactgtgtCGGCCTTCATGATGTAGATTTTCTCTG
20	TX24 t12;9	TTGGTTTTAGAGAGCTAGACACTGAGTTTGGgaccttataaattctcacc
		TTGGTTTTAGAGAGCTAGACACTGAGTTTGGAGATGGGCTGACGAGTGAGCTTTGGGGTTC
		tgtggctcagtgatgcagctAGATGGGCTGACGAGTGAGCTTTGGGGTTC
21	DEL6	CCCaAGAAACAGAAGACTGCGGCTGGTAATGTAGcacagtggtagagtat
		CCCaAGAAACAGAAGACTGCGGCTGGTAATGTAG**AGC**GTGTGTGTGTGTGTGTGTGTGCTTTTTA
		gctttgcttagggtgtgtgtGTGTGTGTGTGTGTGTGTGTGCTTTTTA
22	DEL19-1	CTCCATTGCCTGGGGTGTGACCTGTCccactgtggctggctgca
		CTCCATTGCCTGGGGTGTGACCTGTC**TTT**GAAGGGTTGAGGCCCTGCCACA
		tgcagggagcactgtgtgGAAGGGTTGAGGCCCTGCCACA
22	DEL19-2	CTCCATTGCCTGGGGTGTGACCTGTCccactgtggctggctgca
		CTCCATTGCCTGGGGTGTGACCTGTC**TTTA**AAGGGTTGAGGCCCTGCCACACCTCAACTGA
		tgagcctgcagggagcactgtgtggAAGGGTTGAGGCCCTGCCACACCTCAACTGA
23	DEL22	CTGATTTTGAAGCCAAGAACATACATCAAAGTtaccacattgtgtgaagt
		CTGATTTTGAAGCCAAGAACATACATCAAAGT**C**CATGACTGCAAACACAGAGCAG
		agagattaagtggtggtgCATGACTGCAAACACAGAGCAG

C	Other	
24	DEL5	TTAATCCAAGGTGGACCCCTCTCAAGAAGgtcacctgctaggttttagtt
		TTAATCCAAGGTGGACCCCTCTCAAGA***AG***TGAGAGCAGGAGTTGCTCTAGGGGGTTTAGCAG
		tttgttctagccAGTGAGAGCAGGAGTTGCTCTAGGGGGTTTAGCAG
25	DEL18	ACAGATTTTTTTCCTAGTTTTAAC –140bp– tcttgcccgagctagt
		ACAGATTTTTTTCCTAGTTTTAAC–140bp–CTGCAGTATTTGGTAATCTGTTATC
		ttttaaactgaaaag –140bp– CTGCAGTATTTGGTAATCTGTTATC
26	DEL53	TAGAGATAATGGCATGGCTCTCTGGTTtgccgtaggccaggtgttctcaa
		TAGAGATAATGGCATGGCTCTCTGG***TT***CACTTATCTTAATTTAAAATTCACGGACT
		cagcaggagcctgaaattaTTCACTTATCTTAATTTAAAATTCACGGACT
27	DEL43	GTCTTAAGAATAACAGTGCCCAGGTCCCATTCTAAtatgaatgactccta
		GTCTTAAGAATAACAGTGCCCAGGTCCCATTCTTAAAATTGATTTCTCAGGTCTGCCTCCTTGTC
		aaacaaaagaaaggatgaaAAAATTGATTTCTCAGGTCTGCCTCCTTGTC
28	DEL2	AGCTTCCTTCAGTTTTATTTTGCTCCAGAGTAGCTtaagtgttgctcaag
		AGCTTCCTTCAGTTTTATTTTGCTCCAGAGTAGCTAGATGATTAGAGCGAGCGCTTTGCTCTGCT
		gagcagttggtgaggtggagGATGATTAGAGCGAGCGCTTTGCTCTGCT
29	TX40 t1;12	ATAAGTGATTAGATTAAAATGGTATAAATGACAGATTCttttaatatatattaatt
		ATAAGTGATTAGATTAAAATGGTATAAATGACAGATTCGAGCAGCTGAATAACCCAAACACTTG
		agcatgaggtctgacctgaaGAGCAGCTGAATAACCCAAACACTTG
30	Del8	TTATCTCTCTGCTTAGACATGATTACAAGTGAGAAaaggtactactacctataaca
		TTATCTCTCTGCTTAGACATGATTACAAGTGAGAATTTAACAACCCCACTGAGGTATCCCTTCTA
		caaccccactgtggtatcccTTTAACAACCCCACTGAGGTATCCCTTCTA
31	DEL2	tCCTCCTTCCTCCTCCTCCTTCTCCCTcttcttcttaatttagattattt
		tCCTCCTTCCTCCTCCTCCTTCTCCCTTTCCCTCTGGTACATAATCAGGGTATCACA
		ttgctgtttacttcttataaTTCCCTCTGGTACATAATCAGGGTATCACA
32	DEL12	ATCTACATATCTGGGGACCACCACATCAGACCAtgaagataccagctctc
		ATCTACATATCTGGGGACCACCACATCAGACC***A***GCGAGTTCCAGGACAGCCAGGGCTACATAGA
		agccagcctgatctacagAGCGAGTTCCAGGACAGCCAGGGCTACATAGA
33	DEL13	ATTACaAGTCTTAtGGGGGAAAGCATCATACAGgaagttaatgtaaaagaaaag
		ATTACaAGTCTTAtGGGGGAAAGCATCATAC***AG***ACCTACaACCCAGAATTATACTCT
		tttatctactaaatataaatttagAGACCTACaACCCAGAATTATACTCT
34	DEL14	TTCTTTTTTCCTTGCATACAACAAAACAGctatcaaagggtctttagtgg
		TTCTTTTTTCCTTGCATACAACAAAACAGTCATATGGACAAAAACCTCTGCTCCTGCCTT
		cttgctgtccagctggtctTCATATGGACAAAAACCTCTGCTCCTGCCTT
35	DEL23	TATAAAGCCACCAAACCAAGATACTATTTTggatacctggaagtgattgc
		TATAAAGCCACCAAACCAAGATACTATTTT***T***TCATTGTCTCACAATCACAAACACTT
		tatcactctgataagtttggatattctccTTCATTGTCTCACAATCACAAACACTT
36	DEL27	AAATAGATAGATAAATAACAAATTCATGTATGgattttcaacagaaagat
		AAATAGATAGATAAATAACAAATTCATGTA***TG***CCACACCACGCAGTGCCGAGAATGTCAC
		cagccaaaaacaccacttgcTGCCACACCACGCAGTGCCGAGAATGTCAC
37	DEL28	CAGAAACAGGGACTCATAGACCTGGGATGCATGACAATaacttgtcaatg
		CAGAAACAGGGACTCATAGACCTGGGATGCATGA***CAAT***GTATTTCTTACAATAAATAAGTA
		aatgttttgaaagagaatggcCAATGTATTTCTTACAATAAATAAGTA
38	DUP4	TGAATTTATGAAATTCCTAGGCAaatggatggacctggagggtatcatcc
		TGAATTTATGAAATTCCTAGGCA**TTAAGTGATTA**TGGTACCCATGATGCTCTGCCCAGCACCAC
		ggtgtttacttaagaccctgTGGTACCCATGATGCTCTGCCCAGCACCAC
39	INV2	ACAAATGTGGGAAATATGTCTTAGGAATTATcactgggctcaacattaca
		ACAAATGTGGGAAATATGTCTTAGGAATTAT**ATAT**CTAATGCAAAAACCCACAACTGTAAAACA
		tccgtttgcagtaagtggaaACTAATGCAAAAACCCACAACTGTAAAACA
40	INV10	GATGTCATCACCTTCCCAgggtgtactttgggtttgttca
		GATGTCATCACCTTCCC**A**AAGATGACAAAGATCATGGTGGCCTTTGG
		gtgctgagctctttgagAAAGATGACAAAGATCATGGTGGCCTTTGG
41	INV12	CGTGGACACAGGCGTGCCTGTGCAGgagagagcaagggatggtc
		CGTGGACACAGGCGTGCCTGTG***CAG***(dup)+**TGATT**TTCAGATCTTCAGGGTTGGCAGGAAACC
		tgcaggttctgataagggaaTTCAGATCTTCAGGGTTGGCAGGAAACC
42	INV5	CCACAGGACGCTGCTCGGTCCGCTCCGCTggtcccggactcgcgtgcgct
		CCACAGGACGCTGCTCGGTCCGCT***CCGCT***CATCCCTACCTCACAACGATGGCCT
		gttacattatttccttcttccCCGCTCATCCCTACCTCACAACGATGGCCT
43	INV4	CCGAACTAAGTAGGGCACAGGCCCTTAGAcagatgggaatggcttatagaagtcttcac
		CCGAACTAAGTAGGGCACAGGCCCTTA***GA***GACCCTGTTCAGAATTAATTAAAAATAAGTTTTT
		gacagccagagctacctagtGAGACCCTGTTCAGAATTAATTAAAAATAAGTTTTT
44	INV5	TCTGTCTCAAATGTACACGTCTCTTGTTGGGAGGTatataagccttttta
		TCTGTCTCAAATGTACACGTCTCTTGTTGGGAG***GT***GTGTGTGTGTGTGTGTGTGAGAACTG
		gtgtgtgtgtgtgtgtgtgtGTGTGTGTGTGTGTGTGTGTGTGAGAACTG
45	INV9	GCCTGGGTTGAGTTTGCATAGTTCTTGcagatgccatcataactactatg
		GCCTGGGTTGAGTTTGCATAGT***TCTTGA***AAAACAAAACAACATAACAAAAAG
		ggctacacagagaaaccctgTCTTGAAAAACAAAACAACATAACAAAAAG
46	INV18	GCTGCTACTTTATAACTGTAATTCTGCTGCCGTcacgaactgtaacgtaaaca
		GCTGCTACTTTATAACTGTAATTCTGCTGCC***GT***ATACTAACAATTACATATTATTTCACTT
		gtcattaatacttttaataaGTATACTAACAATTACATATTATTTCACTT
47	INV28	GTGTCTGTCACTATAAAGAGGACCAGAGAGGCcatgtcagttgaagaaat
		GTGTCTGTCACTATAAAGAGGACCAGAG***AGGC***TAACATTATTCTTTATCTAGTGAATT
		ttaattcgggtgacaatcacAGGCTAACATTATTCTTTATCTAGTGAATT

Figure 2.9 Junctional sequence of the SVs. All identified Svs are presented here. In blue and green are the germ line sequences of each breakpoint (upper and lower row, respectively). Middle line represents the junction itself. Bold black are insertion and bold black italic are microhomologies. The numbering correlates to those in table 3. A. endogenous V(D)J rearrangements. B. bona fide RSS to cRSS and cRSS to cRSS. C. all other SVs that could not be assigned to any of the other groups.

DISCUSSION

We used two knockin mouse models, $RAG2^{FS/FS}$ and $RAG2^{del352/del352}$, to demonstrate that in both c-NHEJ-proficient and c-NHEJ-deficient animals, removing RAG2's C-terminus allows aberrant use of a-NHEJ to join physiologic, chromosomal DSBs. These data provide the first evidence that pathway choice operates during normal V(D)J recombination *in vivo.* Is pathway choice physiologically relevant in c-NHEJ-proficient animals? Our data suggest that defects in pathway choice may explain, at least in part, the genomic instability seen in lymphomas from $RAG2^{FS/FS}$;$p53^{-/-}$ mice. However, until a-NHEJ is better defined we cannot rule out involvement of c-NHEJ in such rearrangements.

Additional evidence that RAG2's C-terminus is important for controlling repair pathway choice is provided by analysis of N nucleotide addition. Previous work showing that junctions formed in the absence of Ku80 were devoid of N additions indicated that Ku80 is important for recruiting terminal deoxynucleotidyl transferase (TdT) to DSBs (107,122,123). This does not appear to be the case in the presence of RAG2 C-terminal truncations where coding joints from double mutant mice displayed N regions. We speculate that this may be another feature of abrogated pathway choice control, in which TdT has unregulated access to RAG-mediated DSBs in the absence of RAG2's C-terminus. This is supported by

our observation that in RAG2$^{FS/FS}$ mice, N regions are significantly more prevalent at signal joints (35% in RAG2$^{FS/FS}$ mice vs 20% in wild-type $p<0.006$).

Because RAG2 C-terminal mutants allow joining by a-NHEJ, characteristics of chromosomal a-NHEJ can be studied in both c-NHEJ-proficient and c-NHEJ-deficient animals. Surprisingly, coding joints repaired by a-NHEJ do not exhibit features commonly associated with a-NHEJ, even in the absence of Ku80. The resemblance of a-NHEJ repair to that of c-NHEJ implies that the former might be more prevalent than appreciated and can, like c-NHEJ, repair physiological DSBs in a non-mutagenic fashion. This discovery suggests that we must exercise caution when interpreting features of DNA rearrangements in sequenced tumor genomes. Moreover, the unified sequence features of these junctions (e.g. the lack of a subset of junctions bearing excessive deletions, insertions or microhomologies) imply that they may be formed by a single repair pathway, rather than through a hodgepodge of joining activities. Our data have another important implication, which is that alternatives to c-NHEJ cannot be presumed to have an impact on junction structure. Only upon close scrutiny could we detect any difference between the Ku80-independent coding joints and those generated in wild-type mice (ADRs, discussed in results). Thus, either many repair factors are shared between c-NHEJ and alternative joining mechanisms, or the alternative pathway used to join coding ends is organized to call up a similar compilation of different enzymes that pretty nearly reconstruct the products of c-

NHEJ. We are not the first to question the existence of a single, well-defined alternative NHEJ pathway (35,8), however, this study contributes a definitive illustration of the lack of consistency between different systems.

The striking finding that most coding joints have normal structures is in apparent conflict with both the signal joints recovered in this study as well as with results from our cell-based assay in which the RAG2FS mutant allows coding joints bearing excessive deletions and microhomologies to be formed on an extrachromosomal substrate (36). These apparent discrepancies may be the consequence of a context-dependent aspect of a-NHEJ. There are obvious differences in the requirements in each case: signal ends prior to joining have blunt-ended termini, coding ends need to undergo several processing steps before they can become ligated. There is evidence that the structure of the DNA ends at a DSB can affect their resolutions (124). An extrachromosomal substrate can reflect much about physiological joining, however, there can be a difference between end-joining where cut ends are closely linked and may be able to collide at random, versus ends in the chromosome, where a break must be somehow bridged and stabilized for joining. Lastly, the sequence environment surrounding the DSBs may also influence joining. In immunoglobulin class switch recombination, the almost invariant appearance of microhomology at junctions produced by a-NHEJ may be attributed to the repetitive nature of the switch regions (125,126). This might also be the case in our extrachromosomal

substrate, where a substantial (9 nt) microhomology is present near the break points (36). In antigen receptor loci, which lack this repetitive sequence environment, such repair might not be similarly favored.

Whether the V(D)J joining system refines the outcome of joining by controlling where and when cleavage occurs (ie ‘context’) and by handing off the coding ends to defined pathways, and how this is achieved at a mechanistic level will be relevant to unraveling the mechanisms responsible for preserving genomic integrity during V(D)J recombination, and may illuminate end-joining metabolism in general.

CHAPTER 3

The RAG2 C terminus protects against illegitimate V(D)J recombination

The V(D)J recombination process is a double edged sword. On the one hand it is crucial for the development of the adaptive immune system and its diversity. On the other hand, it inadvertently causes genomic instability contributing to malignant transformations. This illegitimate V(D)J recombination process can be caused through a variety of mechanisms (Fig. 1.3). This chapter will deal with events that are mediated via cryptic RSSs (cRSSs, groups 2-4in Fig. 1.3) because they are recognized as major contributors to lymphoma neoplasms. Aberrant V(D)J recombination by the RAG complex can potentially occur every 600-1000bp in the mouse or human genome respectively (96). Obviously not all cRSSs are the same and some are more effective than others, as reported in systematic functional analyses (127,128). These studies showed that there is more to an RSS than a heptamer or nonamer. The recombination efficiency of an RSS is a summation of all its regions, conserved or non-conserved. It is almost impossible to identify a functional cRSS based on sequence alone. A comprehensive cRSS predication algorithm failed to favorably score multiple cRSS that can undergo recombination in-vitro (96,129).

SKY analysis of tumors from RAG2$^{FS/FS}$; p53 $^{-/-}$ mice showed a high percentage of translocations occurring mostly outside of antigen receptor chromosomes (66% of breaks, Fig.2.8C). As such translocations are not frequent in lymphomas from p53$^{-/-}$ mice, we hypothesized that the C terminus of RAG2 might inhibit cRSS usage to protect the genome integrity. If true, an increased usage of cRSS by the FS allele may provide an additional mechanism by which the lymphomas observed in RAG2$^{FS/FS}$;p53$^{-/-}$ mice evolve. Indeed, our RAG2$^{FS/FS}$;p53$^{-/-}$ tumor data revealed that 27% of the genomic lesions (excluding normal V(D)J rearrangements) were **potentially** mediated via aberrant V(D)J recombination with some involving genes implicated in oncogenesis and genome stability (*Trmt2a* (130), *TNFR2* (131), *CD30* (132), Notch1 (93), Zranb3 (133) and Plxna4 (134)).

Several other results were supportive of this theory. RAG2$^{del352/del352}$ mice were shown to favor some RSSs more than others implying that the C terminus plays a role in RSS selection (112). Moreover, the FS allele lacks the plant homeodomian (PHD) of RAG2 (414-487 amino acid) that binds to trimethylated Histone 3 at lysine 4 (H3K4me3) (89,90). This region destabilizes the RAG-DNA complex impeding cleavage unless bound by H3K4me3 (135). RAG2$^{FS/FS}$ mice are deficient in this level of regulation and hence can cleave in sites irrespective of this epigenetic signal (135).

Functional recombination assays regarding the dynamic between the RAG complex and various cRSSs was only tested in the context of full length RAG2 (FLR2). A consistent observation was that both heptamer and nonamer, with a certain degree of deviation from a consensus sequence, are obligatory for V(D)J activity (127,128). Different sequence changes in either region were tolerated to variable degrees (127). Additionally, biochemical studies using purified core RAG1/2 showed that the heptamer, specifically the 'CAC', is critical for complete cleavage and hairpin generation while the nonamer is important for RSS binding and cannot promote cleavage without a heptamer present (129,136). Nonetheless, understanding of the exact mechanism/s of how illegitimate V(D)J recombination occurs is still lacking.

Illegitimate V(D)J activity has not been functionally tested with different RAG mutants. In light of our tumor data we decided to address this issue. First, we wanted to test whether the FS allele indeed increases aberrant V(D)J activity in a more robust fashion. If these were to hold true, we then sought to examine if various modification to the RGA2 C terminus might be more sensitive to changes in the RSS sequence.

Methods

Notch1 and Bcl11b intragenic deletions

Genomic DNA from thymocytes was prepared using the Roche High Pure PCR template preparation kit according to manufacture recommendations. Nested PCR was preformed to detect intergenic deletion for both genes. PCR protocols and primers were described previously (93,94). Modified primers for Bcl11b are described below. For a negative control we used genomic DNA from R2 KO liver. PCR samples were analyzed by electrophoresis on agarose gels. Identification of a band at the appropriate size was considered an indication of rearrangement. 4 samples from $RAG2^{FS/FS}$ mice and positive RAG2 wild-type were purified, TOPO TA cloned, and Sanger sequenced.

