RAG1/2 induces genomic insertions by mobilizing DNA into RAG1/2-independent breaks

RAG1/2 induziert genomische Insertionen durch die Mobilisierung von DNA in RAG1/2-unabhängige Brüche

Von der Fakultät Energie-, Verfahrens- und Biotechnik der Universität Stuttgart
zur Erlangung der Würde eines Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigte Abhandlung

Vorgelegt von
Philipp Christian Rommel
aus Ostfildern

Hauptberichter: Prof. Dr. Monilola Olayioye
Mitberichter: Prof. M.D. Ph.D. Michel C. Nussenzweig, Prof. Dr. Albert Jeltsch und Prof. Dr.-Ing Ralf Takors
Tag der mündlichen Prüfung: 6. April 2017

Institut für Zellbiologie und Immunologie der Universität Stuttgart

2017

Table of content

List of figures and tables

Abbreviations

AID	activation-induced cytidine deaminase
ALL	acute lymphoblastic leukemia
ATM	ataxia-telangiectasia mutated
BCR	B cell receptor
BOSC23	retroviral packaging cell line
bp	base pair
c-myc	myelocytomatosis oncogene
cRSS	cryptic RSS
CSR	class switch recombination
D segment	diversity gene segment (at *Ig* and *Tcr* loci)
DNA	deoxyribonucleic acid
DSB	double-strand break
EDTA	ethylenediaminetetraacetic acid
EGFP	enhanced green fluorescent protein
ERFS	early replication fragile site
FACS	fluorescence-activated cell sorting
FL	follicular lymphoma
H3K4me3	histone H3 trimethylated at lysine-4
HEPES	N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid
IC-Seq	insertion capture sequencing
Ig	immunoglobulin
Igh/*IGH*	immunoglobulin heavy locus (mouse/human)
Igκ/*IGK*	immunoglobulin κ locus (mouse/human)
Igλ/*IGL*	immunoglobulin λ locus (mouse/human)
IL-7	interleukin-7
J segment	joining gene segment (at *Ig* and *Tcr* loci)
kb, Mb	kilobase, megabase
MCL	mantle cell lymphoma

MHC	major histocompatibility complex
mm9, mm10	Mus musculus genome assembly version 9, 10 (UCSC database)
NHEJ	non-homologous end joining
p53	tumor (suppressor) protein 53
PCR	polymerase chain reaction
RAG	recombination-activating gene
RAG1/2	RAG recombinase (complex of RAG1 and RAG2)
RIC	RSS information content
RNA	ribonucleic acid
RPMI	Roswell Park Memorial Institute medium
RSS	recombination signal sequence
S17	murine bone marrow stroma cell line
TCR	T cell receptor
Tcrα/*TRA*	T cell receptor α locus (mouse/human)
Tcrβ/*TRB*	T cell receptor β locus (mouse/human)
Tcrγ/*TRG*	T cell receptor γ locus (mouse/human)
Tcrδ/*TRD*	T cell receptor δ locus (mouse/human)
TC-Seq	translocation capture sequencing
TE buffer	Tris-EDTA buffer
T_H	helper T cell
T_{reg}	regulatory T cell
Tris	tris(hydroxymethyl)aminomethane
V segment	variable gene segment (at *Ig* and *Tcr* loci)
XLF	XRCC4-like factor

Summary

The RAG recombinase (RAG1/2) plays an essential role in adaptive immunity by mediating V(D)J recombination in developing lymphocytes. In contrast, aberrant RAG1/2 activity promotes lymphocyte malignancies by causing chromosomal translocations and DNA deletions at cancer genes. In addition, RAG1/2 can induce aberrant DNA insertions by transposition and trans-V(D)J recombination, but only few putative such events have been documented *in vivo*. Moreover, those observed in cancer display characteristics that are not compatible with either DNA transposition or trans-V(D)J recombination. Hence, how RAG1/2 causes genomic DNA insertions is still largely unknown.

In this study, I use translocation capture sequencing (TC-Seq) and insertion capture sequencing (IC-Seq) to analyze chromosomal rearrangements in primary murine developing B cells. I identify aberrant RAG1/2-dependent DNA deletions at immunoglobulin genes, whose products are re-inserted at DNA breaks generated by the I-SceI endonuclease on a heterologous chromosome. The existence of similar insertions in human cancer indicates that RAG1/2 also mobilizes genomic DNA into independent physiologic breaks *in vivo*. Thus, my findings reveal a novel pathway through which RAG1/2 causes DNA insertions independent of DNA transposition and trans-V(D)J recombination. Importantly, this pathway has the potential to destabilize the lymphocyte genome by causing aberrant signal-end, hybrid-end and coding-end insertions and shares features with reported oncogenic DNA insertions.

Zusammenfassung

Die RAG Rekombinase (RAG1/2) katalysiert die V(D)J-Rekombination in sich entwickelnden Lymphozyten und spielt daher eine essentielle Rolle in der adaptiven Immunität. Im Gegensatz dazu fördert die anomale Aktivität von RAG1/2 bösartige Lymphozytenerkrankungen durch die Verursachung von chromosomalen Translokationen und DNA-Deletionen in Krebsgenen. Darüber hinaus kann RAG1/2 anomale DNA-Insertionen durch Transposition und Trans-V(D)J-Rekombination erzeugen, wobei jedoch nur wenige solcher vermeintlichen Ereignisse *in vivo* dokumentiert worden sind. Zudem weisen die in Krebs beobachteten Insertionen Eigenschaften auf, die weder mit DNA-Transposition noch mit Trans-V(D)J Rekombination kompatibel sind. Daher ist es immer noch weitgehend unbekannt, wie RAG1/2 genomische Insertionen verursacht.

In dieser Studie nutze ich *translocation capture sequencing (TC-Seq)* und *insertion capture sequencing (IC-Seq)* um chromosomale Rearrangements in primären, sich entwickelnden Maus-B-Zellen zu analysieren. Ich identifiziere anomale RAG1/2-abhängige DNA-Deletionen in Immunglobulin-Genen, deren Produkte in DNA-Brüche eingefügt werden, welche von der I-SceI Endonuklease in einem heterologen Chromosom erzeugt wurden. Die Existenz von ähnlichen Insertionen in menschlichen Krebserkrankungen deutet darauf hin, dass RAG1/2 genomische DNA auch in unabhängige, physiologische Brüche *in vivo* mobilisiert. Somit enthüllen meine Ergebnisse einen neuartigen Pfad, über den RAG1/2 DNA-Insertionen unabhängig von DNA-Transposition und Trans-V(D)J Rekombination verursacht. Von großer Bedeutung ist dabei, dass dieser Pfad das Potential zur Destabilisierung des Lymphozyten-Genoms hat, da er anomale *signal-end*, *hybrid-end* und *coding-end* Insertionen erzeugt, und Merkmale aufweist, die auch bei onkogenen DNA-Insertionen beobachtet wurden.

Preface

This study has been accepted for publication in *The Journal of Experimental Medicine* on December 12th 2016 (Figure 1; Rommel PC, Oliveira TY, Nussenzweig MC and Robbiani DF). In addition, it was presented in part during the Keystone Conference "*B Cells at the Intersection of Innate and Adaptive Immunity*" (Stockholm, Sweden; May/June 2016) and the symposium "*Frontiers in DNA Repair*" (Berlin, Germany; September 2016).

Figure 1: Cover image submitted to *The Journal of Experimental Medicine*.

Artistic depiction of RAG1/2-induced genomic insertions at RAG1/2-independent DNA breaks. RAG1/2 is represented by a computer (THE RAG RECOMBINASE) with screen and keyboard. Each key controls a specific activity of RAG1/2. "EXCISE" and "RELEASE" (buttons in red) are pressed which induces genomic DNA insertions (illustrated on the screen): RAG1/2 (y-shaped structures in red, plotted based on crystal structure; (Kim et al., 2015)) excises and releases DNA fragments (short strands in green) from the genome (long strand in green). Subsequently, those mobilized fragments re-integrate into RAG1/2-independent DNA breaks elsewhere in the genome (broken strand in white). Idea by Philipp C. Rommel, design by Thiago Y. Oliveira.

1. Introduction

1.1. Immunoglobulins

The adaptive immune system has the capacity to initiate effective immune responses against a virtually limitless array of pathogens. B lymphocytes (B cells) are an essential part of this defense system. During an immune response, B cells target pathogens with the help of specialized glycoproteins (immunoglobulins, Igs) which bind to “foreign” antigens with high specificity (Murphy, 2012). Igs exist in two forms that mediate distinct effector functions: they are either expressed as membrane-bound B cell receptors (BCRs) to facilitate the detection of pathogens or secreted as antibodies for host defense. The entire B cell population contains a vast repertoire of antigen specificities (primary Ig repertoire) with each B cell expressing and secreting Igs of only a single specificity (Murphy, 2012). During an immune response, only cells expressing BCRs with matching antigen specificity are activated, clonally expanded and subsequently differentiated into antibody-secreting plasma cells (clonal selection).