Primers:

Bcl11b F1 CTCTCCAATCCTGTGGTCTCTTAC

Bcl11b F2 GGGAACGCTTTTCGGCCTTACTTG

Bcl11b R1 GTCAGCCTAAGGCTACAGCACATTATG

Bcl11b R2 CTCTTCTGCACAGCTTTCCCTCTG

Extrachormosomal assay with human cRSS (LMO2/TAL2)

We used an in vitro system in eukaryotic cells that measures recombination efficiency. The substrates were a gift from Dr. B Nadel and are described in detail in ref 137. Briefly, the substrates have the two sequences to be tested that are

separated by a termination signal (oop). Hence a chloramphenicol acetyltransferase (CAT) gene will only be transcribed in E. Coli upon successful recombination that deleted the opp sequence in cells. The substrate also carries ampicillin (AmpR) resistance that was used to calculate recombination efficiencies. RMP41 fibroblast were transfected with full length Rag1 and the different Rag2 mutants together with a recombination substrate as a read out. 48 Hours post transfection we harvested the cells and obtained DNA by phenol/chloroform extraction. DpnI-digested plasmids were then introduced into chemically competent *E. coli*, and plated on ampicillin (100 µg/ml) and chloramphenicol (10 µg/ml). Absolute recombination efficiency was calculated by dividing the number of double resistant colonies by the total number of colonies (AmpR alone). As a control we transfected the substrate alone without Rag1/2 to account for any RAG independent recombination.

Extrachromosomal assay with modified consensus 12/23 RSSs

We used an established in vitro system in eukaryotic cells that measures recombination efficiency. Substrates were a gift from Dr. J Hesse and are described in detail in ref 127. As described above, the substrates carry an ampicillin (AmpR) resistance and upon a recombination event, resistance to chloramphenicol (CAT-R) is achieved as well. RMP41 fibroblast cells were transfected with full length Rag1 and the different Rag2 mutants with a recombination substrate as a read out. 48h post transfection we harvested the

cells and extracted DNA by phenol/chloroform extraction. DpnI-digested plasmids were then introduced into chemically competent *E. coli* and plated on ampicillin (100 μg/ml) and chloramphenicol (20-30 μg/ml). As a control we transfected the substrate alone without Rag1/2 to account for any RAG independent recombination. Absolute recombination efficiency was calculated by dividing the number of double resistant colonies by the total number of colonies (AmpR alone). The substrates contain the tested RSSs in inversion orientation meaning that double resistance can be acquired either by successful event (INV) or an unfaithful one creating a hybrid joint (HJ) product. HJ is generated upon a loss of one RSS. The cRSS that were tested possessed either a single nucleotide substitution in one RSS or a deletion of the whole heptamer or nonamer while the other is in intact consensus.

Statistical Analysis

A two-tailed unpaired t-test was applied for extrachromosomal recombination efficiency and for all other statistical analysis we used the Fisher exact test of independence.

Results

RAG2$^{FS/FS}$ mice show increased cryptic RSS usage

We wanted to investigate the cRSS usage by the FS allele in a more robust fashion since we only analyzed two tumors. Additionally we wanted to ensure that this phenotype is not dependent on p53 deficiency.

To achieve that we used two approaches:

1. Intrageneic deletions in two genes; *Notch1* and *Bcl11b* that are known to be mediated via the RAG complex and participate in lymphomagenesis (93,94). We examined genomic DNA from healthy thymi of wild-type RAG2 and RAG2$^{FS/FS}$ to avoid any effects of p53 deficiency.

2. Aberrant rearrangement as seen in human lymphomas. Specifically the LMO2/Dd1 (138) and the Db1/TAL2 (139) that were tested in a cell-based assay with the RAG complex (137).

DNA isolated from RAG2$^{FS/FS}$ mouse thymi showed a significant increase in recombination efficiency in both the *Notch1* ($p<0.03$) and *Bcl11b* ($p<0.003$) genes than wild-type RAG2 (Fig.3.1A). Sequence analysis of the PCR products verified that they correspond to the correct genes (Fig3.1B+C). These aberrant events occur in a tissue of a non-tumorigenic mouse. Hence, in the case of a second hit, for example p53, these events can contribute and even initiate lymhomagenesis.

A

	WT	RAG2$^{FS/FS}$
Notch1 (12Kb Del)	1/9	6/10 *
Bcl11b (65Kb Del)	1/8	7/8 **

B

Notch1

GGGAGGAGGATGCCC **CACCTCA-12-ACATGAAGG** **CTTTATAGC-23-CACTGTG** TAGGCTTGGGTGATT

WT

GGGAGGAGGATGCC(-1) GAG GGCTTGGGTGATT(-2)

RAG2$^{FS/FS}$

GGGAGGAGGATGCC(-1) TAGGCTTGGGTGATT
GGGAGGAGGATGCC(-1) TGACCCT TAGGCTTGGGTGATT
GGGAGGAGGATGC(-2) TC GCTTGGGTGATT(-3)
GGGAGGAGGA(-5) GA TAGGCTTGGGTGATT

C

Bcl11b

GGGACACACAGACACACAGACACACA **CACACAC-23-CCCATATCC** **TGTTGTAAG-12-CACTGTG** TGAGATTAATTTCTTTCCTTGTT

WT

GGGACACACAGACACACAGACACACA GGA TAATTTCTTTCCTTGTT(-6)
GGGACACACAGACACACAGACACAC(-1) CT GATTAATTTCTTTCCTTGTT(-3)
GGGACACACAGACACACAGACAC(-3) CCTCC GATTAATTTCTTTCCTTGTT(-3)
GGGACACACAGACACACAGAC(-5) GCACCT GATTAATTTCTTTCCTTGTT(-3)

RAG2$^{FS/FS}$

GGGACACACAGACACACAGACACACA GAGATTAATTTCTTTCCTTGTT(-1)
GGGACACACAGACACACAGACACAC(-1) TAGAGGG TAATTTCTTTCCTTGTT(-6)
GGGACACACAGACACACAGACACAC(-1) CGGGG TTAATTTCTTTCCTTGTT(-5)
GGGACACACAGACACACAGACACAC(-1) GGG GATTAATTTCTTTCCTTGTT(-3)
GGGACACACAGACACACAGACACAC(-1) GGA TAATTTCTTTCCTTGTT(-6)
GGGACACACAGACACACAGACAC(-3) CTCT ATTAATTTCTTTCCTTGTT(-4)
GGGACACACAGACACACAGACAC(-3) C GAGATTAATTTCTTTCCTTGTT(-1)
GGGACACACAGACACACAGACAC(-3) ATTAATTTCTTTCCTTGTT(-4)
GGGACACACAGACACACAGACAC(-3) TGAGATTAATTTCTTTCCTTGTT
GGGACACACAGACACACAGAC(-5) TCGA TAATTTCTTTCCTTGTT(-6)
GGGACACACAGACACACAGAC(-5) TCCCAGATGG ATTTCTTTCCTTGTT(-8)
CTTAGAAGGGAC(-21) ATTAATTTCTTTCCTTGTT(-4)

Figure 3.1 Intragenic deletions in Notch1 and Bcl11b. Genomic DNA from thymocytes of indicated genotypes was used in nested PCR to detect a 12Kb and 65kb deletions in Notch1 and Bcl11b genes respectively. A. Comparison between wild-type and RAG$^{FS/FS}$ mice. * p <0.03, **p<0.003. B-C. Sequence analysis of the PCR products for Notch1 (B) and Bcl11b (C). First line represents the germline sequence with bold letters indicating the cRSS from each side. Capital letters in the middle of the junction are N nt, Bold capital are P nt. wild-type n=1, RAG2$^{FS/FS}$ n=2-4

The same trend was observed with both the LMO2/Dd1 and Db1/TAL2 extrachromosomal substrates in fibroblast cells (Fig.3.2 A+B). The RAG2FS mutant was significantly more competent in recombining these human cRSSs than FLR2 (3.5 to 5 fold with TAL2, pv<0.03 and LMO2 pv<0.05 respectively). These results both support our hypothesis that the C terminus of RAG2 inhibits usage of cRSSs as well as implying that illegitimate V(D)J recombination by the FS allele may provide an additional mechanism by which the lymphomas in $RAG2^{FS/FS};p53^{-/-}$ mice evolved.

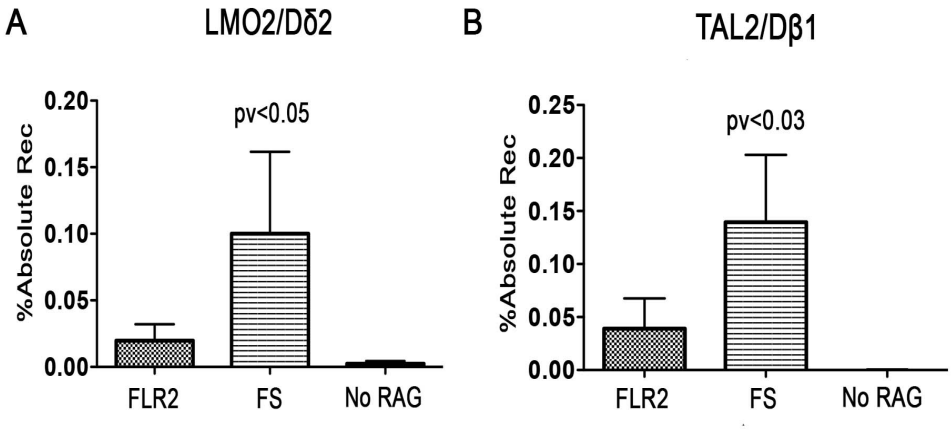


Figure 3.2 Recombination efficincies with human cRSSs. A. LOM2/Dd2 and B. TAL2/Db1 substrates were transfected into RMP41 cell line together with FLR1 and indicated RAG2 version. No RAG is transfection of the substrate alone. Recombination efficiencies were measured 48h after transfection as described in methods. Pv is vs FLR2.

The Dynamic between RAG2 and the RSS

To better understand the role of the RAG2 C terminus in illegitimate V(D)J recombination we decided to examine its 'behavior' in vitro with different cRSSs. We used well-defined recombination substrates that differ from each other only in their RSSs (127, Fig.3.3 &3.4). Our approach was to compare recombination efficiency of full length RAG2 (FLR2) to RAG2 mutants that accelerate lymphomagenesis in a mouse model: RAG2 Core (84), RAG2FS (chapter 2) and RAG2Neut (Coussens M, unpublished). The latter mutant retains the RAG2 C terminus but has neutralized acidic residues in a.a 370-383 (74). Unfortunately the RAG2 Core study was not completed because the version that we used (a.a 1-383) did not correlate with the mutation in the mouse as we discovered midway into the project (chapter 2).

We selected five different extrachromosomal substrates with modified consensus RSS (Fig 3.4). All of those were previously used showing very low levels of recombination with wild-type RAGs (127). The RSSs in these substrates are oriented in an inversional configuration (FIG 3.3). Hence a positive readout can be the result of a faithful inversional event or an unfaithful one such as a HJ (detailed in Methods, Fig 3.5).

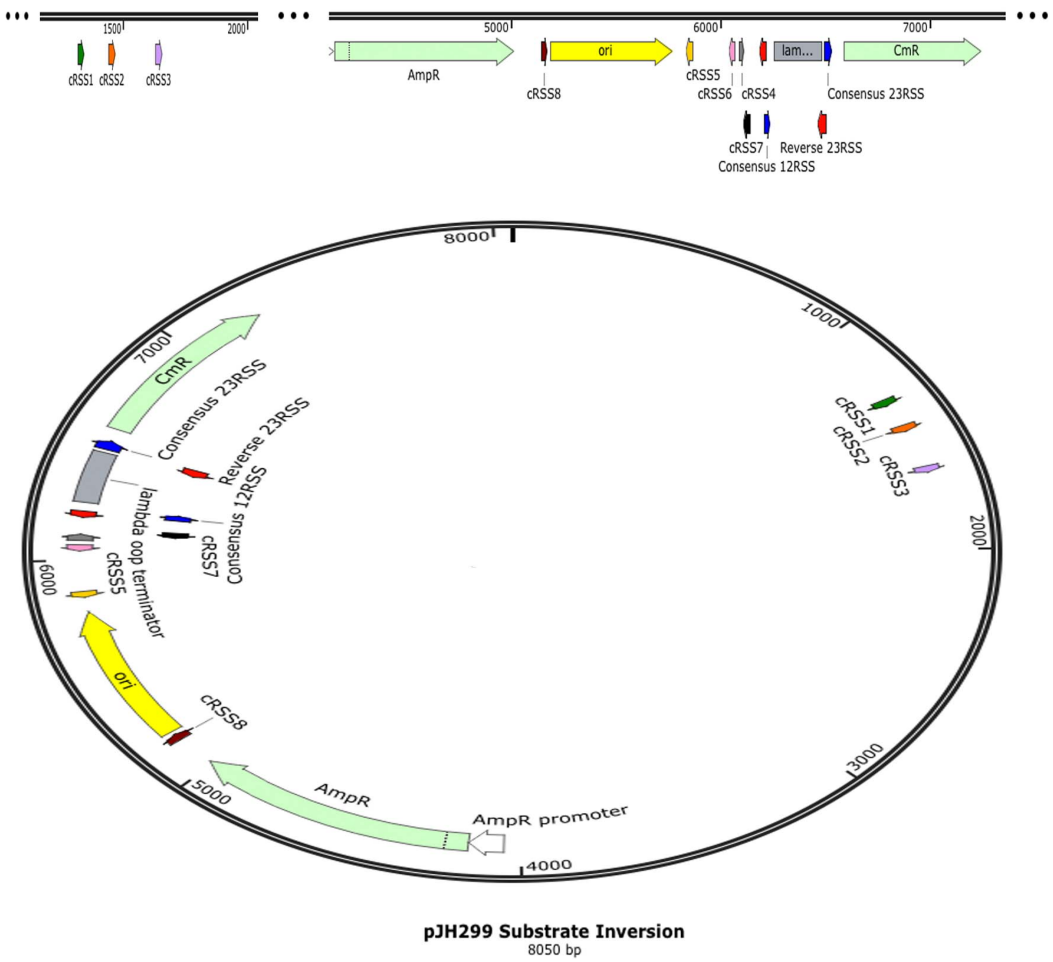


Figure 3.3 pJH299 Plasmid. The vector is preented both as circular and linear for convenience. Important regions are indicated. Different cRSSs are color-coded and their sequence is given in the Fig.6. The direction of the arrow indicates the direction of the RSS. Blue RSSs are the consensus 12/23 RSS. Ori – origin of replication, CAT-R chlorophenicol resistance, AmpR ampiciline resistance.

RSS/cRSS list in pJH299

Consensus 12RSS

CACAGTG CTACAGACTGGA ACAAAAACC

Tested 12 cRSS

12 Heptamer only CACAGTG CTACAGACTGGA **TGTCTCTGA**

12 mHeptamer CA**T**AGTG CTACAGACTGGA ACAAAAACC

12 Nonamer only **TGGCGAT** CTACAGACTGGA ACAAAAACC

Consensus 23RSSs

CACAGTG GTAGTACTCCACTGTCTGGCTGT ACAAAAACC

Tested 23 cRSSs

23 Heptamer only

CACAGTG GTAGTACTCCACTGTCTGGCTGT **TGTCTATGA**

23 mHeptamer

CA**T**AGTG GTAGTACTCCACTGTCTGGCTGT ACAAAAACC

Fortuitous 12 cRSS

cRSS 1	CACAG**C**G	TGTATAATCCAA	**G**TAA**G**TA**T**C
cRSS 2	CAC**G**A**T**G	ACTACTGGTCAT	**T**CA**G**CTA**T**G
cRSS 3	CACA**T**AC	TGCTGGAAGAAG	ACG**A**AA**T**CC
cRSS 4	CAC**C**C**C**A	GGCTTTACACTT	**T**AT**G**CTTCC
cRSS 5	CACAG**A**A	TCAGGGGATAAC	**G**CA**G**G**A**AA**G**
cRSS 6	CACA**T**T**A**	ATTGCGTTGCGC	**T**CA**C**TG**C**CC
cRSS 7	CACAA**T**T	CCAACCAACATA	**C**GA**G**CC**G**GA
cRSS 8	CAC**T**G**A**G	CGTCAGACCCCG	**T**AG**A**AAA**G**A
Rev12RSS	CAC**T**GTG	GTCGACCTGCAG	**C**C**C**AA**G**CTT

Figure 3.4 List of RSS in pJH299 plasmid. Consensus 12/23 RSS are indicated in blue. Tested cRSS are annotated and changes from the consensus 12/23 RSS are indicated in red. 'Fortuitous cRSS' are cRSSs that were used instead of the tested cRSSs. Rev12 RSS – Tested 12 cRSS in the opposite orientation.

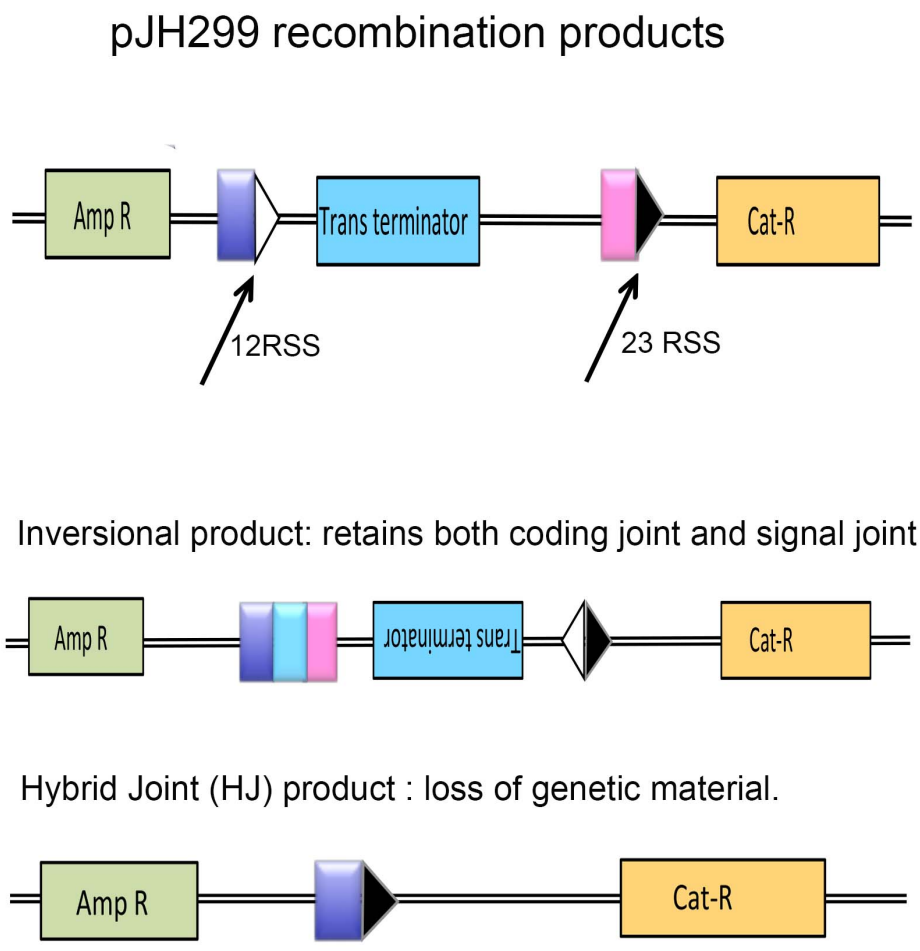


Figure 3.5 pJH299 recombination products. Without RAG activity the transcription terminator sequence does not allow expression of the Chloramphenicol (CAT-R) antibiotic resistance gene. Upon rearrangement this region is either flipped (inversion) or lost (HJ) allowing transcription of the antibiotic resistance gene.

Our theory was that a mutant RAG2 C terminus would lead to increased recombination efficiencies and higher HJ formation with cRSS. If this was to materialize, the type of cRSSs we chose will assist us in better understanding some mechanistic aspects of these erroneous events.

1. A 12 or 23 RSS with only a heptamer: to test for role of the nonamer because a lot of cRSSs usually have poor nonamer conservation than heptamer, if at all.

2. A 12 or 23 RSS with modified heptamer, CA**T**AGTG: to test the requirement for a 'CAC' in the presence of the mutants. If the RAG2 mutants are not as strict as wild-type RAG2, this can affect the identification of cRSSs in our tumor data.

3. A 12RSS with only nonamer: to test whether a perfect nonamer, on its own, can participate in recombination. Biochemical studies showed that an oligo with only a nonamer could be nicked but not fully cleaved (136). It is of interest whether the RAG2 mutants might overcome this block. (23RSS with only nonamer was not tested, a decision based upon our results with a heptamerless 12RSS, as detailed later).

As a first step we needed to assess the recombination efficiency of the RAG2 mutants with the inversional substrate harboring 12 and 23 consensus RSSs. We found that both RAG2FS and Neut mutants have increased recombination efficiency that is significantly better than a FLR2 (Fig 3.6B). This was not due to a hyper recombination activity of the mutants because examination of deletion

events, for instance, showed similar recombination efficiency (Fig 3.6A). We then compared the ratio of inversion (INV) vs HJ junctions generated. Again, both mutants had elevated HJ recombination events vs FLR2 (pv<0.003, Fig 3.6C). We therefore deduced that the C terminus of RAG2 inhibits unfaithful rearrangements that can lead to loss of DNA material. This result is in accordance with increased HJs seen in splenocytes from RAG2 Core mice (84).

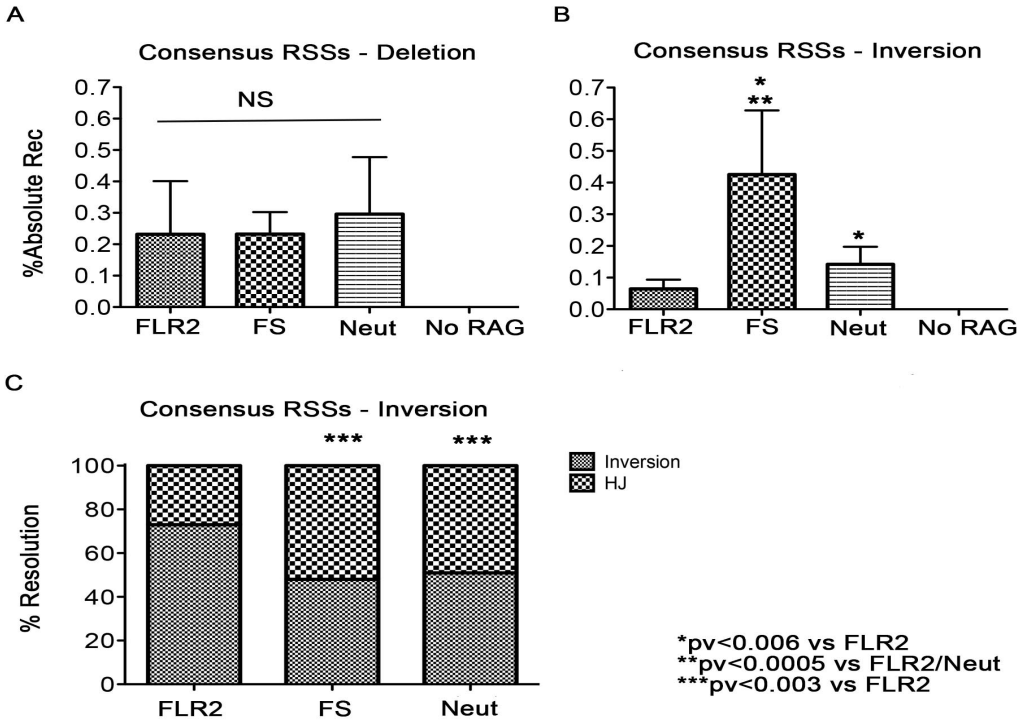


Figure 3.6 pJH299 recombination. Inversional (pJH299) and deletional (pJH290) substrates were transfected into RMP41 cell line together with FLR1 and indicated RAG2 version. 'No RAG' is transfection of the substrate alone. A+B. Recombination efficiencies were measured 48h post transfection as described in methods. C. PCR was carried out to identify individual recombination events via inversion or HJ resolution. 8 PCR products from each RAG2 version were verified by Sanger sequencing.

The significant difference in recombination efficiency with the RAG2 mutants ruled out straightforward comparison of absolute recombination ability between the different RAG2 mutants when using various cRSSs. To overcome this hurdle we had to normalize the absolute recombination result obtained with the tested cRSS to that of consensus RSSs for each RAG2 version.

Testing the different cRSSs we saw that all RAG2 versions have the same recombination efficiencies and that they can recombine a modified 12RSS better than a modified 23RSS (Fig 3.7). Comparing 12 to 23 RSS, Neut showed a significant difference throughout the analysis; FS was variable and FLR2 showed no significance. The lack of significance stems from the large standard deviations caused by the variability between experiments, as these are rare events. Nonetheless, as observed from the graphs the trend exists. Hence, it appears that 1) The C terminus of RAG2 does not affect RSS specificity and 2) A modified 23RSS results in lower recombination efficiency than a modified 12RSS. This phenotype, though not dependent on the C terminus, is exacerbated when mutated. However, the difference between the 12 and 23 RSSs turned out to be more complex than we initially thought (described later).

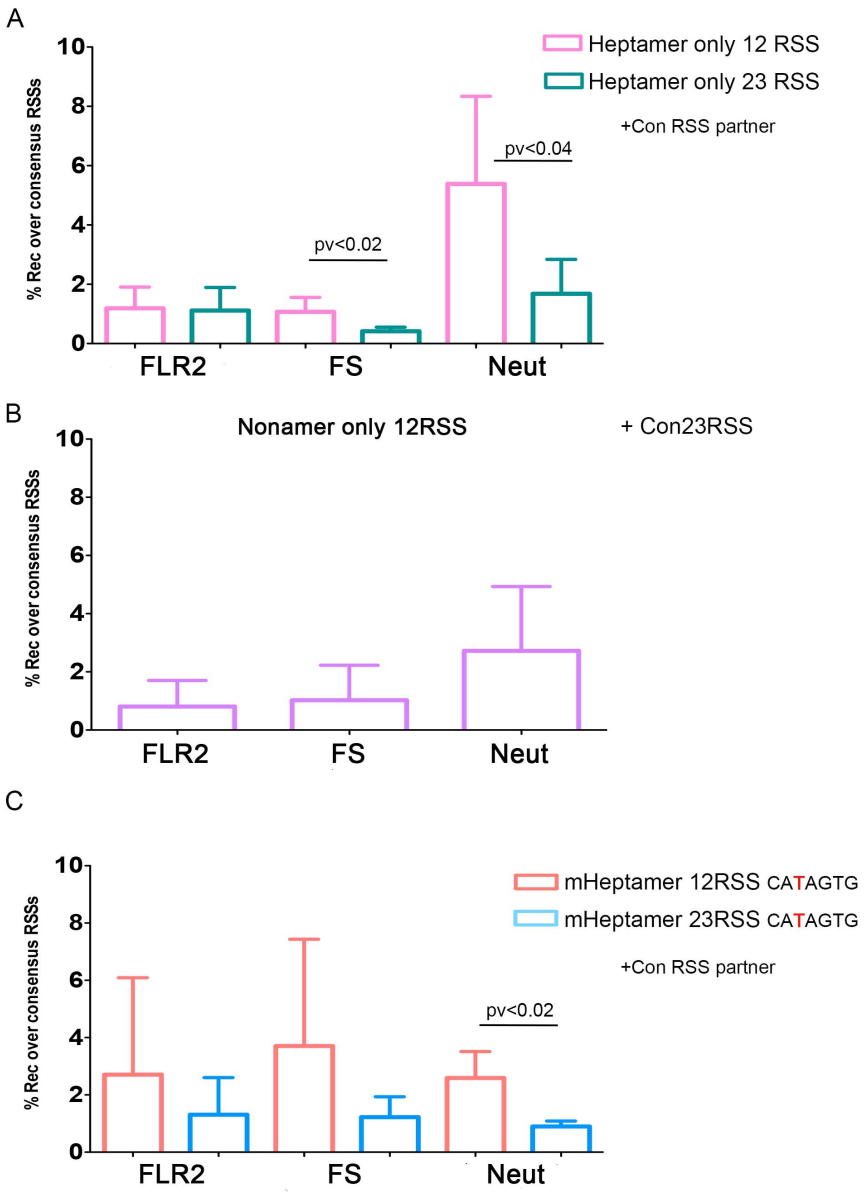


Figure 3.7 Recombination efficiency with different cRSSs. cRSSs substrates were transfected into RMP41 cell line together with FLR1 and indicated RAG2 version. 'No RAG' is transfection of the substrate alone. Recombination efficiencies were measured 48h after transfection as described in methods. A. Heptamer only substrate, the partner RSS is always a consensus one. B. Nonamer only 12 RSS the partner 23RSS is a consensus one C. mHeptamer RSS, the partner RSS is always a consensus one.

We next wanted to know whether the different RAG2 versions might change the ratio of INV to HJ when encountering different cRSSs. To that end we sequenced over 300 junctions from the different substrates and mutants. This approach turned out to be very informative, not just for our initial question, but also for understanding other aspects of the recombination process.

For clarity I grouped the results as 12 vs 23 cRSSs:

23 RSS modifications with consensus 12RSS:

The different versions of RAG2 behaved similarly with altered 23RSSs (Fig 3.8 B+D). Total HJ formation increased for FLR2 and FS but was not significantly different vs the one observed with 12/23 consensus RSS (Table 4). A significant difference was seen only with the mHeptamer 23 RSS (CA**T**AGTG) for FLR2 and FS but not Neut. However, this result was not consistent across the RAG2 versions or with the other cryptic 23 RSS excluding a meaningful interpretation of its significance. Thus, it seems that overall 23 cRSS usage by the various RAG2 versions does not result in increase in unfaithful recombination events.

The sequence analysis revealed more than just junctional resolution. In the case of the 'heptamer only 23RSS' half of the recombination events on average recombined using only a heptamer while the other half recognized the heptamer in an opposite orientation resulting in a coding joint (Fig 3.8B). With mutated heptamer 23RSS almost 100% of all recombination used this atypical heptamer (Fig.3.8D). These data suggest that the RAG complex, independent of whether the RAG2 C-terminus is present or absent, behaves differently in respect to the changes in the RSS regions. A very poor heptamer is better tolerated than a complete absence of a nonamer in a recombination event.

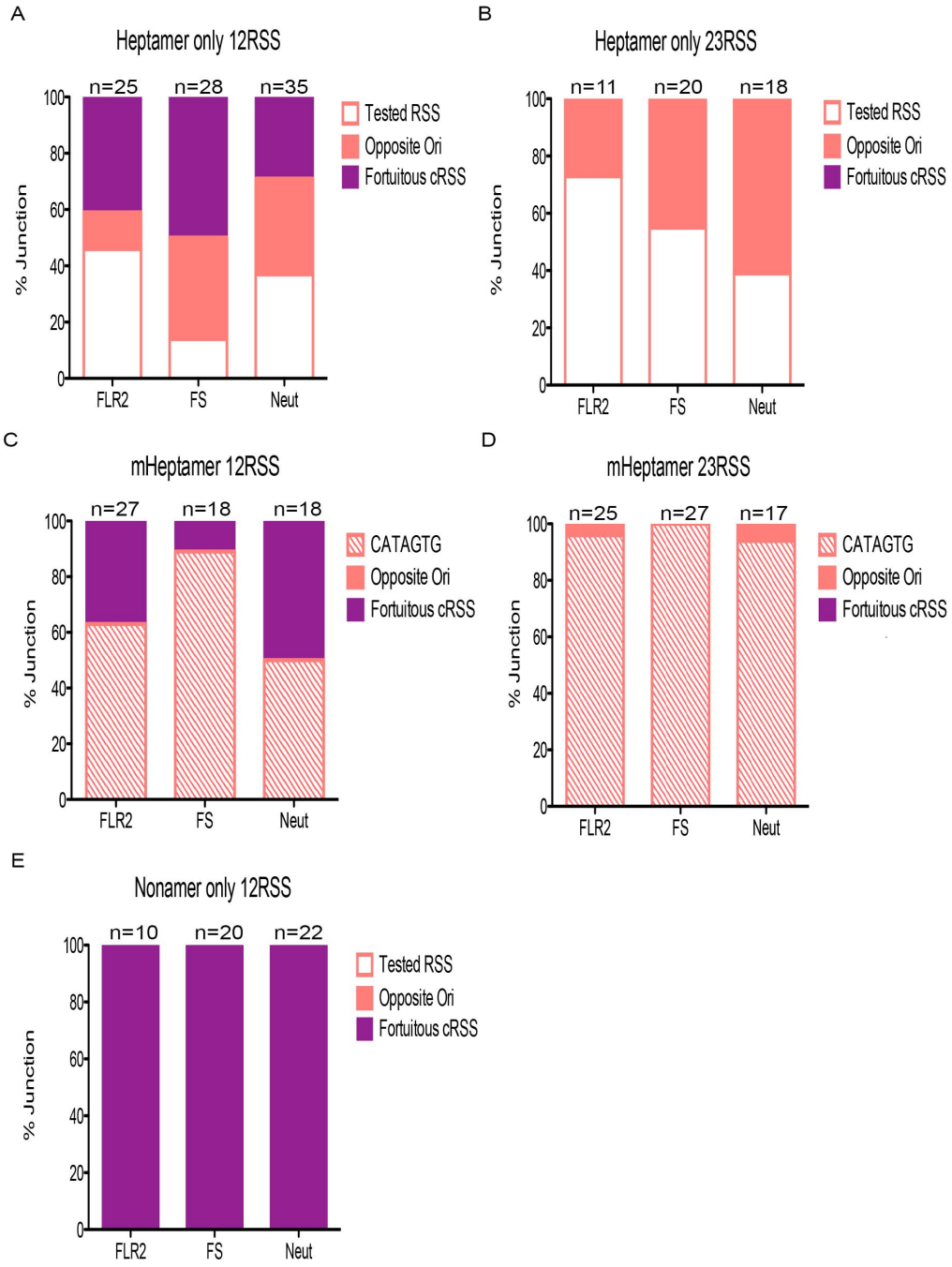


Figure 3.8 Sequence analysis of junction with tested cRSSs. Colonies exhibiting double resistances were grown and Sanger sequenced. Calculation was done out of total events for each situation. A. heptamer only 12RSS, B. heptamer only 23RSS, C. mHeptamer 12RSS, D. mHeptamer 23RSS, E. Nonamer only 12RSS. N indicates the number of junction sequenced.

	pJH299		Change on 12RSS		Change on 23 RSS	
	INV	HJ	INV	HJ	INV	HJ
FLR2	53 (73%)	20 (27%)	17(61%)	11(39%)	17(53%)	15(47%)*
FS	30(47%)	33(53%)	13(65%)	7(35%)	11(29%)	27(71%)*
Neut	28(51%)	27(49%)	13(62%)	9(38%)	14(61%)	9(39%)

Table 4. . Junctions' resolution with the tested cRSS. Inversional (INV) versus hybrid joints (HJ) were assessed by Sanger sequencing. Only junctions form pJH299 were assessed via PCR and several were validated by Sanger. The junctions are grouped by 12 or 23 tested cRSS tested and compared to consensus RSSs (pJH299). Absolute number and percentage are presented. * Significant increase in HJ with mHEptmaer (pv<0.05) but not heptamer only tested cRSS.

12 RSS modification with consensus 23RSS

As seen for the 23RSS, the ratio between the INV/HJ using different 12RSSs did not change when compared to consensus ones (Table 4). Nonetheless, unlike the cryptic 23RSSs, modification in the 12RSS revealed a distinctive behavior of the FS mutant (Fig 3.8 A+C+E).

Using a 'Heptamer only 12RSS' we found that FLR2 and Neut used this cRSS in 46% and 37% of recombination events respectively, while the FS used it significantly less in only 14% of the time (pv<0.02, Fig 3.8A). All other recombination events were due to other cryptic 12RSSs comprised of various fortuitous cRSS in the plasmid, and the tested cRSS itself used in the opposite orientation. The decreased usage of 'Heptamer only 12RSS' by the FS mutant

implies that in the absence of the C terminus the existence of identifiable nonamer is important to allow recombination.

The results we observed for FS with a mutated heptamer on 12RSS were the opposite with regard to other cRSS usage (Fig 3.8C). The FS was able to utilize a poor heptamer, the tested cRSS, for 90% of recombination events. The inability to discriminate a 'CAC' from 'CAT' was most profound with FS ($p<0.03$ vs Neut and $p<0.09$ vs FLR2 where this result was 1 junction away from significance). In the first situation, when a perfect heptamer with **no nonamer** was tested, the FS mutant preferred to recombine different cRSS all together. However, when given a very poor heptamer with a **perfect nonamer** it recombined it very efficiently. Though the substrate had a very poor heptamer, some sort of a heptamer is needed. Analysis of 'Nonamer only 12RSS' showed that all mutants underwent recombination using a fortuitous cryptic 12 RSS in the plasmid (Fig 3.8E). In aggregate, the data from cryptic 12RSSs show that in the absence of the C terminus the existence of identifiable nonamer, even on the expanse of a perfect heptamer, is important to allow recombination.

Our analysis revealed two additional intriguing features that will be discussed in the next section. First, a difference in handling the tested 12 or 23 RSSs. The spectrum of fortuitous cRSSs used with the tested 12 or 23 RSSs was significantly different ($p<0.0001$). In the 'Heptamer only 23RSS' substrate **all** the fortuitous cRSSs stemmed from reading the tested RSS in the opposite

orientation. In the ’Heptamer only 12RSS’ substrate more than half of cRSSs were derived from different regions in the vector (Fig 3.8A vs B). The second feature relates to recognizing the heptamer in an opposite orientation. Given the ‘Heptamer only’ substrates (Fig 3.8 A+B), the RAG complex used the heptamer in both orientations. However, in the case of a mutated heptamer (Fig 3.8 C+D), only 2/132 junctions were formed with the heptamer in the opposite orientation.