Antibodies are comprised of four polypeptide chains, two identical heavy chains (app. 50 kilodaltons) and two identical light chains (app. 25 kilodaltons), which are connected by disulfide bonds and noncovalent interactions to form a roughly Y-shaped glycoprotein (Figure 2A; (Murphy, 2012)). Each chain is comprised of one variable domain and one or multiple constant domains (Figure 2B; (Murphy, 2012)). The variable domains of heavy and light chains form the variable region which contains two identical antigen-binding sites at the tips of the antibody molecule. These facilitate binding of the antibody to specific structures (epitopes) within the antigen and thus determine its specificity. The constant domains of heavy and light chains form the constant region of the antibody. It contains the antibody “stem” which mediates important effector functions for the host defense against pathogens (opsonization, neutralization and activation of the complement system). In humans and most vertebrates (e.g. mice) there are five different heavy chain constant regions which determine the antibody class (isotype): μ (IgM), δ (IgD), γ (IgG), α (IgA) and ε (IgE). Each isotype possesses a distinctive set of effector functions. In addition, there are two types of light chains, λ and κ, which also differ in their constant regions but do not display any functional difference.

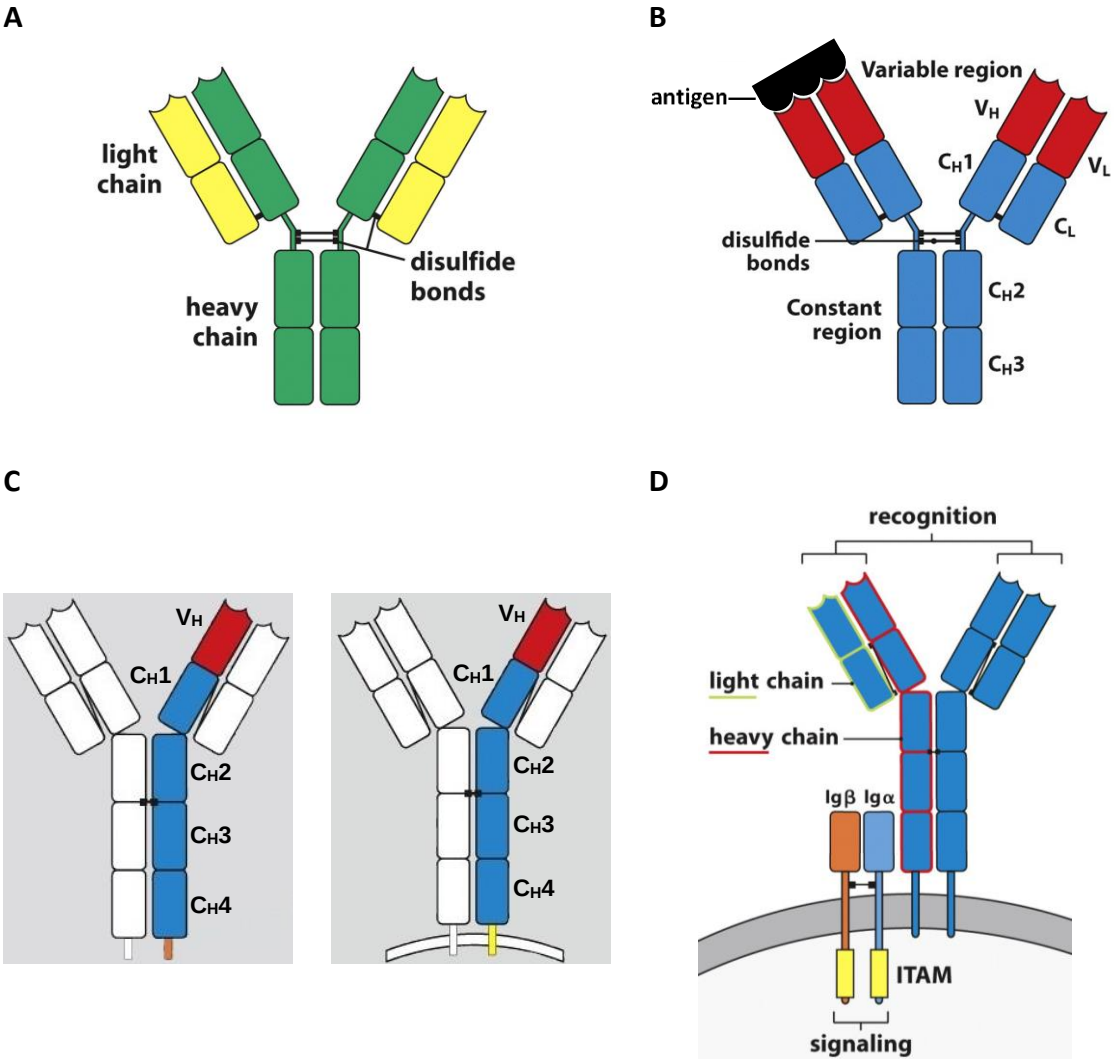

Figure 2: Structure of antibodies and BCRs.

A- Heavy chains (green) and light chains (yellow) of an IgG antibody with their corresponding disulfide bonds.

B- Variable region (red) and constant region (blue) of an IgG antibody. Each heavy chain consists of one variable domain (V_H) and three constant domains (C_H1, C_H2, C_H3) which are numbered from the amino terminus (top) to the carboxy terminus (bottom). Each light chain contains one variable domain (V_L) and one constant domain (C_L). Pairs of V_H and V_L form antigen-binding sites that bind to specific epitopes within antigens (black). C_H2 and C_H3 form the IgG stem that mediates important effector functions for the host defense against pathogens.

C- Comparison between antibodies and BCRs. Heavy chains of IgM antibodies (left) and IgM BCRs (right) are comprised of four constant domains (C_H1 - C_H4, blue) and one variable domain (V_H, red). In IgM antibodies (left), the heavy chain carboxy termini are hydrophilic secretory tails (one shown in orange). In contrast, those of IgM BCRs (right) form hydrophobic transmembrane anchors (one shown in yellow).

D- The BCR complex. BCR heavy chains (one outlined in red) associate with the signaling protein chains Igα and Igβ (light blue and orange, respectively) that transmit cellular signals through immunoreceptor tyrosine-based activation motifs (ITAMs, yellow) upon antigen binding.

A to D modified from (Murphy, 2012).

For each antibody isotype, there is a corresponding BCR. Both are nearly identical in their structure except for their heavy chain carboxy termini (Figure 2C; (Murphy, 2012)). In antibodies, these regions contain secretory tails whereas in BCRs they form transmembrane anchors. Despite structural similarities, BCRs mediate effector functions that are distinct from those of antibodies. The BCR "stem" associates with signaling proteins to form a BCR complex which transmits signals upon antigen binding and thereby mediates development, survival, and clonal expansion of B cells (Figure 2D; (Murphy, 2012)). All mature B cells co-express IgD and IgM on their cell surfaces. Other BCR and antibody isotypes are generated by class switch recombination (CSR), which occurs in antigen-activated B cells and is mediated by the activation-induced cytidine deaminase (AID). In the course of an immune response, AID also "fine-tunes" antibody specificities by inducing mutations at the antigen-binding variable region (somatic hypermutation).

1.2. V(D)J recombination in B cells

The diversity in the primary Ig repertoire is generated by V(D)J recombination, a somatic DNA rearrangement process that occurs during B cell development in the bone marrow (Murphy, 2012). In their germline configuration, Ig chains are organized into three distinct loci (mouse/human): *Igλ*/*IGL*, *Igκ*/*IGK* and *Igh*/*IGH* (λ, κ and heavy chain, respectively). The variable domain of each chain is encoded by different sets of gene segments: variable (V), diversity (D) and joining (J, Figure 3; (Murphy, 2012)).

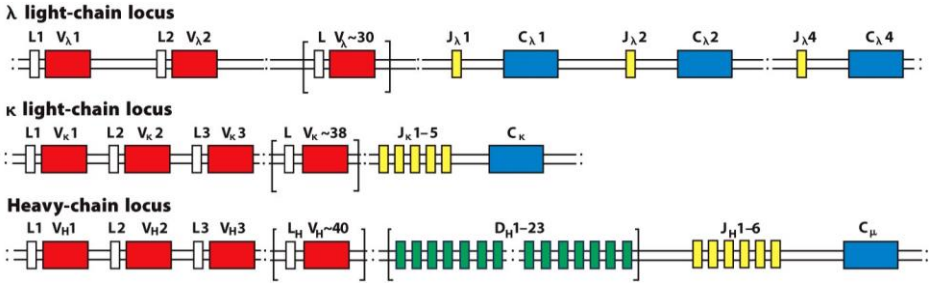

Figure 3: Germline organization of human *IG* loci.

From top to bottom: *IGL* (chromosome 22), *IGK* (chromosome 2), and *IGH* (chromosome 14). Each locus contains a variable number of V (red), D (green) and J (yellow) segments. In addition, there are one or multiple genes encoding for constant regions (C, blue). Upstream leader sequences (L) direct Igs into the cell's secretory pathways. Murine *Ig* loci are located on chromosome 16 (*Igλ*), chromosome 6 (*Igκ*) and chromosome 12 (*Igh*). Modified from (Murphy, 2012).

During V(D)J recombination, the variable domains of heavy and light chains are sequentially assembled to form a functional antigen-binding variable region (Figure 4A; (Murphy, 2012)). The variable domain of the heavy chain is generated by randomly combining a D and a J segment and subsequently by joining a V segment to the combined DJ sequence. The resulting VDJ exon is transcribed and spliced to the downstream constant region to generate the final Ig heavy chain mRNA used for translation. The variable domain of the light chain is generated by a single joining step between a V and a J segment. Analogous to the Ig heavy chain, the final light chain mRNA is generated by splicing of the transcribed VJ exon to the downstream constant region.

V(D)J recombination occurs at specific steps during B-cell development (Figure 4B; (Murphy, 2012)). At the pro-B cell stage, the IgM heavy chain is assembled. In large pre-B cells, the assembled heavy chain is functionally tested by forming a pre-BCR with a surrogate light chain. If the heavy chain is functional, rearrangement at *IGH* stops (allelic exclusion). Otherwise heavy chain rearrangement is repeated using the second *IGH* allele. If that fails as well, the cell dies. In small pre-B cells, the IgM light chain is assembled. Rearrangement starts at either *IGL* or *IGK* and terminates if a productive light chain has been generated (allelic exclusion). Non-productive rearrangements can be rescued by rearranging unused gene segments at the same allele, by using the second light chain allele and finally by restarting the rearrangement at the second light chain locus. Since productive light chain rearrangement terminates V(D)J recombination, B cells express only one type of light chain (isotypic exclusion). In immature B cells, the functional IgM BCR is tested for its reactivity towards “self” antigens (central tolerance). B cells that are self-tolerant leave the bone marrow and complete their development in the peripheral lymphatic organs. Finally, mature B cells co-express IgD and IgM BCRs through alternative splicing of their heavy chain transcripts.

The random combination of different gene segment variants and the pairings between heavy and light chains generate a vast Ig diversity (combinatorial diversity) which is further increased by nucleotide additions/deletions during the joining of gene segments (junctional diversity, see Figure 7). Overall, the primary Ig repertoire of naïve human B cells is estimated to contain at least 10^{11} different BCRs/antibodies (Murphy, 2012).

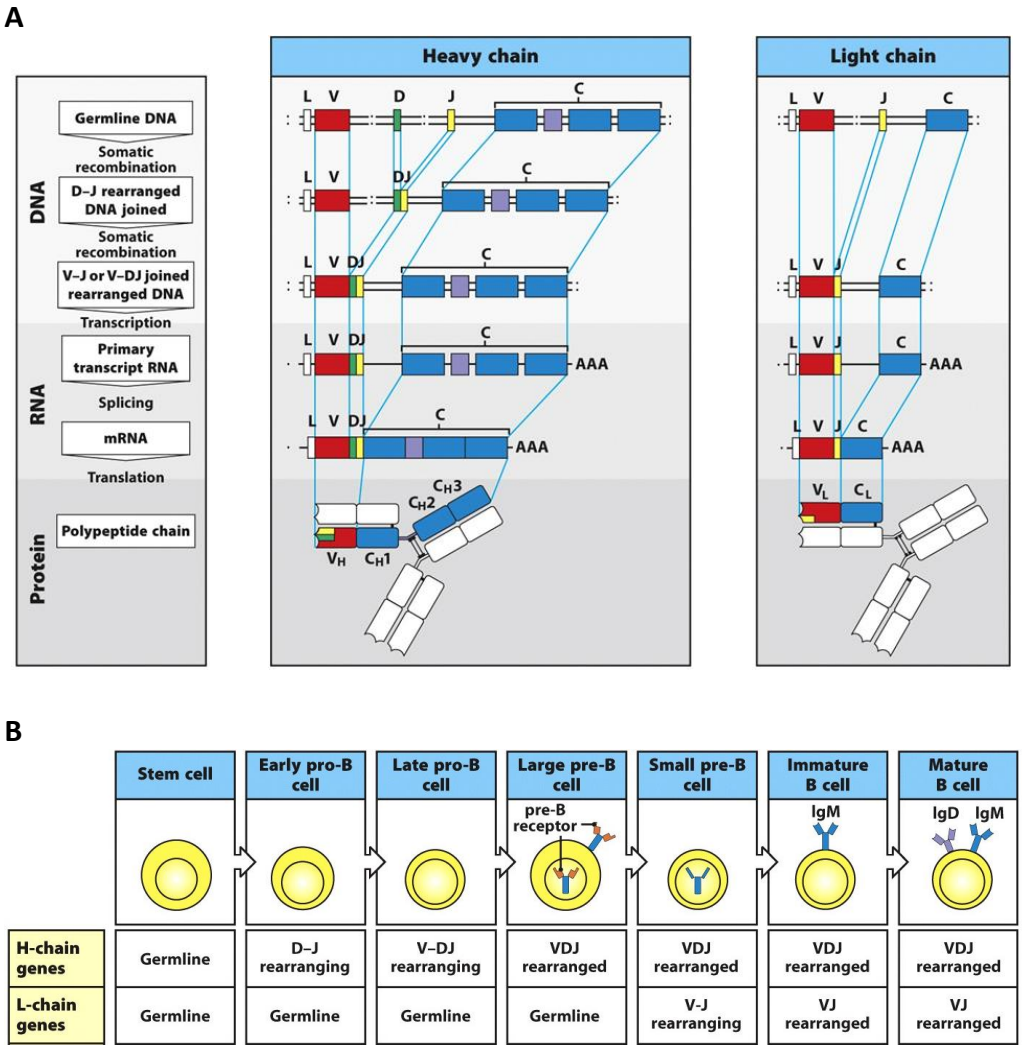

Figure 4: V(D)J recombination during B-cell development.

A- Generation of Ig chains. Heavy chain (middle): The heavy chain variable domain (V_H) is constructed in two steps. First, a D segment (green) is joined to a J segment (yellow). Then, the resulting DJ product is joined to a V segment (red). The constant region (C_H1, C_H2, C_H3) originates from a downstream constant gene (C) with several exons (blue and purple) and is fused to the VDJ product by splicing. It contains a flexible hinge region (purple) that links each antibody arm to its stem. Light chain (right): The light chain variable domain (V_L) is constructed in one step by joining of a V (red) to a J (yellow) segment. The constant region (C_L) originates from a downstream constant gene (C, blue) and is fused to the VJ product by splicing. Leader peptides (L) direct Igs into the cell's secretory pathways and are then cleaved. Modified from (Murphy, 2012).

B- Developmental stages of B cells. From left to right: B cells derive from hematopoietic stem cells. In pro-B cells, the Ig heavy chain (H-chain) is assembled by D-J and V-DJ rearrangement (early and late pro-B cells, respectively). Successful assembly generates pre-B cells which first test the Ig heavy chain by forming a pre-BCR (large pre-B cells) and then assemble the Ig light chain (L-chain) by V-J rearrangement (small pre-B cells). Successful assembly generates immature B cells which express IgM BCRs and, after testing for self-tolerance, migrate to the peripheral lymphoid tissues. Here they become mature B cells which express both IgD and IgM BCRs through alternative splicing. Modified from (Murphy, 2012).

1.3. The RAG recombinase

V(D)J recombination is catalyzed by the RAG recombinase (RAG1/2), a heterotetrameric protein complex encoded by the recombination-activating genes *RAG1* and *RAG2* (Figure 5A; (Kim et al., 2015; Ru et al., 2015). RAG1 is the principal DNA binding and cleavage component of the recombinase. RAG2 is an essential co-factor and consists of a core portion ($RAG2^{core}$) minimally required for its activity and a C-terminal region important for efficiency, fidelity and ordering of V(D)J rearrangements (Figure 5B; (Akamatsu et al., 2003; Curry and Schlissel, 2008; Liang et al., 2002; Talukder et al., 2004; Sekiguchi et al., 2001)). Mice deficient for either RAG1 or RAG2 lack mature lymphocytes and only contain pro-B cells and early T cell progenitors due to their inability to perform V(D)J recombination (for T cells, see Chapter 1.5; (Mombaerts et al., 1992; Shinkai et al., 1992).

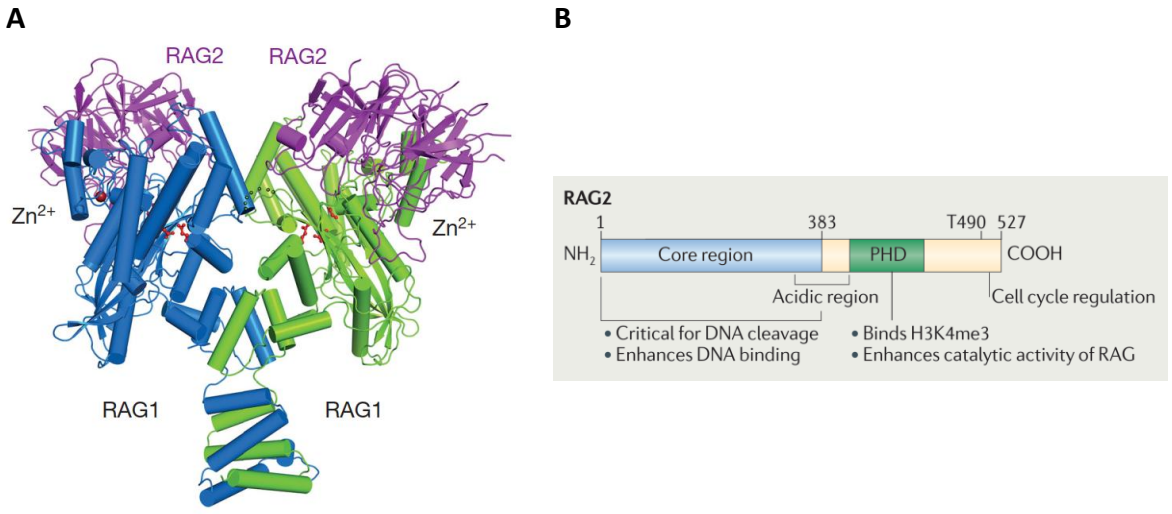

Figure 5: Structure and function of RAG1/2.