Discussion

Illegitimate V(D)J recombination was reported in multiple human and mouse lymphomas (91-95). These events can lead to genomic lesions of all kinds and contribute to or even initiate tumorigenesis. Therefore, a deep understanding of the mechanisms that cause them is very important to enable us to devise ways to inhibit them.

Our results show that the C terminus of RAG2 plays a very important role in preventing aberrant rearrangements by the RAG complex. Such events were found both lymphomas from $RAG2^{FS/FS};p53^{-/-}$ and healthy thymi from $RAG2^{FS/SF}$ mice. Almost a 100% of tested mice had an intrageneic deletion in Nocth1 and Bcl11b genes. This means that erroneous events are not just random event that were positively selected in a tumor, but are a cause of the RAG complex activity. One possibility for an increase in off target events in the absence of the RAG2 C terminus is relaxation in RSS specificity. This was shown to be the case in biochemical studies comparing FLR2 and Core RAG2 with consensus or suboptimal RSS (140). However, our functional examination of several cRSSs with different RAG2 versions showed that the C terminus of RAG2 does not seem to profoundly or systematically relax specificity in our system. This was also the result of a comprehensive analysis comparing cRSSs between $p53^{-/-}$ and $RAG2^{del352/del352};p53^{-/-}$ lymphomas (Mijuskovic M, in preparation). Thus, the frequent illegitimate recombination we observed in our mouse models were due

to other reasons. One explanation can be a difference in chromatin accessibility that allows more and/or different regions of the DNA in truncated RAG2 mice to be cleaved and recombined. For example, a pre-oncogene can be accessible for recombination in the mutated mice but not in wild-type mice (with FLR2) leading to the accelerated lymphomas. A second reason might be differences in the epigenetic make up landscape. As mentioned earlier, the PHD domain of RAG2 inhibits recombination unless H3K4me3 is present (135). The FS allele lacks the PHD domain overriding this level of regulation resulting in more erroneous recombination. This might be the case with the LMO2 and TAL1 cRSSs we tested because extrachromosomal substrates normally have high levels of H3K4me3 (141,142). A third cause can be variation in developmental stage between wild-type RAG2 and mutant RAG2 mice. The latter are blocked at the pro to pre T/B cell stage. Hence, the RAG complex presence is lengthen because the cells are trying to rearrange the antigen receptor locus and progress. This can statistically increase the chance of cleavage at off target sites in these cells. An argument against this is the fact that Core RAG1 mice, which exhibit the most sever block in that pro to pre stage, when crosse to $p53^{-/-}$ are devoid of the genomic instability seen in $RAG2^{del352/del352}$;$p53^{-/-}$ (84). Lastly, it might be that FLR2 activity is similar to the truncated RAG versions in terms of recognition and cleavage of cRSSs but not joining and this is why we cannot detect various structural variants in wild-type mice. The work presented here in chapter one

support this idea via the ability of the FS allele, but not wild-type RAG2, to repair ends via the ‘a-NHEJ’ pathway(s).

Our cRSS analyses shed light not only on the role of the RAG2 C terminus in illegitimate V(D)J recombination but also on aspects of V(D)J recombination in general. First, our results strengthen the notion that a functional RSS is a summation of all its regions rather than each part individually. None of the fortuitous cRSS used had a perfect heptamer nor nonamer. Second, our data support the model of initial synapsis with a 12-RSS by the RAG complex, followed by capture of a 23RSS (143). Only the tested cryptic 12RSSs, but not the 23RSSs were underutilized in favor of a whole different cRSS. This points to the importance of a good 12 RSS probably because it indeed initiates the recombination reaction. Once a good 12RSS is captured, greater variations on a 23RSS can be tolerated. One can argue that the usage of other cryptic 23RSS is not observed because the plasmid is devoid of them. However, at least 2/8 cryptic 12RSS can also qualify as a cryptic 23RSS. Additionally, capture of a 12 RSS in the absence of a C terminus is highly dependent on an identifiable nonamer. In the presence of FLR2, the nonamer does not have much impact on the recombination process. This is also seen in our tumor data in which the heptamer is much more conserved than the nonamer in the cRSSs. Unlike with FLR2, in the absence of the C terminus an RSS with a strong nanomer will prevail. This suggests that good contact of the RAG1 with the DNA in that

scenario is critical, probably compensating for a signal given by the C terminus to proceed with recombination. Third, we show data that can relate to the RAG complex positioning on the RSS. In the presence of a strong nonamer a heptamer can only be read in one orientation. We postulate that the RAG complex is becoming 'locked' in a certain configuration that directs its alignment with the heptamer. This might not be influenced by the RAG2 because contact with the nonamer is mediated via RAG1 (69,144). Not having a nonamer allows more flexibility in the RAG complex positioning on the DNA so a heptamer can be read in both orientations. This is probably due to the palindromic nature of the heptamer; the RAG complex cannot tell one way from the other, as there is no polarity to the sequence. However, this flexibility might come with a price of complex stability with the DNA because almost half of the junctions in 'heptamer only 12RSS' were via fortuitous cRSS in the plasmid. It might be that the RAG complex, in the presence of a good heptamer, 'scans' the DNA for a nonamer or there are multiple steps of association and disassociation until a permissive configuration is achieved. Regardless of the mechanism, the ability to utilize a heptamer in both orientations is very hazardous for genome integrity as such heptamer can be a partner as a 12 or 23 cRSS.

We wanted to examine if cRSS found in our tumors agree with our assumption that a good nonamer on 12RSS is important in the presence of the FS allele. We found most of our annotated cRSSs to agree with it. Only one junction (#21)

might not qualify while another junction (#30) that we did not consider to be mediated via cRSS might actually be one when looking at their nonamers (Table 3).

The increase in illegitimate V(D)J recombination in the absence of the RAG2 C terminus is not due to relaxation in RSS specificity and can be due to a variety of other reasons as mentioned above that should be explored experimentally. Nonetheless, we did observe that RAG2 with truncated C terminus prefers usage of cRSSs with identifiable nonamers.

CHAPTER 4

Discussion and future directions

Findings overview

In this body of work we observe that repair pathway choice is enforced during V(D)J recombination *in vivo,* in both c-NHEJ-proficient and c-NHEJ- deficient mice, and that RAG2's C-terminus is critical for this regulation. Surprisingly, we discover that the way in which a-NHEJ handles broken DNA ends is strongly context-dependent: signal joints show reported features of a-NHEJ, whereas coding joints do not. In fact, joining of RAG-generated coding ends by a-NHEJ in our system produced junctions that were largely indistinguishable from those arising from c-NHEJ. Whole genome sequencing of lymphomas derived from $RAG2^{FS/FS};p53^{-/-}$ mice showed that half of the genomic lesions similarly lack distinctive features of a-NHEJ. However, some lesions, including those in known oncogenes, appear to arise from ectopic recombination between DNA sequences fortuitously resembling RSS. Such genomic lesions are seen in models with wild-type RAG2, but they are both increased and involve, for the most part, distinct genomic regions in our RAG2 C terminally truncated mice. This difference is not due to relaxation in RSS specificity by the truncated RAG2 and may be due to other reasons or combination there of, that need to be further investigated.

Understanding alternative NHEJ

Despite great efforts from multiple labs for almost a decade to try to define aNEHJ, we are still in the dark. The best definition to date is 'repair in the absence of c-NHEJ'. The lack of consistent phenotypes across different models (e.g ku80 vs XRCC4 deficiency (31), V(D)J vs CSR) hinders our ability to define a pathway or even multiple pathways distinct from c-NHEJ. The work described here and others suggest that 'a-NHEJ' probably represents different repair outcomes depending on the environment of the breaks, thus there are different flavors of NHEJ (145). For example, coding joints retrieved from our $RAG2^{FS/FS}$;$KU80^{-/-}$ mice could have been generated via PARP1 in combination with the rest of the c-NHEJ machinery that is present in the cell (47). This might also be the reason our junctions looked similar to junctions formed by c-NHEJ. It would be interesting to test this hypothesis with our double mutant $RAG2^{FS/FS}$;$Ku80^{-/-}$ mice by crossing them to PARP1 null mice to observe if we can still rescue V(D)J recombination.

The importance of defining and understanding a-NHEJ repair has clinical implications. If indeed a-NHEJ is the dominant repair mechanisms leading to tumorigeneic genomic lesions, identifying the proteins that facilitate it would be valuable for cancer treatments and gene therapy. Moreover, it is also necessary to understand what, if any, are the physiological roles and normal functions of the a-NHEJ repair pathway.

Increase in proteins associated with a-NHEJ, PARP1 and Ligase III, has been reported in antiestrogen-resistant human breast cancer (54), positioning such factors as potential therapeutic targets. Indeed, such tumor cells were eradicated by a combination of PARP1 and DNA ligase III inhibitors (54). Interestingly though, single inhibition of either one was not enough to confer significant toxicity (54). PARP1, a potential a-NHEJ factor, is also being clinically tested in patients with breast cancer that are deficient in HR due to mutations in BRCA1/2 genes (146). Although this treatment looks promising in certain patients, the exact mechanism that underlines this effect is still unclear (146). One explanation, in the context of a-NHEJ, would suggest that in the absence of HR, DSBs that are generated could become occupied by PARP1, turning on a-NHEJ repair leading to genomic lesions that contribute to tumorigenesis (9,10,54). Hence, inhibiting PARP1 would leave these breaks unrepaired directing the cell to apoptosis. Alternatively inhibiting a-NHEJ could increase c-NHEJ repair, which was shown to suppress tumorigenesis (9,10). One study showed that PARP1 inhibitors indeed activated c-NHEJ, however, this activation led to genomic instability that then sensitized the lethality in HR deficient tumors (147). The PARP inhibitors example demonstrates the complexity we are facing when trying to decipher the role of a-NHEJ in tumorigenesis. While HR deficient breast cancer cells benefited from treatment with PARP inhibitors alone the antiestrogen-resistant tumors required combined therapy with PARP1 and ligase III inhibitors (53,147). This supports the idea that a-NHEJ probably encompasses several pathways

(68,148). Moreover, in HR deficient cells the c-NHEJ repair mechanism was responsible for genomic instability (147). Hence, a clinical approach to shift DNA repair from a-NHEJ to c-NHEJ might not always be beneficial. Notably, PARP inhibition was beneficial in less than 50% of HR deficient breast cancer cells implying that other factors might need to be targeted to inhibit this tumor (148). Further investifgation of a-NHEJ would most likely uncover some of these factors. So far, most of the factors associated with a-NHEJ also participate in other established repair mechanisms (e.g PARP1 in base excision repair). Therefore, a more detailed investigation of a-NHEJ might discover specific factor(s) that could then be inhibited to eliminate collateral damage caused to other repair mechanisms.

Gene therapy is another field that would benefit from better understanding of a-NHEJ. As there is no room for error when manipulating the genome it is imperative that we pursue the deepest knowledge possible.
'Replacement' of a damaged gene with a corrected version of itself, or 'replacement' of a good gene with a mutated one (to study the gene's function), are the most straightforward applications of gene targeting. These processes are thought to benefit from inhibiting NHEJ repair in favor of HR to decrease random integrations of the vector into the genome (149,150). In this scenario it is not enough and even dangerous to block only c-NHEJ. In the absence of c-NHEJ, a-NHEJ is turned on and can promote chromosomal aberrations (9,10,37,45,48).

Thus, while trying to correct a specific mutation we might actually be promoting tumorigenic event(s).

Another application is targeted disruption of a gene that promotes disease, like CCR5 in HIV infection (151). This process is dependent on c-NHEJ repair to create micro-changes in the gene to render it nonfunctional (151). Just like in the previous scenario, suppression of a-NHEJ would reduce the chances of gross genomic aberration upon DNA cleavage by the nuclease. This scenario emphasizes the need to fully understand whether a-NHEJ is a distinct repair pathway from c-NHEJ or whether it is a combination of pathways that might share components with the c-NHEJ repair machinery. If the latter is true then using this methodology might be more challenging than initially thought.

Until we better understand a-NHEJ repair pathway(s), perhaps a functional definition is more suitable; ku80 dependent, PARP1 dependent etc. In this case each group would comprise all possible outcomes of that condition. Either approach wiould also disqualify, once and for all, certain junctional quality as a feature of one repair pathway. In support of such an approach is a recent paper showing that even in the absence of Ku80 the majority of DSBs in a human cell line were repaired via Ligase IV (152).

There are different experiments that can be pursued to gain deeper understanding of what a-NHEJ is/are. Taking a biochemical approach, we can test purified DNA repair proteins for their end joining activities. In this fashion we

can control the order or composition of the proteins and also test different types of breaks (I-Sce vs V(D)J). However, numerous disadvantages exist (not all proteins can be purified, impaired protein folding, protein modification, no chromatin etc) that ultimately might not grant us with a clearer picture.

A different approach is to carry out RNASeq of different cancer types that show upregulation or downregulation of factors associated with a-NHEJ or c-NHEJ respectively (52,54). Comparison of these data sets to data from a control, non-tumorigenic tissue might result in discovering new proteins that participate in a-NHEJ and/or shed a different light on known proteins that were not initially considered to be involved in a-NHEJ. Alternatively, we can perform whole exome sequencing on these tumors and look for mutations that can confer a selective advantage. Our focus would be on DNA repair genes and additional genes that their activity is still not fully elucidated but might be participating in DNA repair. Potential factors identified via RNASeq or the exome sequencing would be followed by a validation stage to eliminate false positive calls. To that end various established end-joining assays (e.g I-Sce, V(D)J recombination, CSR etc) can be performed to test whether the factor indeed affects end-joining repair. Western blot and immunoprecipitation experiments would be used to demonstrate expression patterns and possible interactions that can assist in elucidating mechanistic elements. Lastly, animal studies with over expression or knockdown of the selected factors should be considered in order to study the targeted factor in a more physiological setting.

Additional experiments that should be explored include characterization of cells deficient for both c-NHEJ and associated a-NHEJ factors. Zhang et al showed that translocation events were still occurring in cells deficient for Ku70 and CtIP (37), implying that additional factors or yet, an unidentified combination of known factors can promote such aberrant end joining. This approach should be tested using different combination of deficiencies in known DNA repair factors. In turn, such data would generate a matrix of potential a-NHEJ cascades/platforms. To illustrate that, we can think of the following scenario: cells deficient for ku70 & ligase III or ku70 & CtIP exhibit equal translocation frequencies, while cells deficient for all three factors show reduced translocation frequencies. This result would suggest that in the absence of ku70, CtIP and ligase III function in different a-NHEJ repair pathways. Any meaningful combination of deficient cell lines should be subjected to microarray analysis to evaluate changes in the transcriptome versus wild-type and singly deficient cells to illuminate factors that might participate in a-NHEJ repair. As translocations are not the only abnormal events in tumors, exploring end joining substrate with a deletion configuration would also be valuable. Comparison of factors identified with the translocation substrate with that of the deletion substrate could teach us about the mechanism(s) that drive(s) these different lesions and possibility aid in cancer treatment. One can theorize that translocation events might be driven by a-NHEJ repair while deletion events may be equally driven by classical or alternative NHEJ repair.

Lastly, our results from RAG2 mouse knockins ($RAG2^{FS/FS}$, $RAG2^{del352/del352}$) show that these mutants have the ability to access a-NHEJ repair pathway(s) both in c-NHEJ proficient and deficient settings. However, we do not know what factors allow or participate in such repair. It would be of interest to cross our knockin mice to mice deficient in potential a-NHEJ factors and to assess coding and signal joint formation. Work done previously in our lab showed that NBS1 is needed for a-NHEJ repair by the RAG2 FS mutant on an extrachromosomal substrate (153). Thus, we can theorize that crossing of $RAG2^{FS/FS}$ to a hypomorphic version of NBS1 would decrease both aberrant signal joint formation and interchromosomal rearrangements between different TCR loci that are considered indicative of a-NHEJ. Such a result would strengthen the role of NBS1, and perhaps the MRN complex as a whole, in a-NHEJ repair (43,153).

In the pursuit to improve one's life through cancer treatment or gene therapy it is vital to gain better insights of what is/are a-NHEJ repair. The knowledge accumulated through experiments described above and others would assist us in designing better strategies to achieve this goal.

Tumorigenic anecdotes through the lens of whole genome sequencing

There are various types of genomic lesions ranging from point mutations to large gross abnormalities like translocations. There is no linear correlation between the size of the lesion and its damage to the cell or the organism as a whole. For example, a point mutation in a critical gene can be more deleterious than a reciprocal translocation. Whole genome sequencing (WGS) of two tumors in this study raises several interesting insights regarding the process of lymphomagenesis.

1) As whole genome sequencing is a luxury of the last few years, a lot of the attention in cancer research was initially concentrated on identification of translocations by cytogenetic analysis. Technically, micro or small lesions are difficult to impossible to identify. Abnormalities such as deletions, insertions, duplication and inversions were generally unaccounted for. As a result, great efforts were aimed at identifying and characterizing translocations formation (10,33,37,48,68). However, as data from WGS is accumulating it appears that intra-chromosomal genomic lesions, specifically deletions, are substantially more frequent in lymphomas than are translocations (Mijuskovic M, in preparation, 97). 67% of lesions found in our $RAG2^{FS/FS};p53^{-/-}$ lymphomas were deletions, 27% inversions and only 6% translocations. Indeed, detection of lesions by WGS is limited by depth of coverage and it is possible that more translocations exist that

are not detected. However, it implies that these translocations are not particularly advantageous for the tumor and are probably more of random ('passenger') lesions than oncogenic ones.

2) Capture of different types of lesions might also shed a different light on the role of c-NHEJ in tumoregenesis. This repair mechanism is thought to suppress genomic lesions due to translocations seen in c-NHEJ deficient backgrounds (9,10). Moreover, sequences of these translocations showed increased usage of microhomologies associated with a-NHEJ (9,11,12). In this study, looking at deletional events retrieved from c-NHEJ proficient animals the majority of junctions looked similar to ones formed via c-NHEJ repair. On the other hand, inversional events exhibited short microhomolgies associated with a-NHEJ in 80% of the junctions. (We could not compare translocations because we only captured two translocation events and they were devoid of microhomologies). Assuming that microhomologies are indeed a feature that characterizes a-NHEJ (which is not the case as shown in this study) this repair mediated only part of the genomic lesions. All other events, including translocations, could not be confidently assigned to any repair pathway. As a result, the role of c-NHEJ in tumorigenesis must not be underestimated because deletions, the majority of the genomic instability, could have been mediate by c-NHEJ rather than a-NHEJ.

3) The granular analysis provided by sequencing of the lesions' breakpoints also opened a great window to understand mechanisms of tumorigenic rearrangements. For example, RSS-like sequence or presence of an AID recognition site in the vicinity of resolved ends might provide strong evidence or footprints of the process that facilitated it. Sequence analysis of 51 human ETV6-RUNX1 ALL determined that illegitimate RAG activity was the predominant driver leading to transformation (97). A comprehensive analysis of 11 thymic lymphomas done in the our lab show that at least 30% of aberrant lesions are mediated via illegitimate V(D)J recombination (this work and Mijuskovic M, in preparation) because we could identify cRSSs in the vicinity of the break. Some of the genes involved in these rearrangements are already implicated in tumorigenesis while others have not been. Hence, such analyses have the power to uncover mechanisms as well as potential new tumorigenic genes that were not considered before. These new genes may become a diagnostic tool and even potentially a therapeutic target. Notably, by looking only at cRSSs, we might be underestimating the involvement of the RAG complex. It has been shown that the RAG complex could also cleave a non-canonical DNA structure (non-B DNA)(154). The t14;18 translocation in follicular B cell lymphoma is thought to arise through cleavage by the RAG complex at a non-B DNA structure on chromosome 18 (154). Hence, it is probable that the RAG complex may have mediated some of the other lesions that were devoid of cRRSs.