A- Crystal structure of RAG1/2 (ribbon diagram). RAG1/2 is comprised of two RAG1 chains (blue and green) and two RAG2 chains (both in magenta). Each RAG1-RAG2 subunit contains an active site (carboxylates shown as red sticks). Zinc ions (Zn^{2+}, dark red spheres) are coordinated by two zinc-binding motifs in RAG1. Modified from (Kim et al., 2015).

B- Schematic overview of RAG2. The core portion of RAG2 (amino acids 1-383, blue) is essential for DNA cleavage of RAG1/2 and also enhances its DNA binding ability. The RAG2 carboxyl-terminal region (amino acids 384-527, orange) enhances the catalytic activity of RAG1/2 and contains a plant homeodomain (PHD, green) that binds to trimethylated lysine 4 on histone H3 (H3K4me3). An acidic region upstream of the PHD also interacts with histones. RAG2 is only stable in G0 or G1 phase cells due to the phosphorylation of a conserved threonine residue (T490) in S, G2 and M phase cells. Modified from (Schatz and Ji, 2011).

1.4. Molecular mechanism of V(D)J recombination

During V(D)J recombination, RAG1/2 recognizes and cleaves conserved recombination signal sequences (RSSs) that flank each V, D, and J gene segment. RSSs are comprised of a conserved palindromic heptamer that is required for DNA cleavage, a degenerate spacer of 12 or 23 base pairs (bp), and a somewhat less-conserved A-rich nonamer that is important for RAG1/2 binding (Figure 6; (Schatz and Ji, 2011; Murphy, 2012)). RSSs with 12- or 23-bp spacers are termed 12RSSs and 23RSSs, respectively.

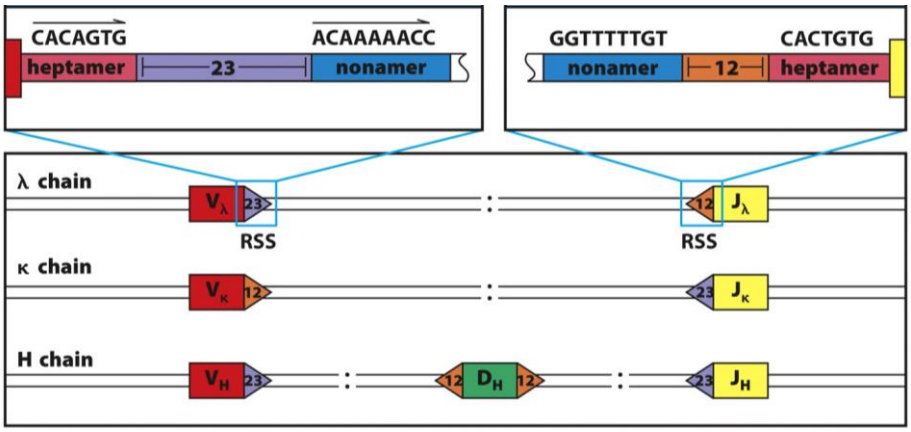

Figure 6: Organization of physiologic RSSs at *IG* loci.

Top: RSSs are comprised of a conserved heptamer (light red) and a somewhat less-conserved nonamer (blue) which are separated by a degenerate spacer of either 12 or 23 bp (orange and purple, respectively). Physiologic consensus sequences are shown on top. RAG1/2 cleavage occurs right upstream of the heptamer, whose first 3 bp are almost perfectly conserved in all physiologic RSSs (Schatz and Ji, 2011). Bottom: RSSs (triangles) flank each V(D)J gene segment (boxes) at *IGL*, *IGK*, and *IGH* (λ, κ, and H chain, respectively). V and J segments (red and yellow, respectively) are flanked by one RSS, whereas D segments (green) are flanked by two RSSs. Physiologic recombination occurs between a 12RSS (orange) and a 23RSS (purple). Modified from (Murphy, 2012).

During V(D)J recombination, RAG1/2 first binds to a single 12- or 23RSS (signal complex) and then captures a "complementary" 23- or 12RSS (paired complex) according to the "12/23 rule" (Figure 7; (Schatz and Ji, 2011; Schatz and Swanson, 2011)). Upon synapsis, RAG1/2 introduces DNA double-strand breaks (DSBs) between coding sequences and flanking RSSs by making a single-strand nick which is used to catalyze a *trans*-esterification that produces a hairpin-sealed coding end and a blunt-cut signal end. After cleavage, RAG1/2 remains associated with paired coding and signal ends in a post-cleavage complex, thereby scaffolding their repair by non-

homologous end joining (NHEJ). Coding ends are fused to produce V(D)J exons that form the Ig variable region and ligation of signal ends generates non-coding signal joints.

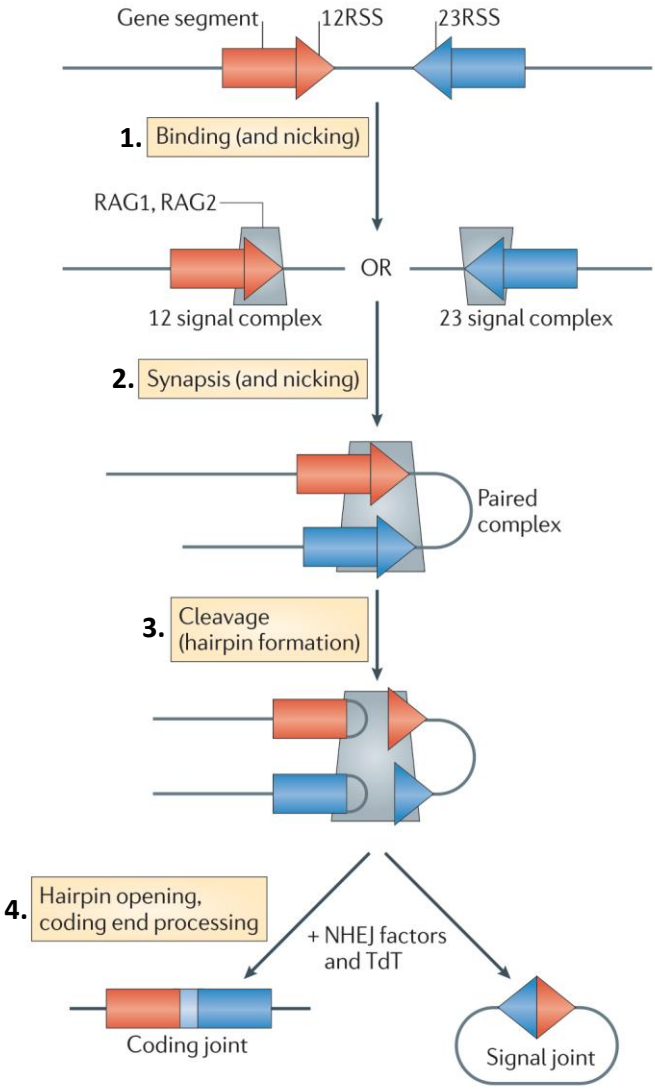

Figure 7: Mechanism of V(D)J recombination.

From top to bottom: Each V(D)J gene segment (red and blue boxes) is flanked by a corresponding 12- or 23RSS (red and blue triangles, respectively). In the first step, RAG1/2 (grey box) forms a 12 or 23 signal complex by binding to a 12- or 23RSS, respectively. Second, RAG1/2 captures a "complementary" 23- or 12RSS (synapsis). Within the paired complex RAG1/2 introduces a DNA single strand break (nicking) between each gene segment and its flanking RSS (not shown). Alternatively, nicking might already occur within the signal complex in the previous step. Third, RAG1/2 introduces DSBs through transesterification which generates hairpin-sealed coding ends (left) and blunt-cut signal ends (right). Fourth, NHEJ factors supported by the RAG1/2 post-cleavage complex join cleaved ends to generate coding and signal joints. The formation of coding joints requires hairpin opening and DNA processing, which frequently causes nucleotide deletions and additions. Moreover, non-template nucleotides (light blue) are added by the terminal deoxynucleotidyl transferase (TdT). Modified from (Schatz and Ji, 2011).

Depending on the orientation of involved RSSs, RAG1/2 catalyzes either deletional (convergent RSSs) or inversional (head-to-tail RSSs) rearrangements. During deletional V(D)J recombination signal joints are released as episomes (Figure 8A), whereas they remain in the genome during inversional recombination (Figure 8B; (Helmink and Sleckman, 2012)).

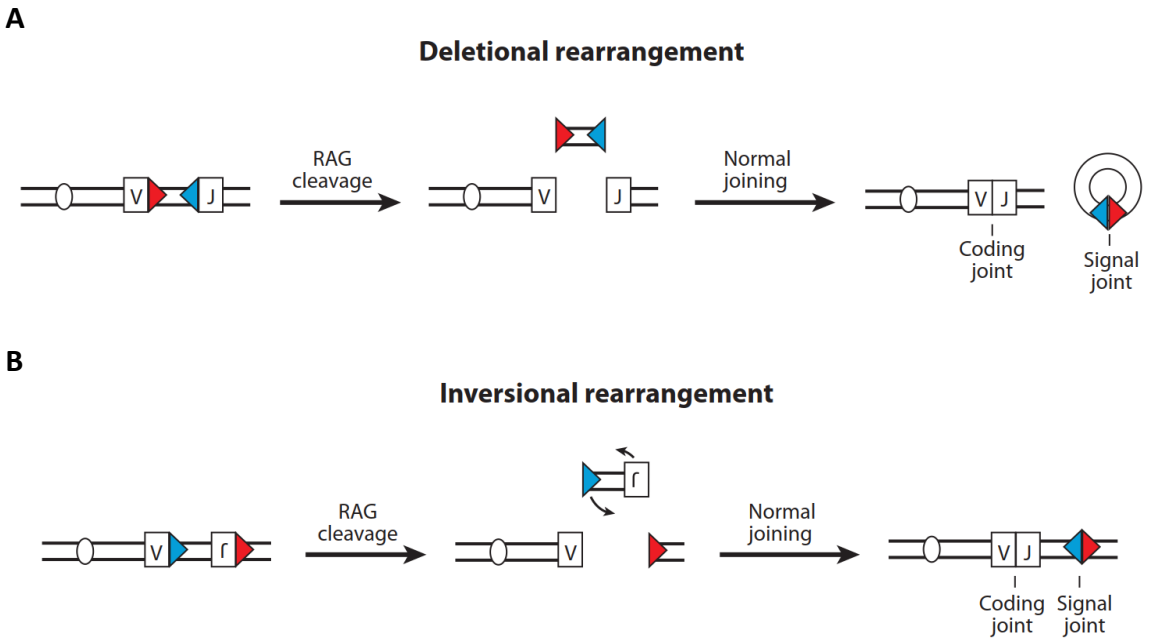

Figure 8: Deletional and inversional V(D)J rearrangements.

A- Deletional V(D)J rearrangement. From left to right: Recombination of convergent RSSs (red and blue triangles) induces DNA deletion which generates genomic coding and episomal signal joints.

B- Inversional V(D)J rearrangement. From left to right: Recombination of head-to-tail RSSs (red and blue triangles) induces DNA inversion which generates genomic coding and signal joints.

A to B modified from (Helmink and Sleckman, 2012).

1.5. V(D)J recombination in T cells

T lymphocytes (T cells) form the second part of the adaptive immune system. They express immunoglobulin-like T cell receptors (TCRs) that resemble membrane-bound antibody fragments (Figure 9A; (Murphy, 2012)). TCRs are comprised of two transmembrane glycoprotein chains (TCRα and TCRβ) whose variable domains are assembled by RAG1/2-mediated V(D)J recombination during T-cell development in the thymus (Figure 9B; (Murphy, 2012)). In contrast to Igs, TCRs only bind to antigens that have been partly degraded inside host cells and are presented by the major histocompatibility complex (MHC) proteins on host cell surfaces

(Murphy, 2012). Moreover, T cells do not secrete soluble equivalents of their TCRs upon activation. Instead they are specialized on cell-cell interactions and mediate distinct effector functions based on their class and subtype (Murphy, 2012). Helper T cells (T_H) provide necessary co-stimulatory signals to activate antigen-stimulated B cells (T_H1 and T_H2 subtypes) or infected macrophages (T_H1 subtype) and recruit neutrophils (T_H17 subtype). Cytotoxic T cells directly engage and kill host cells that are infected with intracellular pathogens (e.g. viruses). Regulatory T cells (T_{reg}) suppress the activity of other lymphocytes and thereby help to control immune responses.

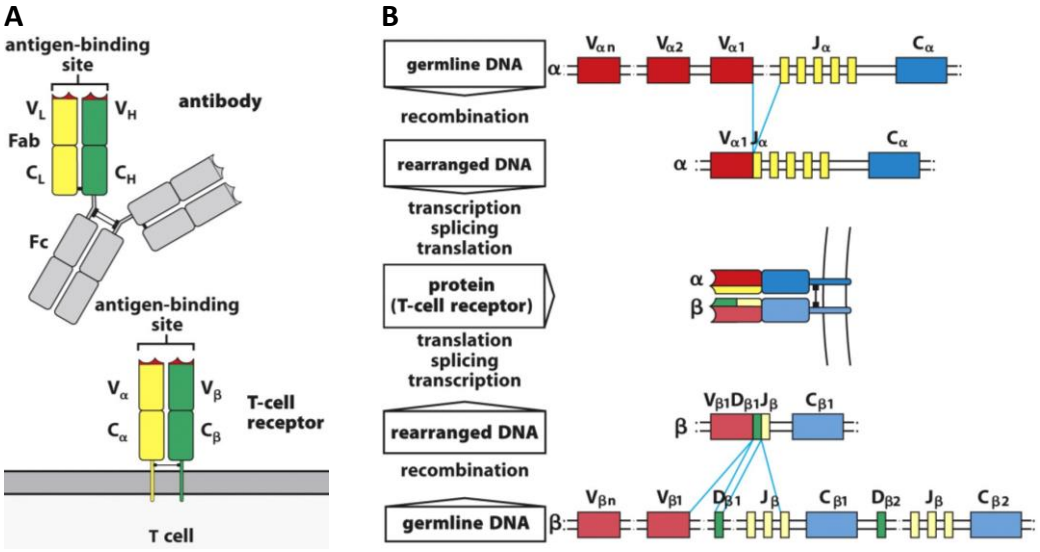

Figure 9: TCRs and V(D)J recombination.

A- Comparison between antibodies and TCRs. Top, an antibody with its two identical arms (each termed fragment antigen binding, Fab) and its stem (termed fragment crystallizable, Fc). Bottom, a TCR expressed on a T cell. Both Fab and TCR are comprised of two polypeptide chains (yellow and green) that each contain one variable domain (V_L, V_H and V_α, V_β, respectively) and one constant domain (C_L, C_H and C_α, C_β, respectively). The variable domains of the TCR form an antigen-binding site (red) similar to that of the Fab (red). However, TCRs only bind to antigen fragments presented by host MHC proteins (not shown). Modified from (Murphy, 2012).

B- V(D)J recombination in T cells. TCRs are encoded by variable (V, red), diversity (D, green) and joining (J, yellow) gene segments and constant genes (C, blue) at the *Tcrα*/*TRA* (top) and *Tcrβ*/*TRB* (bottom) loci (mouse/human, respectively). During V(D)J recombination, the variable domain of the TCRβ chain (bottom) is generated by D-J and V-DJ rearrangements. The resulting VDJ exon is transcribed and spliced to a downstream Cβ gene to generate the final TCRβ mRNA used for translation. The variable domain of the TCRα chain (top) is generated by a single joining step between a V and a J segment. Analogous to the TCRβ chain, the final TCRα mRNA is generated by splicing of the transcribed VJ exon to the Cα gene. A subset of T cells bear an alternative TCR comprised of γ and δ chains which are encoded at the *Tcrγ*/*TRG* and *Tcrδ*/*TRD* loci, respectively (mouse/human, respectively; not shown). Modified from (Murphy, 2012).

1.6. Aberrant V(D)J recombination

1.6.1. RAG1/2-mediated chromosomal translocations and aberrant DNA deletions

In addition to its essential role in adaptive immunity, RAG1/2 has been implicated in the genesis of chromosome translocations and aberrant DNA deletions associated with lymphoid malignancy (Lieber, 2016; Roth, 2003). Mice deficient for ataxia-telangiectasia mutated kinase (ATM) or both the tumor suppressor protein 53 (p53) and components of the NHEJ machinery develop RAG1/2 dependent chromosome translocations associated with pro-B cell lymphomas (Alt et al., 2013; Nussenzweig and Nussenzweig, 2010). In humans, RAG1/2 is implicated in the genesis of follicular lymphoma (FL), mantle cell lymphoma (MCL), and acute lymphoblastic leukemia (ALL), all of which carry genome aberrations in the proximity of RSSs in antigen receptor genes or non-physiologic cryptic RSSs (cRSSs) with conserved heptamer motifs (Alt et al., 2013; Küppers and Dalla-Favera, 2001; Nussenzweig and Nussenzweig, 2010). Predicted cRSSs are broadly distributed throughout the genome and so are RAG1/2 binding sites as assayed by chromatin immunoprecipitation (Ji et al., 2010; Lewis et al., 1997; Merelli et al., 2010; Teng et al., 2015). Consistent with the idea that RAG1/2 can induce DNA damage at cRSSs, it causes chromosomal deletions, and in the context of ATM deficiency also translocations, between engineered RSSs and genomic cRSSs in primary pro-B cells and pro-B cell lines (Hu et al., 2015). The reported off-target mechanism involves directional, linear tracking of RAG1/2 within chromosomal loop domains to locate RSS/cRSS pairs (Hu et al., 2015).