4) The evidence of aberrant V(D)J recombination in genomic lesions ties us back to the question of c-NHEJ and tumorigenesis. It is established that the c-NHEJ repair machinery joins V-D-J ends. Thus, V(D)J ends generated outside of the antigen receptor loci are more than likely to be joined by this repair mechanism as well. Hence, the aberrant V(D)J events we observed position c-NHEJ as a repair pathway that can predispose to tuomoregensis much like a-NHEJ is thought to be. Moreover, 50% of these events involved genes implicated in tumorigenesis further supporting the contribution of c-NHEJ repair to the lymphomas in our mouse models.

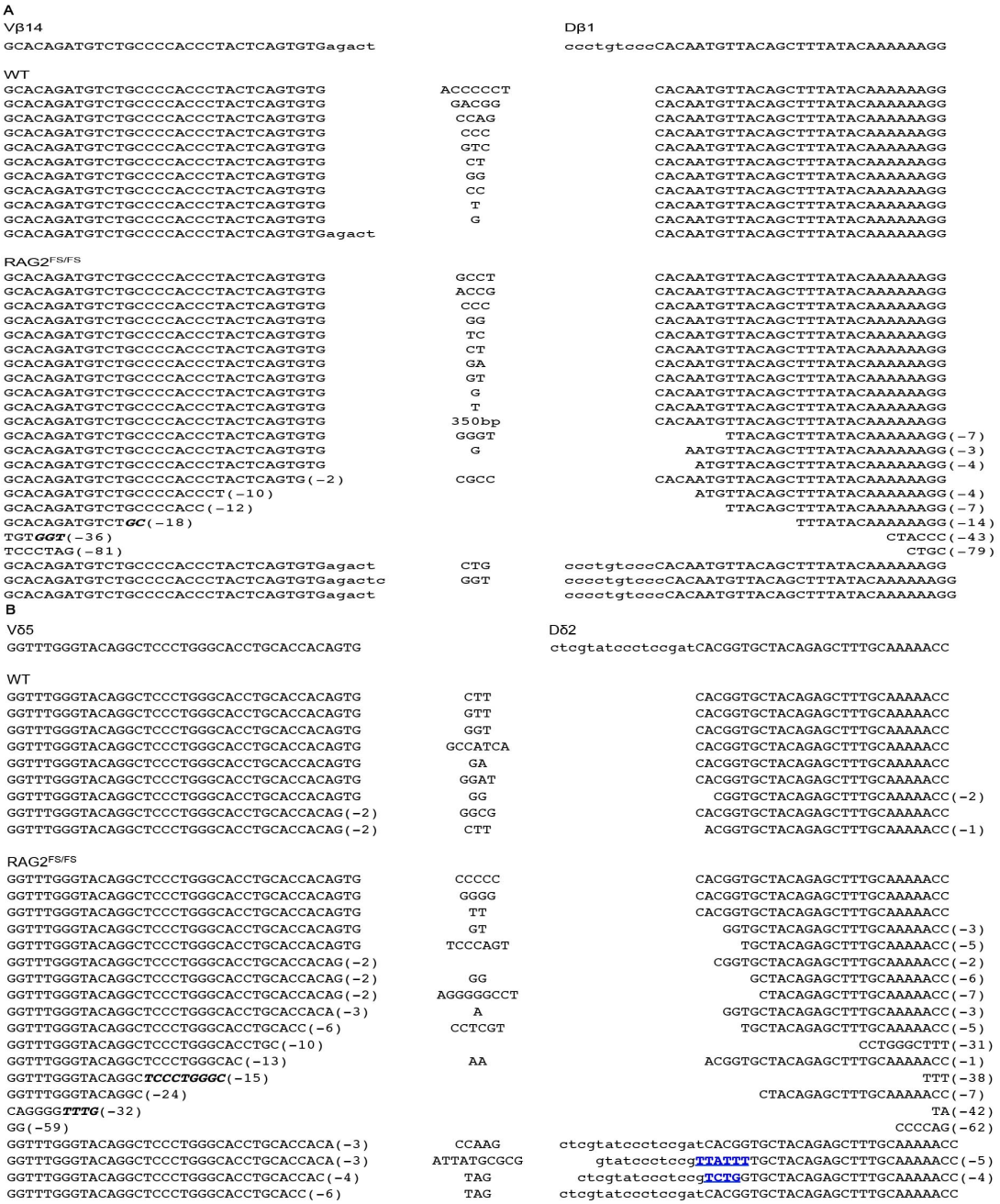
5) The substantial off target activity of the RAG complex, genome wide, should be a warning sign when attempting gene targeting. Several nucleases such as Transcription activator-like effector nuclease (TALEN) and Zinc Finger Nucleases (ZFN) are being studied for genome engineering. Although they are sequence-specific, like the RAG recombinase, off target activity of these nucleases has been reported (155). However, these off target sites were predicted a priori either by identifying sequence homology to a different gene or by prediction algorithms (155). In light of our WGS results, such biased approaches to test for off target activity are neither satisfactory nor prudent. Hence, it is of great importance that WGS be performed on cells engineered with these nucleases to assess, in an unbiased fashion, their full potential for off target activity. Preliminary work in our

lab show that WGS of a single clone engineered with TALENs revealed numerous unpredictable off target sites (Lindsay C, Roth lab).

WGS in our studies and ones carried out in human lymphomas (97) demonstrates its importance in better understanding tumor pathology. Hence, it might be valuable if tumors that share common features would be sequenced to try to identify patterns of genomic aberrations that can be informative. Once a large enough data set is compiled we then could screen for only targeted variants that were deemed important. WGS can also assist in understanding cancer evolution and metastasis. Sequencing of a tumor upon diagnosis and comparing it to tumor that reemerges could shed light on acquisition of chemo-resistance. Another scenario can be a comparison of primary vs. metastatic foci to understand changes in the tumor that conferred metastasis. It is obvious that in parallel to this genomic approach, important genomic lesions should be investigated biologically via microarray analysis, for example, to explore if and how they influence the biology of the tumor. Such analysis would help determine whether a certain variant can be used as a diagnostic and/or therapeutic target. From a clinical pint of view, WGS is still an expansive endeavor, yet advancement in technology and big data analysis are continuously decreasing cost and turnover time. Implementing WGS in a clinical setting has other challenges: quality of specimens, training health care professionals, meaningful communication of results to physicians/patients, and standardizing an

appropriate use of the technology. Nonetheless, if WGS is essential to provide the optimal care we should not let these challenges be more than a temporary obstacle.

Supplementary information



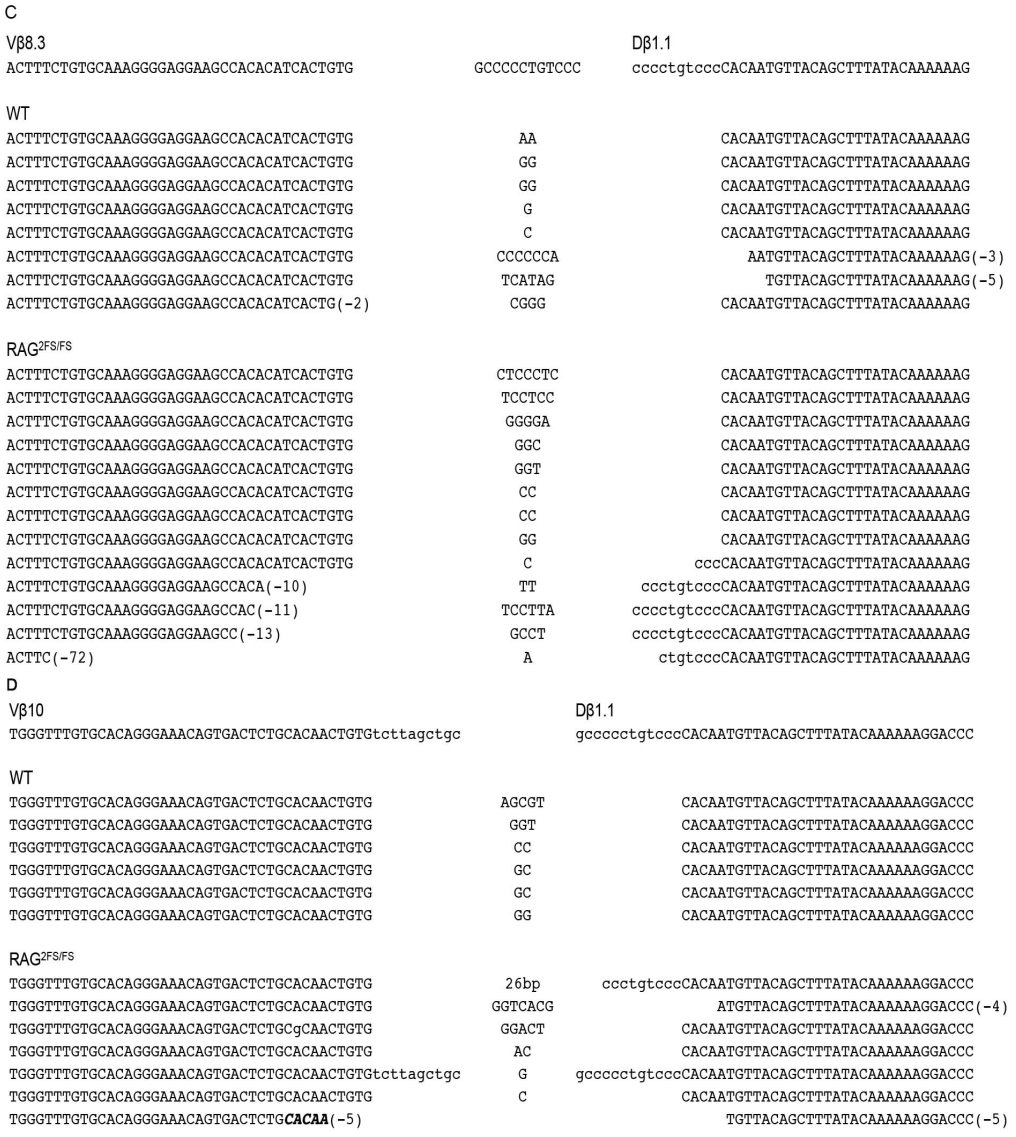


Figure S1. Signal joints sequences from wild-type and RAG2$^{FS/FS}$ mice

A. SJ from Vβ14-Dβ1. B. SJ from Vδ5-Dδ2. C. SJ from Vβ8.3-Dβ1.1 D. SJ from Vβ10-Dβ1. Germ-line sequence of each locus is shown at the top. Capital letters indicate the RSS and small letters were indicted are coding end region. Capital letters in the middle of the junction indicate N nt, deletions are indicated in parentheses, small letter indicates sequences from the coding region (miscleavage), capital bold italics are microhomology and underlined blue are open shut junctions.

A

		Dβ1/2			
Vβ8.2	GTACTTCTGTGCCAGCGGTGATG	GGGACAGGGGGC		AGTCAAAACACCTTGTACTTTGGGCCA	Jβ2.4
Vβ8.1	TATATTTCTGTGCCAGCAGTGATG	GGGACTGGGGGGGC		AACCAAGACACCCAGTACTTGGGCCA	Jβ2.5
Vβ7	TGTACTTCTGTGCTAGCAGTTTATC			CTCCTATGAACAGTACTTCGGTCCCGG	Jβ2.7
Vβ6	TTTTTCTCTGTGCCAGCAGTATAG				

WT

GTACTTCTGTGCCAGCGGTGATG	TC			AAGACACCCAGTACTTTGGGCCA(-4)
TTTTTCTCTGTGCCAGCAGTATAG		(-4)CTGGGG(-4)		GACACCCAGTACTTTGGGCCA(-6)
TATATTTCTGTGCCAGCAGTGATG	G	GGGACAGGG(-3)		AAGACACCCAGTACTTTGGGCCA(-4)
TATATTTCTGTGCCAGCAGTGATG	**C**	GGGGGG	AC	GAACAGTACTTCGGTCCCGG(-7)
TGTACTTCTGTGCTAGCAGTTTATC		(-4)CAGGG(-3)		ATGAACAGTACTTCGGTCCCGG(-5)
GTACTTCTGTGCCAGCGGTGAT(-1)			AATC	TATGAACAGTACTTCGGTCCCGG(-4)
GTACTTCTGTGCCAGCGGTGAT(-1)	A	(-6)GGGGGGGC	**GCC**GG	CACCCAGTACTTTGGGCCA(-8)
GTACTTCTGTGCCAGCGGTGAT(-1)		(-4)CTGGG(-5)	AA	GAACAGTACTTCGGTCCCGG(-7)
TTTTTCTCTGTGCCAGCAGTATA(-1)		(-5)AGGGG(-2)	A	ATGAACAGTACTTCGGTCCCGG(-4)
TTTTTCTCTGTGCCAGCAGTATA(-1)	CC	GGGGC	**G**	CCAAGACACCCAGTACTTTGGGCCA(-2)
TATATTTCTGTGCCAGCAGTGAT(-1)	C	(-4)CAGGGGG(-1)	GGTTAGC	ACAGTACTTCGGTCCCGG(-9)
GTACTTCTGTGCCAGCGGTGA(-2)		(-4)CTGGGGGGC	**GCC**GG	CCAAGACACCCAGTACTTTGGGCCA(-2)
TATATTTCTGTGCCAGCAGTGA(-2)		(-4)CTGGGGGGG(-1)		ATGAACAGTACTTCGGTCCCGG(-5)
TGTACTTCTGTGCTAGCAGTTTA(-2)		GGGGGC	G	ACCAAGACACCCAGTACTTTGGGCCA(-1)
GTACTTCTGTGCCAGCGGTG(-3)	T	(-4)CTGG(-6)	T**CT**	AGTCAAAACACCTTGTACTTTGGTGCG
GTACTTCTGTGCCAGCGGTG(-3)		GC	CCAA	TCAAAACACCTTGTACTTTGGTGCG(-2)
TGTACTTCTGTGCTAGCAGTTT(-3)		(-5)AGGGGGC	**G**	ACCAAGACACCCAGTACTTTGGGCCA(-1)
TGTACTTCTGTGCTAGCAGTT(-4)		(-3)ACAGGGG(-2)	T	CAAGACACCCAGTACTTTGGGCCA(-3)
TGTACTTCTGTGCTAGCAG(-6)	CC	(-2)GACAGGG(-3)	ATGG	GAACAGTACTTCGGTCCCGG(-7)
TGTACTTCTGTGCTAGCAG(-6)	CCT**CC**	GGGACTGGGGG(-3)	AT	CTATGAACAGTACTTCGGTCCCGG(-3)
TGTACTTCTGTGCTAGC(-8)		T		CAAGACACCCAGTACTTTGGGCCA(-3)
TGTACTTCTGTGCTAGC(-8)	T	(-5)TGGGG(-4)	A	AAGACACCCAGTACTTTGGGCCA(-4)
TGTACTTCTGTGCTA(-10)	CCCAG	(-5)TGGGGGG(-2)		ATGAACAGTACTTCGGTCCCGG(-5)

RAG2$^{FS/FS}$

GTACTTCTGTGCCAGCGGTGATG	**C**GT	(-4)CAGGG(-3)	TCCCC	AAGACACCCAGTACTTTGGGCCA(-4)
GTACTTCTGTGCCAGCGGTGATG	**C**	GG		GAACAGTACTTCGGTCCCGG(-7)
TATATTTCTGTGCCAGCAGTGATG	CAG	(-5)AGGG(-3)	ACATCT	AGACACCCAGTACTTTGGGCCA(-5)
TTTTTCTCTGTGCCAGCAGTATAG		(-4)ACGGG(-3)	TGGG	TATGAACAGTACTTCGGTCCCGG(-4)
TTTTTCTCTGTGCCAGCAGTATAG		(-3)ACTGGGGGG(-1)	ACGAGGA	GAACAGTACTTCGGTCCCGG(-7)
TTTTTCTCTGTGCCAGCAGTATAG		(-5)AGGGGG(-1)	G	TGAACAGTACTTCGGTCCCGG(-6)
GTACTTCTGTGCCAGCGGTGAT(-1)	T	(-4)CAGGGGG(-2)	GATAAG	CACCCAGTACTTTGGGCCA(-8)
GTACTTCTGTGCCAGCGGTGAT(-1)	CG	(-4)CTGGGGGG(-2)		ACCAAGACACCCAGTACTTTGGGCCA(-1)
GTACTTCTGTGCCAGCGGTGA(-2)	AGA	(-5)TGGGG(-4)	AT	TATGAACAGTACTTCGGTCCCGG(-4)
TATATTTCTGTGCCAGCAGTGA(-2)	C	GGGG	GATTGATCGGGG	ACCAAGACACCCAGTACTTTGGGCCA(-1)
TATATTTCTGTGCCAGCAGTGA(-2)	A	(-1)GGACTGGGGGGGC	TA	GAACAGTACTTCGGTCCCGG(-7)
TTTTTCTCTGTGCCAGCAGTAT(-2)	AAC	GGC	AG	AAGACACCCAGTACTTTGGGCCA(-4)
TGTACTTCTGTGCTAGCAGTTT(-3)	C	(-2)GACAG(-5)		CTCCTATGAACAGTACTTCGGTCCCGG
TGTACTTCTGTGCTAGCAGTTT(-3)	GTCT	(-4)CAGGG(-3)	ATGAG	GTACTTCGGTCCCGG(-12)
TGTACTTCTGTGCTAGCAGTTT(-3)		(-2)GACAGGGGGC	AGG	GACACCCAGTACTTTGGGCCA(-6)
TGTACTTCTGTGCTAGCAGTTT(-3)	G	AGGGGGC	**G**G	GAACAGTACTTCGGTCCCGG(-7)
TATATTTCTGTGCCAGCAGT(-4)	CCCT	(-2)GACAGGG(-3)	AGG	GAACAGTACTTCGGTCCCGG(-7)
TTTTTCTCTGTGCCAGCAGT(-4)	ACG	(-3)ACAGG(-4)		CTATGAACAGTACTTCGGTCCCGG(-3)
TTTTTCTCTGTGCCAGCAGT(-4)	GC	(-1)GGACTGGGGGGGC	**G**T	TGAACAGTACTTCGGTCCCGG(-6)
TGTACTTCTGTGCTAGCAGTT(-4)	C	GGGG		CAAAACACCTTGTACTTTGGTGCG(-3)
TGTACTTCTGTGCTAGCAGTT(-4)	C	(-3)ACTGGGG(-4)		AACCAAGACACCCAGTACTTGGGCCA
TTTTTCTCTGTGCCAGCAG(-5)		(-3)ACTGGGGGGGC	**GC**GGG	GACACCCAGTACTTTGGGCCA(-6)
TGTACTTCTGTGCTAGCAGT(-5)	CCT	(-3)ACAGGG(-3)		AACCAAGACACCCAGTACTTGGGCCA
TGTACTTCTGTGCTAGCAGT(-5)	AT**CC**	GGGACAGGGGGC	T	CACCCAGTACTTTGGGCCA(-8)
TGTACTTCTGTGCTAGCAGT(-5)		(-4)CTGG(-6)	ACAGT	TATGAACAGTACTTCGGTCCCGG(-4)
TGTACTTCTGTGCTAGCAGT(-5)	AC	(-2)GACTGGGGGGGC	**GC**TG	GAACAGTACTTCGGTCCCGG(-7)
TGTACTTCTGTGCTAGCAG(-6)	CTT	GGGACAGGG(-3)	A	GAACAGTACTTCGGTCCCGG(-7)
TGTACTTCTGTGCTAGCAG(-6)	CCCA	(-3)ACAGGG(-3)	TT	TGAACAGTACTTCGGTCCCGG(-6)
TGTACTTCTGTGCTAGCAG(-6)	CA	(-3)ACTGGGGGGGC	**GC**GGT	CCAAGACACCCAGTACTTTGGGCCA(-2)
TGTACTTCTGTGCTAGCAG(-6)	CCCC	GGGGG		CTATGAACAGTACTTCGGTCCCGG(-3)
TTTTTCTCTGTGCCAGCAGTATAGcacagtgga			A	CTATGAACAGTACTTCGGTCCCGG(-3)