1.6.2. RAG1/2-mediated DNA insertions

Biochemical experiments as well as episomal assays in cell lines and yeast suggest that RAG1/2 can mediate DNA insertions through transposition (Agrawal et al., 1998; Chatterji et al., 2006; Clatworthy et al., 2003; Elkin et al., 2003; Hiom et al., 1998; Lee et al., 2002; Neiditch et al., 2002; Posey et al., 2006; Reddy et al., 2006; Tsai et al., 2003). During DNA transposition, RAG1/2 excises RSS-flanked DNA fragments from donor sequences and mediates their re-integration at target sites through transesterification (“cut and paste”, Figure 10A). The insertion process utilizes the free hydroxyl groups of the cleaved RSSs to attack the phosphodiester bonds at the

target loci. The resulting staggered DNA breaks cause characteristic target-site duplications. Interestingly, similar reactions are a well-characterized feature of bacterial transposases which is why RAG1/2 is thought to have evolved from an ancient "RAG transposon". Moreover, it has also been proposed that this "RAG transposon" contributed to the evolution of modern *IG* and *TCR* loci by fragmenting the precursors of antibody receptor genes through recurrent DNA transpositions (Agrawal et al., 1998; Hiom et al., 1998).

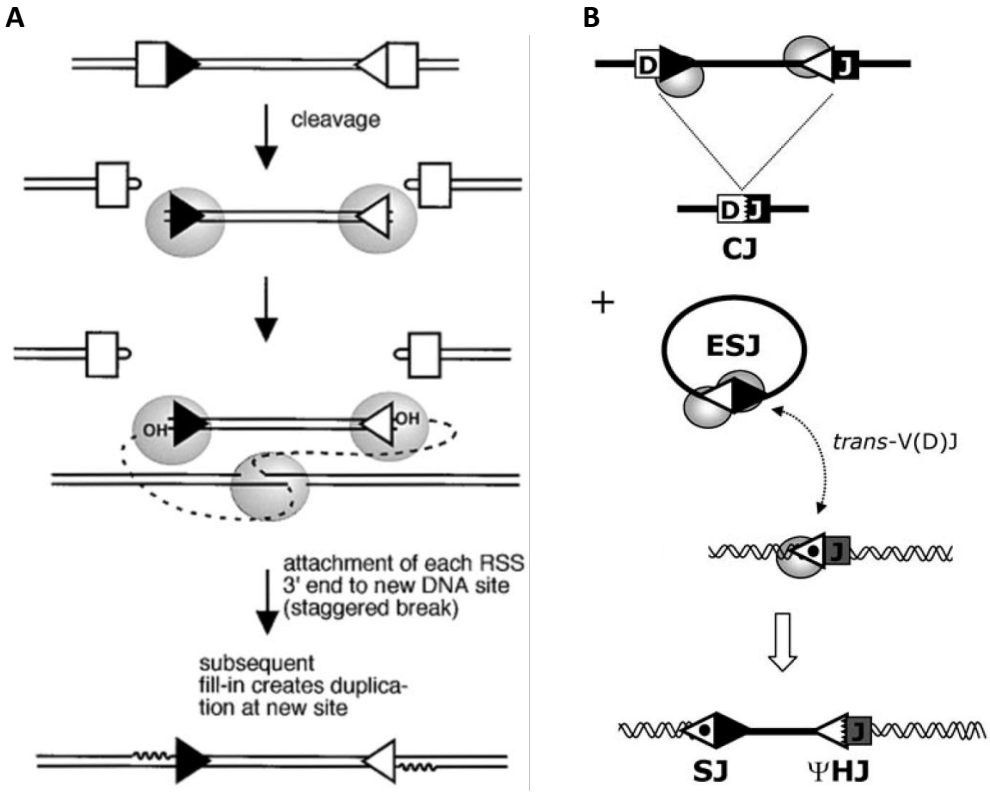

Figure 10: RAG1/2-mediated DNA insertion through transposition and trans-V(D)J recombination.

A- RAG1/2-mediated DNA transposition. RAG1/2 (grey spheres) excises a DNA fragment flanked by RSSs (black and white triangles) from a donor site and subsequently mediates its re-integration into unrelated target DNA by transesterification using the free 3' hydroxyl groups (OH) of the cleaved RSSs. The insertion process generates staggered DNA breaks at the target site which leads to characteristic target-site duplications (jagged lines). Modified from (Lewis and Wu, 2000).

B- Trans-V(D)J recombination. RAG1/2 (grey spheres) excises DNA flanked by RSSs (black and white triangles) from a donor site which generates a genomic coding joint (CJ) and an episomal signal joint (ESJ). Subsequently, RAG1/2 recombines the ESJ *in trans* with another V(D)J gene segment which leads to its re-integration and the formation of a genomic signal joint (SJ) and a "pseudo-hybrid" joint (ΨHJ). The latter is characterized by extensive DNA processing (nucleotide deletions and additions, jagged line). Modified from (Vanura et al., 2007).

In addition, experiments with reporter cell lines indicate that RAG1/2 can mediate DNA insertions through trans-V(D)J recombination (Reddy et al., 2006). During trans-V(D)J recombination, RAG1/2 excises RSS-flanked DNA as episomal signal joints and then mediates their re-integration at endogenous RSSs or cRSSs through ongoing V(D)J recombination (Figure 10B).

1.7. Translocation capture sequencing

In 2011, Klein et al. published a next-generation sequencing technique to capture and sequence rearranged genomic DNA (translocation capture sequencing, TC-Seq; (Klein et al., 2011; Oliveira et al., 2012)). In their system, a restriction site for the I-SceI endonuclease, which is normally absent from the mouse genome, is introduced at a specific locus, e.g. at the first intron of the myelocytomatosis oncogene *c-myc* (*Myc*^{I}). Subsequent expression of I-SceI induces a unique DSB at *Myc*^{I} which serves as "bait" to capture concurrent DNA breaks in the genome (Figure 11A). Chromosomal rearrangements between the I-SceI break and the genome are amplified by semi-nested polymerase chain reaction (PCR), deep-sequenced and analyzed computationally (Figures 11B and 11C).

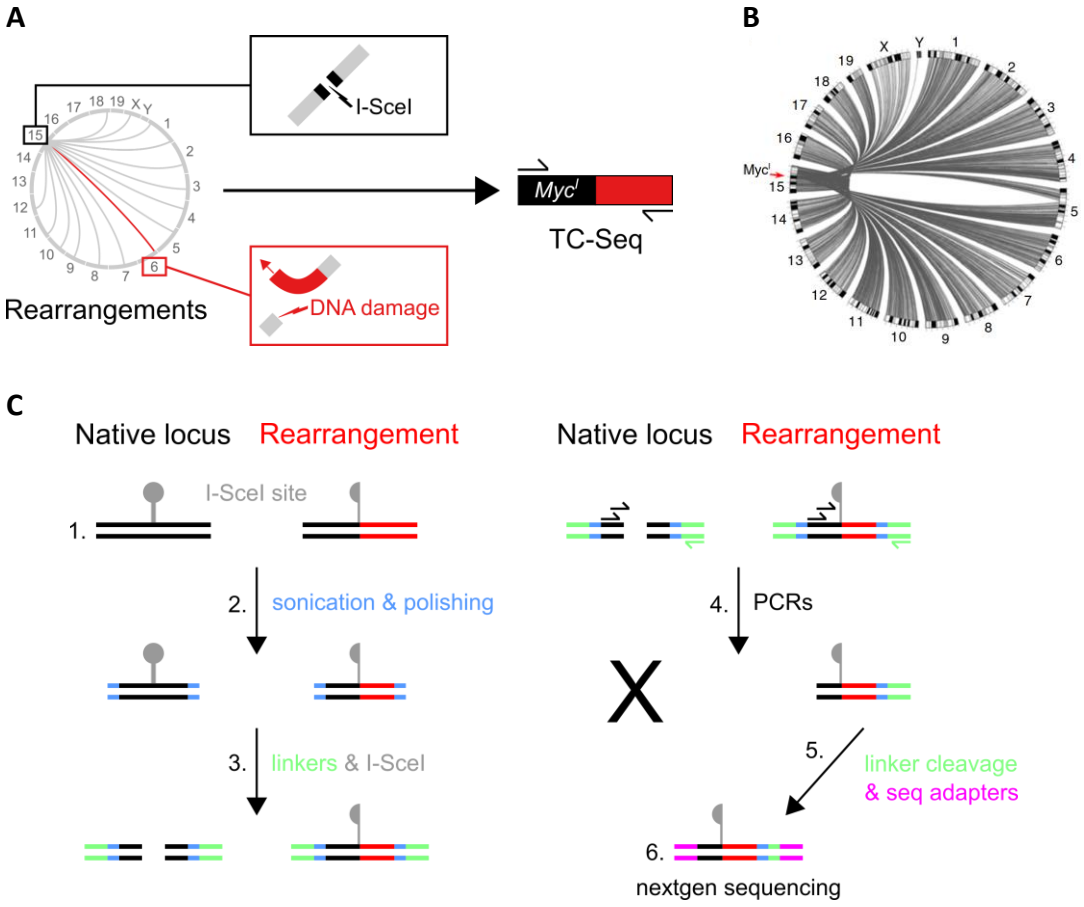

Figure 11: Preparation and analysis of TC-Seq libraries.

A- Basic principle of TC-Seq. Left: The mouse genome is represented as circle (grey, labeled with chromosomes). I-SceI induces a DSB at *Myc^I* on chromosome 15 (black box) while independent DNA damage factors cause concurrent DNA breaks throughout the genome, e.g. on chromosome 6 (red box). Breaks at the I-SceI site and the genome recombine and form rearrangements (lines inside the circle). Right: During TC-Seq, rearrangements to *Myc^I* are amplified by semi-nested PCR (black arrows), deep-sequenced and analyzed computationally (see also Figure 11C).

B- Genome-wide rearrangements detected by TC-Seq. Circos plot of chromosomal rearrangements to *Myc^I* in activated, splenic AID$^{-/-}$ B cells. Each line represents a unique recombination event between the DSB at *Myc^I* (red arrow) and a break in the genome. Modified from (Klein et al., 2011).

C- TC-Seq library preparation (simplified). 1. The I-SceI restriction site (grey pin) at *Myc^I* (black) is intact at the native but not at the rearranged locus (left and right, respectively) 2. Genomic DNA is fragmented by sonication and single-stranded overhangs are removed (polishing) to generate blunt ends (blue). Since sonication generates random DNA breaks, the produced ends are used to identify unique rearrangements during the computational analysis. 3. PCR-linkers (green) are ligated and DNA fragments are digested with I-SceI to degrade native DNA fragments (see left). 4. DNA fragments are amplified by semi-nested ligation-mediated PCR (black and green arrows). Digested DNA is not enriched (X). 5. PCR-linkers are cleaved by restriction digest. A small linker fragment (green) remains and is used as control barcode during the computational analysis. Finally, sequencing adapters (pink) are ligated. 6. DNA fragments are deep-sequenced using paired-end next-generation sequencing and analyzed computationally.

1.8. Aims of the thesis

RAG1/2 has been implicated in the genesis of chromosome translocations and aberrant DNA deletions associated with lymphoid malignancy (see Chapter 1.6.1). In addition, RAG1/2 has the capacity to induce aberrant DNA insertions through transposition and trans-V(D)J recombination (see Chapter 1.6.2). However, in contrast to translocations and aberrant deletions, only few putative RAG1/2-mediated genomic insertions have been documented *in vivo* (Curry et al., 2007; Messier et al., 2003; Vanura et al., 2007). Moreover, those observed in cancer reveal features that are neither compatible with DNA transposition nor trans-V(D)J recombination (Navarro et al., 2015). Thus, how RAG1/2 causes genomic DNA insertions still remains largely unknown.

The general aim of my thesis was to characterize the potential of RAG1/2 to promote genome destabilization and lymphomagenesis on a genome-wide scale. My initial experiments were geared at identifying RAG1/2^{core}-induced chromosome translocations throughout the genome using TC-Seq. Upon detecting a peculiar rearrangement pattern in a subset of events, my major objectives were:

1. to analyze if RAG1/2^{core} mediates aberrant DNA deletions that are compatible with the observed rearrangement pattern (see Chapter 3.3)
2. to screen the TC-Seq data for bona fide insertions (see Chapter 3.4)
3. to develop a novel assay to specifically detect chromosomal insertions from donor sites throughout the genome (see Chapter 3.5)
4. to verify the occurrence of aberrant DNA insertions under physiologic conditions in the presence of RAG1/2 wild type (see Chapter 3.5)
5. to screen for similar insertions at physiologic DNA breaks *in vivo* (see Chapter 3.6)
6. to explore if related insertions might contribute to human cancer (see Chapter 3.6 and Discussion)

2. Materials and methods

2.1. Mice

Mutant mice used in this study include *RAG2*$^{-/-}$*Myc*$^{l/l}$ (B6(Cg)-*Rag2*$^{tm1.1Cgn}$/J, The Jackson Laboratory and (Robbiani et al., 2008)), *ROSA*^{erISCEI}*Myc*$^{l/l}$*Igh*$^{l/l}$ and *ROSA*^{erISCEI}*Myc*$^{l/l}$*Igh*$^{l/l}$*AID*$^{-/-}$ (Robbiani et al., 2015). All mice were in a C57BL/6 background or backcrossed to it for at least 10 generations. All experiments were performed in agreement with protocols approved by the Rockefeller University Institutional Animal Care and Use Committee.

2.2. Retroviruses

Murine RAG2 (RAG2^{full}) and RAG2^{core} sequences were amplified from mouse genomic DNA using primers p2/p6 and p3/p6, respectively (Table S5). I-SceI was amplified from pMX-I-SceI-EGFP using primers p4/p5 (Table S5; (Robbiani et al., 2008)). Overlap extension PCRs of the above products with primers p2/p4 and p3/p4 generated I-SceI-P2A-RAG2^{full} and I-SceI-P2A-RAG2^{core}, respectively (Table S5). Finally, both constructs were cloned into pMX-EGFP to generate pMX-I-SceI-P2A-RAG2^{full}-EGFP and pMX-I-SceI-P2A-RAG2^{core}-EGFP, respectively (Figure S1).

2.3. Cell culture and infection for TC-Seq

Pro-B cells were isolated from tibias, femurs and humeri of *RAG2*$^{-/-}$*Myc*$^{l/l}$ mice at 4-10 weeks of age by immunomagnetic enrichment with anti-B220 MicroBeads (Miltenyi Biotech). Cells were cultured at 2.0 x 10^6 cells/ml in the presence of IL-7 (5 ng/ml, SIGMA) in complete RPMI (RPMI-1640 supplemented with L-glutamine (GIBCO), sodium pyruvate (GIBCO), antibiotic/antimycotic (GIBCO), HEPES (GIBCO), 55 µM beta-mercaptoethanol (GIBCO), and 10% fetal calf serum (Hyclone)). IL-7 was replenished on day 2. On days 3 and 4, cell supernatants were replaced with retroviral supernatants resulting from co-transfection (Fugene-6, Roche) of BOSC23 cells with pCL-Eco and pMX-I-SceI-P2A-RAG2^{core}-EGFP or pMX-I-SceI-EGFP plasmids 3 days before (Robbiani et al., 2008). Spinoculation was at 1111 g for 1.5h in the presence of 2.5 µg/ml polybrene, 5 ng/ml IL-7 and 20 mM HEPES. After 6-8h at 37°C, on day 3 retroviral supernatants were replaced with original supernatants, while on day 4 cells were collected for IL-7 washout

and re-plating in fresh complete RPMI. Cells were harvested after 2.5 days of IL-7 depletion, sorted for EGFP-expression with a FACSAria instrument (Becton Dickson), pelleted, and snap-frozen on dry ice. Samples infected with pMX-I-SceI-P2A-RAG2^{core}-EGFP are referred to as RAG2^{core} and those infected with pMX-I-SceI-EGFP are referred to as RAG2$^{-/-}$.

2.4. Cell culture for IC-Seq

Bone marrow B cells were isolated from tibias, femurs and humeri of *ROSA*^{erISCEI}*Myc*$^{l/l}$*Igh*$^{l/l}$ and *ROSA*^{erISCEI}*Myc*$^{l/l}$*Igh*$^{l/l}$*AID*$^{-/-}$ mice at 6-8.5 months of age by immunomagnetic enrichment with anti-B220 MicroBeads (Miltenyi Biotech). Cells were pooled and cultured at 2.0 x 10^6 cells/ml in the presence of IL-7 (5 or 10 ng/ml, SIGMA) and Tamoxifen (1 µM, SIGMA) in complete RPMI. On day 1, cells were collected for IL-7 washout and re-plated in fresh complete RPMI with 1 µM Tamoxifen. On day 2, cultures were harvested and cell pellets snap-frozen on dry ice.

2.5. TC-Seq library preparation

TC-Seq libraries of RAG2^{core} and RAG2$^{-/-}$ pro-B cells were prepared in duplicates from each of 50 million sorted cells, as previously described (Klein et al., 2011; Robbiani et al., 2015) with the exception that sonication of genomic DNA was performed with Covaris S220 (power 105, duty factor 5%, cycles 200, time 35s, water level 12, temperature 7°C) yielding a core of DNA fragments between 500-850 bp. Each library was sequenced twice using Illumina MiSeq (300 cycles, paired-end).