B

Vβ10		Dβ1/2		Jβ2.1
TCTCTGTGCCAGCAGCTAAGA		GGGACAGGGGGC		TAACTATGCTGAGCAGTTCTTCG
		GGGACTGGGGGGGC		

WT

TCTCTGTGCCAGCAGCTAAG(-1)	**C**	GGGAC		AACTATGCTGAGCAGTTCTTCG(-1)
TCTCTGTGCCAGCAGCTAA(-2)	AGG	(-4)CAGGGGG(-1)	**TTA**	TAACTATGCTGAGCAGTTCTTCG
TCTCTGTGCCAGCAGCTA(-3)	TT	(-5)TGGGGG(-3)		ATGCTGAGCAGTTCTTCG(-5)
TCTCTGTGCCAGCAGCTA(-3)	**CC**	GGGACAGGGGGC	**GC**	GCTGAGCAGTTCTTCG(-7)
TCTCTGTGCCAGCAGCTA(-3)	TCCT	(-2)GACA(-6)	ACGA	TATGCTGAGCAGTTCTTCG(-4)
TCTCTGTGCCAGCAGCTA(-3)		TGGGGGGG	GGT	ACTATGCTGAGCAGTTCTTCG(-2)
TCTCTGTGCCAGCAGCT(-4)		(-4)CTGGGGGGG(-1)	A	TATGCTGAGCAGTTCTTCG(-4)
TCTCTGTGCCAGCAGCT(-4)	CGCC	GGGC	CT	CTATGCTGAGCAGTTCTTCG(-3)
TCTCTGTGCCAGCAGCT(-4)	CCCGG	(-4)CTGGGGGGG(-1)		AACTATGCTGAGCAGTTCTTCG(-1)
TCTCTGTGCCAGCAGCT(-4)	T	(-2)GACTGGGGGG(-2)		AACTATGCTGAGCAGTTCTTCG(-1)
TCTCTGTGCCAGCAGC(-5)	GC	(-3)ACTGGGGGGGC	**GC**TGA	ATGCTGAGCAGTTCTTCG(-5)
TCTCTGTGCCAGCAG(-6)		(-3)ACTG(-7)	AG	AACTATGCTGAGCAGTTCTTCG(-1)
TCTCTGTGCCAGCAG(-6)	TCAAG	GGGACAGGGG(-2)	AGAG	ATGCTGAGCAGTTCTTCG(-5)
TCTCTGTGCCAGCAG(-6)	TC	(-1)GGACTGGGGGGG(-1)	G	CTATGCTGAGCAGTTCTTCG(-3)
TCTCTGTGCCAGCAG(-6)	TCCC	(-2)GACTGGGGGG(-2)	TC	AACTATGCTGAGCAGTTCTTCG(-1)
TCTCTGTGCCAGCA(-7)		(-3)ACTGGGGGGG(-1)		CTATGCTGAGCAGTTCTTCG(-3)

RAG2$^{FS/FS}$

TCTCTGTGCCAGCAGCTAAGA	AG	GGGACTGGGGG(-3)	A	AACTATGCTGAGCAGTTCTTCG(-1)
TCTCTGTGCCAGCAGCTAAG(-1)		GGGG	TCTTC	CTATGCTGAGCAGTTCTTCG(-3)
TCTCTGTGCCAGCAGCTAAG(-1)	T**C**	GGGACAGGGG(-2)	T	TGAGCAGTTCTTCG(-9)
TCTCTGTGCCAGCAGCTA(-3)		(-4)CAGGGGG(-1)		ACTATGCTGAGCAGTTCTTCG(-2)
TCTCTGTGCCAGCAGCTA(-3)	TT	(-5)TGGGGGGG(-1)	GG	TAACTATGCTGAGCAGTTCTTCG
TCTCTGTGCCAGCAGCTA(-3)	T			ACTATGCTGAGCAGTTCTTCG(-2)
TCTCTGTGCCAGCAGCTA(-3)	TCT	GGGACT(-8)	A	AACTATGCTGAGCAGTTCTTCG(-1)
TCTCTGTGCCAGCAGCT(-4)	T	(-1)GGACAG(-5)		TAACTATGCTGAGCAGTTCTTCG
TCTCTGTGCCAGCAGCT(-4)	T	GGGACAGG(-4)		AACTATGCTGAGCAGTTCTTCG(-1)
TCTCTGTGCCAGCAGCT(-4)		GGGACA(-6)	A	AACTATGCTGAGCAGTTCTTCG(-1)
TCTCTGTGCCAGCAGCT(-4)	CCTTC	(-1)GGACTGG(-6)	**A**	TAACTATGCTGAGCAGTTCTTCG
TCTCTGTGCCAGCAGCT(-4)	C**CCC**	GGGACAGG(-4)		ATGCTGAGCAGTTCTTCG(-5)
TCTCTGTGCCAGCAGCT(-4)	CTCGT	(-4)CAGGGGG(-1)	**A**	TAACTATGCTGAGCAGTTCTTCG
TCTCTGTGCCAGCAGCT(-4)	T	(-5)AGGGGGC	TCG	ATGCTGAGCAGTTCTTCG(-5)
TCTCTGTGCCAGCAGCT(-4)	T	(-5)TGGGGGG(-2)	CCGGGA	ATGCTGAGCAGTTCTTCG(-5)
TCTCTGTGCCAGCAGCT(-4)		(-7)GGGGGGC	CG	ACTATGCTGAGCAGTTCTTCG(-2)
TCTCTGTGCCAGCAGCT(-4)	C	(-4)CAGGGGGC		CTATGCTGAGCAGTTCTTCG(-3)
TCTCTGTGCCAGCAGC(-5)	CC**CCC**	GGGACA(-6)	AT	ACTATGCTGAGCAGTTCTTCG(-2)
TCTCTGTGCCAGCAGC(-5)	**CC**	GGGACAGGGG(-4)		TATGCTGAGCAGTTCTTCG(-4)
TCTCTGTGCCAGCAGC(-5)	A**CC**	GGGACTGGGGGGG	T	CTATGCTGAGCAGTTCTTCG(-3)
TCTCTGTGCCAGCAGC(-5)	A	GGGACAG(-5)	C	AACTATGCTGAGCAGTTCTTCG(-1)
TCTCTGTGCCAGCAGC(-5)	AA**CC**	GGGACT(-8)	T	TAACTATGCTGAGCAGTTCTTCG
TCTCTGTGCCAGCAGC(-5)	CCC	(-2)GACTGGGGG(-3)	AT	CTATGCTGAGCAGTTCTTCG(-3)
TCTCTGTGCCAGCAGC(-5)	CCTC	(-6)GGGGGGG(-1)	A	AACTATGCTGAGCAGTTCTTCG(-1)
TCTCTGTGCCAGCAGC(-5)	CGTGAC	GGG		ATGCTGAGCAGTTCTTCG(-5)
TCTCTGTGCCAGCAGC(-5)	CCC	(-2)GACAGGG(-3)	**A**	TAACTATGCTGAGCAGTTCTTCG
TCTCTGTGCCAGCAGC(-5)	C**CCC**	GGGACT(-8)	TG	AACTATGCTGAGCAGTTCTTCG(-1)
TCTCTGTGCCAGCAG(-6)	TTT	GGGAC(-7)	CGAAA	TATGCTGAGCAGTTCTTCG(-4)
TCTCTGTGCCAGCAG(-6)	GT**CC**	GGGACAGGGGG	GG	ATGCTGAGCAGTTCTTCG(-5)
TCTCTGTGCCAGCAG(-6)	GAA	(-2)GACAGG(-4)	TT	ACTATGCTGAGCAGTTCTTCG(-2)
TCTCTGTGCCAGC(-8)	**CC**	GGGACTGGGG(-4)		CTATGCTGAGCAGTTCTTCG(-3)
TCTCTGTGCCAGC(-8)		(-5)TGGGGGG(-2)	T	CTATGCTGAGCAGTTCTTCG(-3)

C

Vβ14		Dβ1		Jβ1.1
CCTCTGTGCCTGGAGTCT		GGGACAGGGGGC		CAAACACAGAAGTCTTCTTTGG
WT				
CCTCTGTGCCTGGAGTCT	**AG**TG	CAGG		CAAACACAGAAGTCTTCTTTGG
CCTCTGTGCCTGGAGTCT	**AG**	GGGACAGGGGG	T	CAAACACAGAAGTCTTCTTTGG
CCTCTGTGCCTGGAGTCT	G	AGGG	**AAG**	CAAACACAGAAGTCTTCTTTGG
CCTCTGTGCCTGGAGTCT		GG	TTCCTTTAC	CAAACACAGAAGTCTTCTTTGG
CCTCTGTGCCTGGAGTCT	**A**TT	GGGG		AACACAGAAGTCTTCTTTGG(−2)
CCTCTGTGCCTGGAGTCT	**A**	GACA	ATA	AGAAGTCTTCTTTGG(−7)
CCTCTGTGCCTGGAGTCT	**AGA**	GC	T	GTCTTCTTTGG(−11)
CCTCTGTGCCTGGAGTC(−1)	CAG**ACC**	GGGACAG	C	AACACAGAAGTCTTCTTTGG(−2)
CCTCTGTGCCTGGAGT(−2)	TCA	GGGACAGGGGGC	GGAGG	AGTCTTCTTTGG(−10)
CCTCTGTGCCTGGA(−4)		ACAGGGG		AAACACAGAAGTCTTCTTTGG(−1)
CCTCTGTGCCT(−7)	**C**	GGGACAGGG	AAGG	ACAGAAGTCTTCTTTGG(−5)
RAG2$^{FS/FS}$				
CCTCTGTGCCTGGAGTCT	**AG**GC	AC	TA	CAAACACAGAAGTCTTCTTTGG
CCTCTGTGCCTGGAGTCT	**AG**	AGGGG	**TC**	CAAACACAGAAGTCTTCTTTGG
CCTCTGTGCCTGGAGTCT	G	AGGG	**AAG**	CAAACACAGAAGTCTTCTTTGG
CCTCTGTGCCTGGAGTCT	**AA**T	AGGG		AAACACAGAAGTCTTCTTTGG(−1)
CCTCTGTGCCTGGAGTCT		AGGGGG	G	AACACAGAAGTCTTCTTTGG(−2)
CCTCTGTGCCTGGAGTCT	**A**CC	GACAG	T	CAGAAGTCTTCTTTGG(−6)
CCTCTGTGCCTGGAGTCT	**A**	GACAGGG	AC	CAGAAGTCTTCTTTGG(−6)
CCTCTGTGCCTGGAGTCT	TTA	ACAGGGGGC	AA	AAGTCTTCTTTGG(−9)
CCTCTGTGCCTGGAGTC(−1)	GAAGGG	AGGGG	T	CACAGAAGTCTTCTTTGG(−4)
CCTCTGTGCCTGGAGTC(−1)		GACAG	ATGT	CAGAAGTCTTCTTTGG(−6)
CCTCTGTGCCTGGAGT(−2)	GT	AG	TAAC	CACAGAAGTCTTCTTTGG(−4)
CCTCTGTGCCTGGAG(−3)		CAGGGGG	**G**	CAAACACAGAAGTCTTCTTTGG
CCTCTGTGCCTGG(−5)	GAT	ACAGG	TTT	AGAAGTCTTCTTTGG(−7)
CCTCTGTGCC(−8)	CT	CAGGG	G**TG**	CAAACACAGAAGTCTTCTTTGG

D

Dβ2

tttttgtatcacgatgtaacattgtgGGGACTGGGGGGGC CTCCTATGAACAGTACTTCGGTCCCGG Jβ2.6
AACCAAGACACCCAGTACTTTGGGCCA Jβ2.5
AGTCAAAACACCTTGTACTTTGGTGCG Jβ2.4

WT		
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGGC	**GC**TTGGGGC	CCAAGACACCCAGTACTTTGGGCCA(−2)
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGGC	**G**ATGGG	CAAGACACCCAGTACTTTGGGCCA(−3)
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGGC	**GC**TT	AGACACCCAGTACTTTGGGCCA(−5)
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGGC	**GC**T	TGAACAGTACTTCGGTCCCGG(−6)
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGGC	CC	CACCCAGTACTTTGGGCCA(−8)
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGG(−1)	T	TCCCGG(−21)
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGG(−1)	G	CTATGAACAGTACTTCGGTCCCGG(−3)
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGG(−1)	GT	TATGAACAGTACTTCGGTCCCGG(−4)
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGG(−1)	GG	CCAAGACACCCAGTACTTTGGGCCA(−2)
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGG(−1)		AAGACACCCAGTACTTTGGGCCA(−4)
tttttgtatcacgatgtaacattgtgGGGACTGGGGGG(−2)	T	TGAACAGTACTTCGGTCCCGG(−6)
tttttgtatcacgatgtaacattgtgGGGACTGGGGGG(−2)	AGGCG	GACACCCAGTACTTTGGGCCA(−6)
tttttgtatcacgatgtaacattgtgGGGACTGGGGG(−3)		CCAAGACACCCAGTACTTTGGGCCA(−2)
tttttgtatcacgatgtaacattgtgGGGACTGGG(−5)	CC**TT**	AACCAAGACACCCAGTACTTTGGGCCA
tttttgtatcacgatgtaacattgtgGGGACTGGG(−5)	**TT**	AACCAAGACACCCAGTACTTTGGGCCA
RAG2$^{FS/FS}$		
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGGC	ACGAG	CCAAGACACCCAGTACTTTGGGCCA(−2)
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGGC	**GC**AGG	CCAAGACACCCAGTACTTTGGGCCA(−2)
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGGC	CGC	CCAAGACACCCAGTACTTTGGGCCA(−2)
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGGC		TCCTATGAACAGTACTTCGGTCCCGG(−1)
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGGC	CT**T**	AGTCAAAACACCTTGTACTTTGGTGCG
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGGC		CCAAGACACCCAGTACTTTGGGCCA(−2)
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGGC		CCAAGACACCCAGTACTTTGGGCCA(−2)
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGGC		GACACCCAGTACTTTGGGCCA(−6)
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGG(−1)	AA**ACT**	AGTCAAAACACCTTGTACTTTGGTGCG
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGG(−1)	G	ACACCTTGTACTTTGGTGCG(−7)
tttttgtatcacgatgtaacattgtgGGGACTGGGGGGG(−1)		AGACACCCAGTACTTTGGGCCA(−5)
tttttgtatcacgatgtaacattgtgGGGACTGGGGGG(−2)	TG	AACACCTTGTACTTTGGTGCG(−6)
tttttgtatcacgatgtaacattgtgGGGACTGGGGG(−3)	AT	TATGAACTTCGGTCCCGG(−4)
tttttgtatcacgatgtaacattgtgGGGACTGGGGG(−3)		TTTGGGCCA(−18)
tttttgtatcacgatgtaacattgtgGGGACTGGGG(−4)		AACCAAGACACCCAGTACTTTGGGCCA
tttttgtatcacgatgtaacattgtgGGGACTGGG(−5)	AG	GACACCCAGTACTTTGGGCCA(−6)
tttttgtatcacgatgtaacattgtgGGGACTGG(−6)	TGGATG	AAAACACCTTGTACTTTGGTGCG(−4)

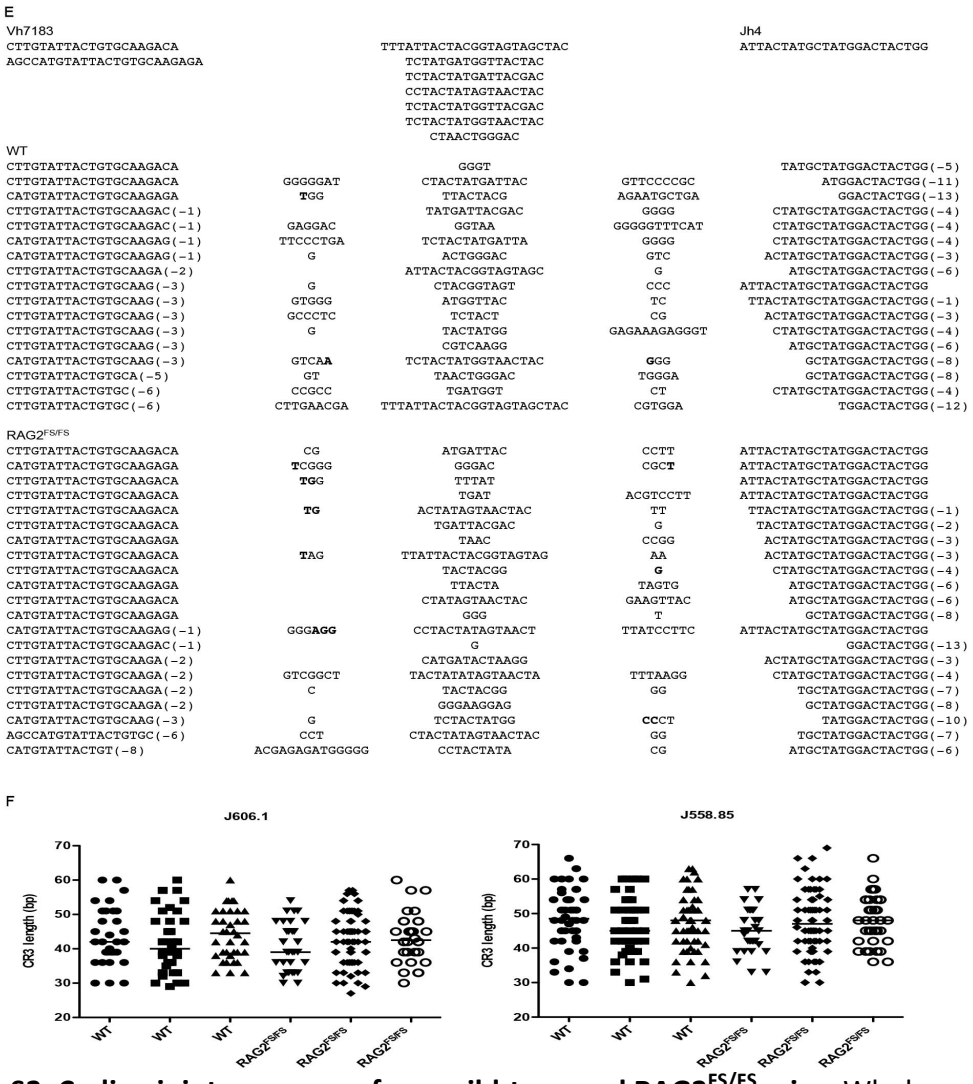


Figure S2. Coding joint sequences from wild-type and RAG2$^{FS/FS}$ mice. Whole thymocytes or BM cells from 2-3 mice were analyzed by PCR amplification of the indicated loci followed by Topo cloning. Germ line sequence is indicated at the top of each locus. Capital letters in the middle of the junction indicate N nt, capital bold are P nt, deletions are indicated in parentheses, small letter indicates sequences from the coding region (miscleavage). A. Vβ6/7/8-Jβ2 B. Vβ10-Jβ2.1 C. Vβ14-J1.1. Blue and red represents the ADRs D. Dβ2-Jβ2 (bold small letters represent the 12 RSS) E. Vh1783-Jh4. F.VH CDR3 length distribution analyzed by CDR3 spectratyping on spleen DNA (54). CDR3 length (bp) is plotted for wild-type and RAG2$^{FS/FS}$ mice. Each symbol represents a single rearrangement of J606.1-JH2 (left panel) or J558.82-JH2 (right panel). Each column is an individual mouse of a given genotype (n=3 per genotype). Horizontal bars indicate the median CDR3 length for all of the rearrangements sampled in a single mouse.

BIBLIOGRAPHY

1. Mitelman,F., Johansson,B. and Mertens,F. (2007) The impact of translocations and gene fusions on cancer causation. Nat. Rev. Cancer., 7, 233-245.

2. Inaki,K. and Liu,E.T. (2012) Structural mutations in cancer: Mechanistic and functional insights. Trends Genet., 28, 550-559.

3. Johnson RD, Jasin M (2000) Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. EMBO J. 3;19(13):3398-407.

4. Symington LS, Gautier J (2011) Double-strand break end resection and repair pathway choice. Annu Rev Genet. 45:247-71

5. Sartori AA, Lukas C, Coates J, Mistrik M, Fu S, Bartek J, Baer R, Lukas J, Jackson SP (2007) Human CtIP promotes DNA end resection. Nature. 22;450(7169):509-14.

6. Jensen RB, Carreira A, Kowalczykowski SC (2010) Purified human BRCA2 stimulates RAD51-mediated recombination. Nature. 467(7316):678-83.

7. Moynahan ME, Jasin M (2010) Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. Nat Rev Mol Cell Biol. (3):196-207.

8. Lieber,M.R. (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu. Rev. Biochem., 79, 181-211.

9. Zhu,C., Mills,K.D., Ferguson,D.O., Lee,C., Manis,J., Fleming,J., Gao,Y., Morton,C.C. and Alt,F.W. (2002) Unrepaired DNA breaks in p53-deficient cells lead to oncogenic gene amplification subsequent to translocations. Cell, 109, 811-821

10. Difilippantonio,M.J., Zhu,J., Chen,H.T., Meffre,E., Nussenzweig,M.C., Max,E.E., Ried,T. and Nussenzweig,A. (2000) DNA repair protein Ku80 suppresses chromosomal aberrations and malignant transformation. Nature, 404, 510-514.

11. Wang JH, Alt FW, Gostissa M, Datta A, Murphy M, Alimzhanov MB, Coakley KM, Rajewsky K, Manis JP, Yan CT (2008) Oncogenic transformation in the absence of Xrcc4 targets peripheral B cells that have undergone editing and switching. J Exp Med. 205(13):3079-90

12. Difilippantonio MJ, Petersen S, Chen HT, Johnson R, Jasin M, Kanaar R, Ried T, Nussenzweig A (2002) Evidence for replicative repair of DNA double-strand breaks leading to oncogenic translocation and gene amplification. J Exp Med.196(4):469-80.

13. Mimori T, Hardin JA (1986) Mechanism of interaction between Ku protein and DNA. J Biol Chem. 261(22):10375-9.

14. Blier PR, Griffith AJ, Craft J, Hardin JA (1993) Binding of Ku protein to DNA. Measurement of affinity for ends and demonstration of binding to nicks. J Biol Chem. 268(10):7594-601.

15. Ma Y, Lu H, Tippin B, Goodman MF, Shimazaki N, Koiwai O, Hsieh CL, Schwarz K, Lieber MR (2004) A biochemically defined system for mammalian nonhomologous DNA end joining. Mol Cell. 16(5):701-13.

16. Lieber MR (2008) The mechanism of human nonhomologous DNA end joining. J Biol Chem. 283(1):1-5

17. Gell D, Jackson SP (1999) Mapping of protein-protein interactions within the DNA-dependent protein kinase complex. Nucleic Acids Res. 27(17):3494-502.

18. Yannone SM, Khan IS, Zhou RZ, Zhou T, Valerie K, Povirk LF (2008) Coordinate 5' and 3' endonucleolytic trimming of terminally blocked blunt DNA double-strand break ends by Artemis nuclease and DNA-dependent protein kinase. Nucleic Acids Res. 36(10):3354-65.

19. Ma YM, Pannicke U, Schwarz K, Lieber MR. (2002) Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. Cell 108:781–94

20. Moon AF, Garcia-Diaz M, Batra VK, Beard WA, Bebenek K, Kunkel TA, Wilson SH, Pedersen LC (2007) The X family portrait: structural insights into biological functions of X family polymerases. DNA Repair (Amst). 6(12):1709-25.

21. Teo SH, Jackson SP (1997) Identification of Saccharomyces cerevisiae DNA ligase IV: involvement in DNA double-strand break repair. EMBO J. 16(15):4788-95.

22. Yano K, Morotomi-Yano K, Akiyama H (2009) Cernunnos/XLF: a new player in DNA double-strand break repair. Int J Biochem Cell Biol.;41(6):1237-40.

23. Li Z, Otevrel T, Gao Y, Cheng HL, Seed B, Stamato TD, Taccioli GE, Alt FW (1995) The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. Cell. 83(7):1079-89.

24. Grawunder U, Wilm M, Wu X, Kulesza P, Wilson TE, Mann M, Lieber MR (1997) Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. Nature. 388(6641):492-5.

25. Touvrey C, Couedel C, Soulas P, Couderc R, Jasin M, de Villartay JP, Marche PN, Jouvin Marche E, Candéias SM (2008) Distinct effects of DNA-PKcs and Artemis inactivation on signal joint formation in vivo. Mol Immunol. 45(12):3383-91.

26. Gu J, Lu H, Tippin B, Shimazaki N, Goodman MF, Lieber MR (2007) XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. EMBO J. 26(4):1010-23.

27. Gu J, Lu H, Tsai AG, Schwarz K, Lieber MR (2007) Single-stranded DNA ligation and XLF-stimulated incompatible DNA end ligation by the XRCC4-DNA ligase IV complex: influence of terminal DNA sequence. Nucleic Acids Res. 35(17):5755-62.

28. Zhang,L., Reynolds,T.L., Shan,X. and Desiderio,S. (2011) Coupling of V(D)J recombination to the cell cycle suppresses genomic instability and lymphoid tumorigenesis. Immunity, 34, 163-174.

29. Barlow C, Hirotsune S, Paylor R, Liyanage M, Eckhaus M, Collins F, Shiloh Y, Crawley JN, Ried T, Tagle D, Wynshaw-Boris A (1996) Atm-deficient mice: a paradigm of ataxia telangiectasia. Cell. 86(1):159-71.

30. Petiniot LK, Weaver Z, Vacchio M, Shen R, Wangsa D, Barlow C, Eckhaus M, Steinberg SM, Wynshaw-Boris A, Ried T, Hodes RJ (2002) RAG-mediated V(D)J recombination is not essential for tumorigenesis in Atm-deficient mice.. Mol Cell Biol. 22(9):3174-7.

31. Kabotyanski,E.B., Gomelsky,L., Han,J.O., Stamato,T.D. and Roth,D.B. (1998) Double-strand break repair in Ku86- and XRCC4-deficient cells. Nucleic Acids Res., 26, 5333-5342.

32. Perrault,R., Wang,H., Wang,M., Rosidi,B. and Iliakis,G. (2004) Backup pathways of NHEJ are suppressed by DNA-PK. J. Cell. Biochem., 92, 781-794.

33. Liang,F., Romanienko,P.J., Weaver,D.T., Jeggo,P.A. and Jasin,M. (1996) Chromosomal double-strand break repair in Ku80-deficient cells. Proc. Natl. Acad. Sci. U. S. A., 93, 8929-8933.

34. Liang F, Jasin M (1996) Ku80-deficient cells exhibit excess degradation of extrachromosomal DNA. J Biol Chem. 271(24):14405-11.

35. Boboila,C., Alt,F.W. and Schwer,B. (2012) Classical and alternative end-joining pathways for repair of lymphocyte-specific and general DNA double-strand breaks. Adv. Immunol., 116, 1-49.

36. Corneo,B., Wendland,R.L., Deriano,L., Cui,X., Klein,I.A., Wong,S.Y., Arnal,S., Holub,A.J., Weller,G.R., Pancake,B.A., et al. (2007) Rag mutations reveal robust alternative end joining. Nature, 449, 483-486.

37. Zhang,Y. and Jasin,M. (2011) An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. Nat. Struct. Mol. Biol., 18, 80-84.

38. Boboila C, Oksenych V, Gostissa M, Wang JH, Zha S, Zhang Y, Chai H, Lee CS, Jankovic M, Saez LM, Nussenzweig MC, McKinnon PJ, Alt FW, Schwer B. (2012) Robust chromosomal DNA repair via alternative end-joining in the absence of X-ray repair cross-complementing protein 1 (XRCC1). Proc Natl Acad Sci U S A. Feb 14;109(7):2473-8.

39. Audebert M, Salles B, Calsou P (2004) Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. J Biol Chem 279: 55117–55126.

40. Wang M, Wu W, Wu W, Rosidi B, Zhang L, et al. (2006) PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. Nucleic Acids Res 34: 6170–6182.

41. Audebert M, Salles B, Weinfeld M, Calsou P (2006) Involvement of polynucleotide kinase in a poly(ADP-ribose) polymerase-1-dependent DNA double-strand breaks rejoining pathway. J Mol Biol 356: 257–265.

42.Robert I, Dantzer F, Reina-San-Martin B (2009) Parp1 facilitates alternative NHEJ, whereas Parp2 suppresses IgH/c-myc translocations during immunoglobulin class switch recombination. J Exp Med 206: 1047–1056.

43. Rass E, Grabarz A, Plo I, Gautier J, Bertrand P, et al. (2009) Role of Mre11 in chromosomal nonhomologous end joining in mammalian cells. Nat Struct Mol Biol 16: 819–824.

44. Bennardo N, Cheng A, Huang N, Stark JM (2008) Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. PLoS Genet 4: e1000110.

45. Wang H, Rosidi B, Perrault R, Wang M, Zhang L, et al. (2005) DNA ligase III as a candidate component of backup pathways of nonhomologous end joining. Cancer Res 65: 4020–4030.

46. Simsek,D., Brunet,E., Wong,S.Y., Katyal,S., Gao,Y., McKinnon,P.J., Lou,J., Zhang,L., Li,J., Rebar,E.J., et al. (2011) DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. PLoS Genet., 7, e1002080.

47. Spagnolo L, Barbeau J, Curtin NJ, Morris EP, Pearl LH (2012) Visualization of a DNA-PK/PARP1 complex. Nucleic Acids Res. 40(9):4168-77

48. Simsek D, Jasin M (2010) Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. Nat Struct Mol Biol. 17(4):410-6

49. Guirouilh-Barbat,J., Huck,S., Bertrand,P., Pirzio,L., Desmaze,C., Sabatier,L. and Lopez,B.S. (2004) Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells. Mol. Cell, 14, 611-623.

50. Simbulan-Rosenthal CM, Haddad BR, Rosenthal DS, Weaver Z, Coleman A, Luo R et al (1999) Chromosomal aberrations in PARP(–/–) mice: genome stabilization in immortalized cells by reintroduction of poly(ADP-ribose) polymerase cDNA. Proc Natl Acad Sci USA 96: 13191–13196.

51. Shibata A, Kamada N, Masumura K, Nohmi T, Kobayashi S, Teraoka H et al (2005) Parp-1 deficiency causes an increase of deletion mutations and insertions/rearrangements in vivo after treatment with an alkylating agent. Oncogene 24: 1328–1337

52. Bentley,J., L'Hote,C., Platt,F., Hurst,C.D., Lowery,J., Taylor,C., Sak,S.C., Harnden,P., Knowles,M.A. and Kiltie,A.E. (2009) Papillary and muscle invasive bladder tumors with distinct genomic stability profiles have different DNA repair fidelity and KU DNA-binding activities. Genes Chromosomes Cancer, 48, 310-321.

53. Bentley,J., Diggle,C.P., Harnden,P., Knowles,M.A. and Kiltie,A.E. (2004) DNA double strand break repair in human bladder cancer is error prone and involves microhomology-associated end-joining. Nucleic Acids Res., 32, 5249-5259.

54. Tobin,L.A., Robert,C., Nagaria,P., Chumsri,S., Twaddell,W., Ioffe,O.B., Greco,G.E., Brodie,A.H., Tomkinson,A.E. and Rassool,F.V. (2012) Targeting abnormal DNA repair in therapy-resistant breast cancers. Mol. Cancer. Res., 10, 96-107.

55. Cortizas,E.M., Zahn,A., Hajjar,M.E., Patenaude,A.M., Di Noia,J.M. and Verdun,R.E. (2013) Alternative end-joining and classical nonhomologous end-joining pathways repair different types of double-strand breaks during class-switch recombination. J. Immunol., 191, 5751-5763.

56. Roth D, Wilson JH (1985) Relative rates of homologous and nonhomologous recombination in transfected DNA. Proc Natl Acad Sci USA. 82:3355–3359

57. Kim JS, Krasieva TB, Kurumizaka H, Chen DJ, Taylor AM, Yokomori K (2005) Independent and sequential recruitment of NHEJ and HR factors to DNA damage sites in mammalian cells. J Cell Biol. 170(3):341-7.

58. Helmink,B.A., Tubbs,A.T., Dorsett,Y., Bednarski,J.J., Walker,L.M., Feng,Z., Sharma,G.G., McKinnon,P.J., Zhang,J., Bassing,C.H., et al. (2011) H2AX prevents CtIP-mediated DNA end resection and aberrant repair in G1-phase lymphocytes. Nature, 469, 245-249.

59. Escribano-Díaz C, Orthwein A, Fradet-Turcotte A, Xing M, Young JT, Tkáč J, Cook MA, Rosebrock AP, Munro M, Canny MD, Xu D, Durocher D. (2013) A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. Mol Cell. 49(5):872-83.

60. Jiang G, Plo I, Wang T, Rahman M, Cho JH, Yang E, Lopez BS, Xia F (2013) BRCA1-Ku80 protein interaction enhances end-joining fidelity of chromosomal double-strand breaks in the G1 phase of the cell cycle. J Biol Chem. 288(13):8966-76

61. Allen C, Kurimasa A, Brennemann MA, Chen DJ, Nickoloff JA (2002) DNA-dependent protein kinase suppresses double-strand break-induced and spontaneous homologous recombination. Proc Natl Acad Sci USA. 19;99(6):3758-63

62. Pierce AJ, Hu P, Han M, Ellis N, Jasin M (2001) Ku DNA end-binding protein modulates homologous repair of double-strand breaks in mammalian cells. Genes Dev. 15;15(24):3237-42.

63. Young Eun Choi, Yunfeng Pan, Eunmi Park, Panagiotis Konstantinopoulos, Subhajyoti De, Alan D'Andrea, Dipanjan Chowdhury (2014) MicroRNAs down-regulate homologous recombination in the G1 phase of cycling cells to maintain genomic stability. eLife 2014;10.7554/eLife.02445

64. Chen BP, Chan DW, Kobayashi J, et al. (2005) Cell cycle dependence of DNA-dependent protein kinase phosphorylation in response to DNA double strand breaks. J Biol Chem 280:14709–14715

65. Bothmer A, Robbiani DF, Feldhahn N, Gazumyan A, Nussenzweig A, Nussenzweig MC (2010) 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. J Exp Med. 207(4):855-65

66. Dimitrova N, Chen YC, Spector DL, de Lange T (2008) 53BP1 promotes non-homologous end joining of telomeres by increasing chromatin mobility. Nature. 456(7221):524-8

67. Teng X, Dayhoff-Brannigan M, Cheng WC, Gilbert CE, Sing CN, Diny NL, Wheelan SJ, Dunham MJ, Boeke JD, Pineda FJ, Hardwick JM (2013) Genome-wide consequences of deleting any single gene. Mol Cell. 52(4):485-94

68. Deriano,L. and Roth,D.B. (2013) Modernizing the nonhomologous end-joining repertoire: Alternative and classical NHEJ share the stage. Annu. Rev. Genet., 47, 433-455.

69. Schatz,D.G. and Swanson,P.C. (2011) V(D)J recombination: Mechanisms of initiation. Annu. Rev. Genet., 45, 167-202.

70. Curry JD, Geier JK, Schlissel MS (2005) Single-strand recombination signal sequence nicks in vivo: evidence for a capture model of synapsis. Nat. Immunol. 6:1272–79

71. Jones JM, Gellert M (2002) Ordered assembly of the V(D)J synaptic complex ensures accurate recombination. EMBO J. 21:4162–71

72. Mundy CL, Patenge N, Matthews AGW, Oettinger MA (2002) Assembly of the RAG1/RAG2 synaptic complex. Mol. Cell. Biol. 22:69–77

73. McBlane JF, van Gent DC, Ramsden DA, Romeo C, Cuomo CA, et al. (1995) Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. Cell 83:387–95

74. Coussens,M.A., Wendland,R.L., Deriano,L., Lindsay,C.R., Arnal,S.M. and Roth,D.B. (2013) RAG2's acidic hinge restricts repair-pathway choice and promotes genomic stability. Cell. Rep., 4, 870-878.

75. S.M. Lewis (1994) P nucleotides, hairpin DNA and V(D)J joining: making the connection. Semin. Immunol., 6 pp. 131–141

76. Landau NR, Schatz DG, Rosa M, Baltimore D (1987) Increased frequency of N-region insertion in a murine pre-B-cell line infected with a terminal deoxynucleotidyl transferase retroviral expression vector. Mol Cell Biol. 7(9):3237-43.

77. Ramsden D A, Gellert M (1995) Formation and resolution of double-strand break intermediates in V(D)J rearrangement. Genes Dev. 9(19):2409-20.

78.Roth D B, Nakajima P B, Menetski J P, Bosma M J, Gellert M (1992) Double-strand breaks associated with V(D)J recombination at the TCR delta locus in murine thymocytes. Curr Top Microbiol Immunol. 182:115-24.

79. Schlissel M, Constantinescu A, Morrow T, Baxter M, Peng A (1993) Double-strand signal sequence breaks in V(D)J recombination are blunt, 5'-phosphorylated, RAG-dependent, and cell cycle regulated. Genes Dev. 7(12B):2520-32.

80. Lewis S, Gifford A, Baltimore D (1985) DNA elements are asymmetrically joined during the site-specific recombination of kappa immunoglobulin genes. Science. 228(4700):677-85.

81. Neiditch MB, Lee GS, Huye LE, Brandt VL, Roth DB (2002) The V(D)J recombinase efficiently cleaves and transposes signal joints. Mol Cell. 9(4):871-8.

82. Vanura K, Montpellier B, Le T, Spicuglia S, Navarro JM, Cabaud O, Roulland S, Vachez E, Prinz I, Ferrier P, Marculescu R, Jäger U, Nadel B (2007) In vivo reinsertion of excised episomes by the V(D)J recombinase: a potential threat to genomic stability. PLoS Biol. 5(3):e43.

83. Lewis, S., Hesse, J.E., Mizuuchi, K., and Gellert, M. (1988). Novel strand exchanges in V(D)J recombination. Cell 55, 1099-1107.

84. Deriano L, Chaumeil J, Coussens M, Multani A, Chou Y, Alekseyenko AV, Chang S, Skok JA, Roth DB (2011) The RAG2 C terminus suppresses genomic instability and lymphomagenesis. Nature. 471(7336):119-23

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

86. R.M. Cobb, K.J. Oestreich, O.A. Osipovich, E.M. Oltz (2006) Accessibility control of V(D)J recombination. Adv Immunol, 91 pp. 45–109

87. Sleckman BP, Gorman JR, Alt FW (1996) Accessibility control of antigen-receptor variable-region gene assembly: role of cis-acting elements. Annu Rev Immunol. 14:459-81. Review.

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

89. A.G. Matthews, A.J. Kuo, S. Ramon-Maiques, S. Han, K.S. Champagne, D. Ivanov et al. (2007) RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. Nature, 450 pp. 1106–111

90. Liu Y, Subrahmanyam R, Chakraborty T, Sen R, Desiderio S (2007) A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. Immunity (4):561-71.

91. Kirsch IR, Morton CC, Nakahara K, Leder P (1982) Human immunoglobulin heavy chain genes map to a region of translocations in malignant B lymphocytes. Science. 216(4543):301-3

92. Taub R, Kirsch I, Morton C, Lenoir G, Swan D, Tronick S, Aaronson S, Leder P (1982) Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells. Proc Natl Acad Sci U S A. 79(24):7837-41.

93. Tsuji,H., Ishii-Ohba,H., Katsube,T., Ukai,H., Aizawa,S., Doi,M., Hioki,K. and Ogiu,T (2004) Involvement of illegitimate V(D)J recombination or microhomology-mediated nonhomologous end-joining in the formation of intragenic deletions of the Notch1 gene in mouse thymic lymphomas. Cancer Res., 64, 8882-8890.

94. Sakata,J., Inoue,J., Ohi,H., Kosugi-Okano,H., Mishima,Y., Hatakeyama,K., Niwa,O. and Kominami,R. (2004) Involvement of V(D)J recombinase in the generation of intragenic deletions in the Rit1/Bcl11b tumor suppressor gene in gamma-ray-induced thymic lymphomas and in normal thymus of the mouse. Carcinogenesis, 25, 1069-1075.

95. Marculescu,R., Vanura,K., Montpellier,B., Roulland,S., Le,T., Navarro,J.M., Jager,U., McBlane,F. and Nadel,B. (2006) Recombinase, chromosomal translocations and lymphoid neoplasia: Targeting mistakes and repair failures. DNA Repair (Amst), 5, 1246-1258.

96. Lewis,S.M., Agard,E., Suh,S. and Czyzyk,L. (1997) Cryptic signals and the fidelity of V(D)J joining. Mol. Cell. Biol., 17, 3125-3136.

97. Papaemmanuil E, Rapado I, Li Y, Potter NE, Wedge DC, Tubio J, Alexandrov LB, Van Loo P, Cooke SL, Marshall J et al. (2014) RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. Nat Genet. 46, 116-25

98. Messier TL, O'Neill JP, Hou SM, Nicklas JA, Finette BA (2003) In vivo transposition mediated by V(D)J recombinase in human T lymphocytes. EMBO J. 22(6):1381-8.

99. Curry,J.D. and Schlissel,M.S. (2008) RAG2's non-core domain contributes to the ordered regulation of V(D)J recombination. Nucleic Acids Res., 36, 5750-5762.

100. Welzel N, Le T, Marculescu R, Mitterbauer G, Chott A, Pott C, Kneba M, Du MQ, Kusec R, Drach J, Raderer M, Mannhalter C, Lechner K, Nadel B, Jaeger U (2001) Templated nucleotide addition and immunoglobulin JH-gene utilization in t(11;14) junctions: implications for the mechanism of translocation and the origin of mantle cell lymphoma. Cancer Res. 15;61(4):1629-36.

101. Falk,I., Nerz,G., Haidl,I., Krotkova,A. and Eichmann,K. (2001) Immature thymocytes that fail to express TCRbeta and/or TCRgamma delta proteins die by apoptotic cell death in the CD44(-)CD25(-) (DN4) subset. Eur. J. Immunol., 31, 3308-3317.

102. Jäger U, Böcskör S, Le T, Mitterbauer G, Bolz I, Chott A, Kneba M, Mannhalter C, Nadel B (2000) Follicular lymphomas' BCL-2/IgH junctions contain templated nucleotide insertions: novel insights into the mechanism of t(14;18) translocation. Blood. 95(11):3520-9.

103. Frank KM, Sekiguchi JM, Seidl KJ, Swat W, Rathbun GA, Cheng HL, Davidson L, Kangaloo L, Alt FW (1998) Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV. Nature. 1998 396(6707):173-7.

104. Gao Y, Sun Y, Frank KM, Dikkes P, Fujiwara Y, Seidl KJ, Sekiguchi JM, Rathbun GA, Swat W, Wang J, Bronson RT, Malynn BA, Bryans M, Zhu C, Chaudhuri J, Davidson L, Ferrini R, Stamato T, Orkin SH, Greenberg ME, Alt FW (1998) A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. Cell. 95(7):891-902.

105. Bosma GC, Custer RP, Bosma MJ (1983) A severe combined immunodeficiency mutation in the mouse. Nature. 301(5900):527-30.

106. Rooney S, Sekiguchi J, Zhu C, Cheng HL, Manis J, Whitlow S, DeVido J, Foy D, Chaudhuri J, Lombard D, Alt FW (2002) Leaky Scid phenotype associated with defective V(D)J coding end processing in Artemis-deficient mice.. Mol Cell. (6):1379-90.

107. Bogue,MA., Wang,C., Zhu,C. and Roth,DB. (1997) V(D)J recombination in Ku86-deficient mice: distinct effects on coding, signal, and hybrid joint formation. Immunity, 1, 37-47.

108. Nussenzweig,A., Chen,C., da Costa Soares,V., Sanchez,M., Sokol,K., Nussenzweig,M.C. and Li,G.C. (1996) Requirement for Ku80 in growild-typeh and immunoglobulin V(D)J recombination. Nature, 382, 551-555.

108. Jones,J.M. and Simkus,C. (2009) The roles of the RAG1 and RAG2 "non-core" regions in V(D)J recombination and lymphocyte development. Arch. Immunol. Ther. Exp. (Warsz), 57, 105-116.

110. Talukder,S.R., Dudley,D.D., Alt,F.W., Takahama,Y. and Akamatsu,Y. (2004) Increased frequency of aberrant V(D)J recombination products in core RAG-expressing mice. Nucleic Acids Res., 32, 4539-4549.

111. Mansour,W.Y., Rhein,T. and Dahm-Daphi,J. (2010) The alternative end-joining pathway for repair of DNA double-strand breaks requires PARP1 but is not dependent upon microhomologies. Nucleic Acids Res., 38, 6065-6077.

112. Liang,H.E., Hsu,L.Y., Cado,D., Cowell,L.G., Kelsoe,G. and Schlissel,M.S. (2002) The "dispensable" portion of RAG2 is necessary for efficient V-to-DJ rearrangement during B and T cell development. Immunity, 17, 639-651.

113. Mijuskovic,M., Brown,S.M., Tang,Z., Lindsay,C.R., Efstathiadis,E., Deriano,L. and Roth,D.B. (2012) A streamlined method for detecting structural variants in cancer genomes by short read paired-end sequencing. PLoS One, 7, e48314.

114. Sadofsky,M.J., Hesse,J.E. and Gellert,M. (1994) Definition of a core region of RAG-2 that is functional in V(D)J recombination. Nucleic Acids Res., 22, 1805-1809.

115. Cuomo,C.A. and Oettinger,M.A. (1994) Analysis of regions of RAG-2 important for V(D)J recombination. Nucleic Acids Res., 22, 1810-1814.

116. Akamatsu,Y., Monroe,R., Dudley,D.D., Elkin,S.K., Gartner,F., Talukder,S.R., Takahama,Y., Alt,F.W., Bassing,C.H. and Oettinger,M.A. (2003) Deletion of the RAG2 C terminus leads to impaired lymphoid development in mice. Proc. Natl. Acad. Sci. U. S. A., 100, 1209-1214.

117. Schulte-Uentrop,L., El-Awady,R.A., Schliecker,L., Willers,H. and Dahm-Daphi,J. (2008) Distinct roles of XRCC4 and Ku80 in non-homologous end-joining of endonuclease- and ionizing radiation-induced DNA double-strand breaks. Nucleic Acids Res., 36, 2561-2569.

118. Mahajan,K.N., Nick McElhinny,S.A., Mitchell,B.S. and Ramsden,D.A. (2002) Association of DNA polymerase mu (pol mu) with ku and ligase IV: Role for pol mu in end-joining double-strand break repair. Mol. Cell. Biol., 22, 5194-5202.

119. Lebedeva,N.A., Rechkunova,N.I., Dezhurov,S.V., Khodyreva,S.N., Favre,A., Blanco,L. and Lavrik,O.I. (2005) Comparison of functional properties of mammalian DNA polymerase lambda and DNA polymerase beta in reactions of DNA synthesis related to DNA repair. Biochim. Biophys. Acta, 1751, 150-158.

120. Sobol,R.W. and Wilson,S.H. (2001) Mammalian DNA beta-polymerase in base excision repair of alkylation damage. Prog. Nucleic Acid Res. Mol. Biol., 68, 57-74.

121. Clark,JM. (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. Nucleic Acids Res., 16, 9677-86.

122. Purugganan,MM., Shah,S., Kearney,JF. and Roth,DB. (2001) Ku80 is required for addition of N nucleotides to V(D)J recombination junctions by terminal deoxynucleotidyl transferase. Nucleic Acids Res., 29, 1638-46.

123. Boubakour-Azzouz,I., Bertrand,P., Claes,A., Lopez,BS., and Rougeon,F. (2012) Terminal deoxynucleotidyl transferase requires KU80 and XRCC4 to promote N-addition at non-V(D)J chromosomal breaks in non-lymphoid cells. Nucleic Acids Res., 40, 8381-91.

124. Schipler A, Iliakis G. (2013) DNA double-strand-break complexity levels and their possible contributions to the probability for error-prone processing and repair pathway choice. Nucleic Acids Res. 41, 7589-605

125. Boboila,C., Yan,C., Wesemann,D.R., Jankovic,M., Wang,J.H., Manis,J., Nussenzweig,A., Nussenzweig,M. and Alt,F.W. (2010) Alternative end-joining catalyzes class switch recombination in the absence of both Ku70 and DNA ligase 4. J. Exp. Med., 207, 417-427.

126. Yan,C.T., Boboila,C., Souza,E.K., Franco,S., Hickernell,T.R., Murphy,M., Gumaste,S., Geyer,M., Zarrin,A.A., Manis,J.P., et al. (2007) IgH class switching and translocations use a robust non-classical end-joining pathway. Nature, 449, 478-482.

127. Hesse JE, Lieber MR, Mizuuchi K, Gellert M (1989) V(D)J recombination: a functional definition of the joining signals. Genes Dev. 3(7):1053-61.

128. Lee AI, Fugmann SD, Cowell LG, Ptaszek LM, Kelsoe G, Schatz DG (2003) A functional analysis of the spacer of V(D)J recombination signal sequences. PLoS Biol. 1(1):E1.

129. Merelli I, Guffanti A, Fabbri M, Cocito A, Furia L, Grazini U, Bonnal RJ, Milanesi L, McBlane F (2010) RSSsite: a reference database and prediction tool for the identification of cryptic Recombination Signal Sequences in human and murine genomes. Nucleic Acids Res. 38 (Web Server issue):W262-7

130. Hicks,D.G., Janarthanan,B.R., Vardarajan,R., Kulkarni,S.A., Khoury,T., Dim,D., Budd,G.T., Yoder,B.J., Tubbs,R., Schreeder,M.T., et al. (2010) The expression of TRMT2A, a novel cell cycle regulated protein, identifies a subset of breast cancer patients with HER2 over-expression that are at an increased risk of recurrence. BMC Cancer, 10, 108-2407-10-108.

131. So,T. and Croft,M. (2013) Regulation of PI-3-kinase and akt signaling in T lymphocytes and other cells by TNFR family molecules. Front. Immunol., 4, 139.

132. Oflazoglu,E., Grewal,I.S. and Gerber,H. (2009) Targeting CD30/CD30L in oncology and autoimmune and inflammatory diseases. Adv. Exp. Med. Biol., 647, 174-185.

133. Ciccia A, Nimonkar AV, Hu Y, Hajdu I, Achar YJ, Izhar L, Petit SA, Adamson B, Yoon JC, Kowalczykowski SC, Livingston DM, Haracska L, Elledge SJ (2012) Polyubiquitinated PCNA recruits the ZRANB3 translocase to maintain genomic integrity after replication stress. Mol Cell. 47(3):396-409.

134. Kigel B, Rabinowicz N, Varshavsky A, Kessler O, Neufeld G (2011) Plexin-A4 promotes tumor progression and tumor angiogenesis by enhancement of VEGF and bFGF signaling. Blood. 118(15):4285-96.

135. Grundy GJ, Yang W, Gellert M. Autoinhibition of DNA cleavage mediated by RAG1 and RAG2 is overcome by an epigenetic signal in V(D)J recombination. Proc Natl Acad Sci U S A. 107(52):22487-92

136. Ramsden DA, McBlane JF, van Gent DC, Gellert M. (1997) Distinct DNA sequence and structure requirements for the two steps of V(D)J recombination signal cleavage. EMBO J. 15(12):3197-206.

137. Marculescu R, Le T, Simon P, Jaeger U, Nadel B (2002) V(D)J-mediated translocations in lymphoid neoplasms: a functional assessment of genomic instability by cryptic sites. J Exp Med. 195(1):85-98.

138. Cheng JT, Yang CY, Hernandez J, Embrey J, Baer R (1990) The chromosome translocation (11;14)(p13;q11) associated with T cell acute leukemia. Asymmetric diversification of the translocational junctions..J Exp Med. 171(2):489-501.

139. Tycko B, Reynolds TC, Smith SD, Sklar J (1989) Consistent breakage between consensus recombinase heptamers of chromosome 9 DNA in a recurrent chromosomal translocation of human T cell leukemia. J Exp Med. 169(2):369-77.

140. Shimazaki N, Askary A, Swanson PC, Lieber MR (2012) Mechanistic basis for RAG discrimination between recombination sites and the off-target sites of human lymphomas. Mol Cell Biol. 32(2):365-75

141. Okitsu CY, Hsieh CL (2007) DNA methylation dictates histone H3K4 methylation. Mol Cell Biol. 7:2746-57

142. Okitsu CY, Hsieh JC, Hsieh CL (2010) Transcriptional activity affects the H3K4me3 level and distribution in the coding region. Mol Cell Biol. 12:2933-46.

143. Swanson PC, Kumar S, Raval P (2009) Early steps of V(D)J rearrangement: insights from biochemical studies of RAG-RSS complexes. Adv Exp Med Biol. 650:1-15. Review.

144. Swanson PC, Desiderio S (1999) RAG-2 promotes heptamer occupancy by RAG-1 in the assembly of a V(D)J initiation complex. Mol Cell Biol. 19(5):3674-83.

145. Pannunzio NR, Li S, Watanabe G, Lieber MR (2014) Non-homologous end joining often uses microhomology: Implications for alternative end joining.. DNA Repair (Amst) 17:74-80.

146. De Lorenzo SB, Patel AG, Hurley RM, Kaufmann SH (2013) The Elephant and the Blind Men: Making Sense of PARP Inhibitors in Homologous Recombination Deficient Tumor Cells. Front Oncol. 3:228. Review.

147. Patel AG, Sarkaria JN, Kaufmann SH (2011) Nonhomologous end joining drives poly(ADP-ribose) polymerase (PARP) inhibitor lethality in homologous recombination-deficient cells. Proc Natl Acad Sci U S A. 108(8):3406-11.

148. Balmaña J, Domchek SM, Tutt A, Garber JE. (2011) Stumbling blocks on the path to personalized medicine in breast cancer: the case of PARP inhibitors for BRCA1/2-associated cancers. Cancer Discov. 1:29-34.

149. Iiizumi S, Kurosawa A, So S, Ishii Y, Chikaraishi Y, Ishii A, Koyama H, Adachi N (2008) Impact of non-homologous end-joining deficiency on random and targeted DNA integration: implications for gene targeting. Nucleic Acids Res. 19:6333-42.

150. Kurosawa A, Saito S, So S, Hashimoto M, Iwabuchi K, Watabe H, Adachi N (2013) DNA ligase IV and artemis act cooperatively to suppress homologous recombination in human cells: implications for DNA double-strand break repair. PLoS One. 8(8):e72253.

151. Perez EE, Wang J, Miller JC, Jouvenot Y, Kim KA, Liu O, Wang N, Lee G, Bartsevich VV, Lee YL, Guschin DY, Rupniewski I, Waite AJ, Carpenito C, Carroll RG, Orange JS, Urnov FD, Rebar EJ, Ando D, Gregory PD, Riley JL, Holmes MC, June CH (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat Biotechnol. 26(7):808-16

152. Oh S, Harvey A, Zimbric J, Wang Y, Nguyen T, Jackson PJ, Hendrickson EA (2014) DNA ligase III and DNA ligase IV carry out genetically distinct forms of end joining in human somatic cells. DNA Repair (Amst). pii: S1568-7864(14)00137-2

153. Deriano L, Stracker TH, Baker A, Petrini JH, Roth DB (2009) Roles for NBS1 in alternative nonhomologous end-joining of V(D)J recombination intermediates. Mol Cell. 34(1):13-25

154. Raghavan SC, Swanson PC, Ma Y, Lieber MR (2005) Double-strand break formation by the RAG complex at the bcl-2 Mbr and at other non-B DNA structures in vitro. Mol Cell Biol 25:5904–5919.

155. Mussolino C, Alzubi J, Fine EJ, Morbitzer R, Cradick TJ, Lahaye T, Bao G, Cathomen T (2014). TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. Nucleic Acids Res. 42(10):6762-73.