2.6. IC-Seq library preparation

IC-Seq libraries of bone marrow B-cells were prepared in duplicates from 40 million and 60 million cultured cells. Genomic DNAs were extracted with phenol-chloroform following Proteinase K digestion, washed twice with 70% ethanol and resuspended in TE buffer (Invitrogen). For the first PCR, 1 µg of DNA was amplified in each reaction with Phusion polymerase (NEB) and the *Myc*^{l} flanking primers p247/p251 with the following conditions: 98°C for 2 min; 35x (98°C for 10 s, 72°C for 1:30 min); 72°C for 5 min (Table S5). Pooled PCR reactions

were column purified (Macherey-Nagel) and high molecular weight products (1500-5000 bp) were isolated by agarose gel electrophoresis. Extracted DNA was digested with I-SceI (NEB) and column purified (Macherey-Nagel). In the second PCR, 25 ng DNA were amplified in each reaction with Phusion polymerase (NEB) and primers p274a/p275a, p274b/p275b, p274c/p275c and p274d/p275d with the following conditions: 98°C for 2 min; 3x (98°C for 10 s, 65°C for 30 s, 72°C for 1 min); 32x (98°C for 10 s, 72°C for 1:15 min); 72°C for 5 min (Table S5). PCR products were pooled and high molecular weight amplicons (280-3000 bp) were isolated by agarose gel electrophoresis. Extracted DNA was digested with I-SceI (NEB) and column purified (Macherey-Nagel). To add index adapters for sequencing, the PCR was similar as the second PCR but with primers pNextflex common/pNextflex index5 or pNextflex common/pNextflex index6 with the following conditions: 98°C for 2 min; 3x (98°C for 10 s, 67°C for 30 s, 72°C for 1 min); 32x (98°C for 10 s, 72°C for 1:15 min); 72°C for 5 min (Table S5). PCR products were pooled and high molecular weight amplicons (350-2000 bp) isolated by agarose gel electrophoresis. Extracted DNA was digested with I-SceI (NEB), column purified (Macherey-Nagel) and high molecular weight products (300-2000 bp) were isolated once more by agarose gel electrophoresis. Extracted DNA was sequenced twice using Illumina NextSeq (150 cycles, paired-end).

2.7. TC-Seq analysis

Two independent libraries were sequenced twice and the data pooled for analysis using a novel pipeline to identify rearrangement and insertion breakpoints. First, sequencing reads were trimmed for high quality with seqtk (error rate threshold of 0.01; Broad Institute) and those with primer sequences from the first PCR or less than 5 bp of *Myc*^{I} following the nested primer sequence were discarded. Second, reads were mapped against *Myc*^{I} with its repetitive regions masked using SMALT (v0.7.6, parameters: -c 11 -x -O; Sanger Institute). Paired reads that both aligned to *Myc*^{I} at their 5' end were analyzed in "insertion mode", otherwise they were processed in "rearrangement mode".

In rearrangement mode (Figures 13, 15, 16 and 17), bases aligning to *Myc*^{I} were clipped from either the beginning or the end of the reads and the remaining sequences were mapped to the mouse genome (mm10) with SMALT (parameters: -O -r -1). Only alignments with at least 36 bp and a Phred score of 20 were accepted. Reads with the same sheared ends, which derive from

sonication during library preparation, were merged into one event and single reads were preserved. Rearrangements that did not yield breakpoints were discarded. Finally, reads that crossed the I-SceI site by more than 3 bp were excluded.

In insertion mode (Figures 20, 21 and 22), bases aligning to *Myc*^{I} were clipped from both ends of the reads and the remaining sequences were mapped to the mouse genome (mm10) with SMALT (parameters: -O -r -1). Only alignments with at least 36 bp and a Phred score of 20 were accepted. Pairs with incorrect genomic orientation (+/+ and -/-) were excluded. The alignment of insertions yielded either both genomic breakpoints (double junctions) or only one (single junctions). Because of saturation at *Myc*^{I}, events were merged if they possessed all of the following features: identical shears, genomic breakpoints within 5 bp and same orientation. Events based on single reads were preserved. Finally, reads that crossed the I-SceI site by more than 3bp were excluded.

2.8. IC-Seq analysis

Data from two independent libraries were pooled and analyzed similar to the “insertion mode” in TC-Seq, with minor modifications (Figure 24). Only genomic alignments with at least 25 bp and a Phred score of 20 were accepted. Insertions were merged if they possessed genomic breakpoints within 5 bp of each other and occurred in the same orientation. Finally, reads that crossed the I-SceI site by more than 3 bp were excluded.

2.9. Analysis of rearrangements (TC-Seq) and insertions (IC-Seq)

To characterize chromosomal rearrangements and insertions derived from distal regions (Figures 13C, 13D, 13E, 20C, 20D and 20E), the following portions of the genome were excluded: 50 kilobases (kb) or 20 kb surrounding the I-SceI site at *Myc*^{I} (rearrangements or insertions, respectively), 2 kb surrounding cryptic I-SceI sites (consensus [TCA][AT]GGGATA[AC]CAGG[GCT][TC][ATC][AG][TAC]), *RAG2* (likely representing retroviral integrations), 3 megabases (Mb) at each centromere and chromosome M (mitochondrial DNA).

To determine the enrichment at genic regions (Figures 13C and 20C), the portion of DNA from -2 kb of the most upstream transcription start site to the end of the last exon was considered as

genic. For transcription analysis (Figures 13D and 20D), RNA-seq data (Revilla-i-Domingo et al., 2012) were mapped with STAR aligner (v2.4.2a, default parameters; (Dobin et al., 2013)) using the mouse genome (mm10) and removing multiple alignments. Transcripts were quantified and annotated using cufflinks (v2.2.1, cuffdiff parameters: --upper-quartile-norm --dispersion-method per-condition; (Trapnell et al., 2013)) and Ensembl annotation (release 80). Transcription groups were defined using the mclust R package: silent (0 FPKM), trace (0.000000522291-2.8443 FPKM), low (2.84555-11.9418 FPKM), medium (11.9476-47.115 FPKM) and high (47.1191-74.211 FPKM). To detect enrichment within ERFS (Figures 13E and 20E), previously reported sites (Barlow et al., 2013) were lifted over from mouse genome mm9 to mm10 (UCSC LiftOver tool).

2.10. Detection of rearrangement breakpoint clusters (TC-Seq)

RAG1/2^{core}-dependent breakpoint clusters were detected by a three-step process. First, RAG2^{core} and RAG2$^{-/-}$ TC-Seq libraries were screened for local enrichment of rearrangement breakpoints to identify breakpoint hotspots (at least 3 breakpoints and a combined P value of less than 10^{-8}, (Klein et al., 2011)). To prevent potential sonication artifacts, hotspots were excluded if their sheared ends are either within less than 18 bp of each other or overlap with simple repeat regions. Second, breakpoint hotspots were defined as RAG1/2^{core}-dependent if they did not display any RAG2$^{-/-}$ breakpoints or sheared ends within +/- 1 kb distance. Third, breakpoint clusters containing 3 or more events within up to 25 bp distance of each other were identified within each RAG1/2^{core} hotspot. Off-target clusters were manually filtered based on the location of recurrent breakpoints near CA motifs that were shared by at least 3 clusters (CACA, CACC, CACT and CAGA). Simple CA-repeat regions were excluded. Putative cRSS sequences were manually detected and analyzed using Geneious (Kearse et al., 2012) and RSSsite (http://www.itb.cnr.it/rss; (Merelli et al., 2010)). Sequences of physiologic RSSs were obtained from IMGT (http://www.imgt.org/) and published RSS data sets (Cowell et al., 2002). Annotation of V(D)J segments was based on Ensembl (release 80). Rearrangements crossing the I-SceI site were still allowed during the detection of breakpoint hotspots and clusters, but afterwards manually removed from all sites in the final data.

2.11. Analysis of insertions in human tumors

A novel pipeline was established to search whole genome sequences for insertions derived from *IG*/*TCR* loci. First, *IG*/*TCR* baits were generated that correspond to regions spanning 150 bp upstream and downstream from each physiologic RSS cleavage site of human V and J segments (Ensembl, release 84). D segments were excluded and repeat regions were masked. Second, whole genome sequences from published human cancer datasets (Table S4; (Holmfeldt et al., 2013; Okosun et al., 2014; Zhang et al., 2012)) were mapped with bwa mem (v0.7.12-r1039, default parameters) using the *IG*/*TCR* baits as references. Third, paired reads aligning to the baits were mapped against the human genome (hg38) using bwa mem (v0.7.12-r1039, default parameters). Only alignments with a Phred score of at least 20 were accepted. Finally, reads containing junctions (chimeric alignments) were filtered to yield insertions which were then manually verified using Geneious (Kearse et al., 2012). The analysis of publicly available human cancer datasets was classified as exempt activity by The Rockefeller University Institutional Review Board.

2.12. Deletion PCR assays

Genomic DNAs of TC-Seq (RAG2^{core} and RAG2$^{-/-}$) and IC-Seq (RAG1/2 wild type) cultures were used for deletion PCR assays. Duplicates for RAG2^{core} and RAG2$^{-/-}$ originated from cell cultures with modified conditions: control was infected with pMX-EGFP; cells were transferred onto irradiated S17 stroma cells after IL-7 washout on day 4 and depleted for 1.5 days. In order to detect small and rare deletion events, nested PCRs with a “poison” primer were performed (Edgley et al., 2002). For PCRI, 100 ng (Jκ1/2, Jκ4/5) or 200 ng (Vκ3-1) genomic DNA was amplified in 20 µl reactions with HotStarTaq polymerase (Qiagen). For PCRII, 1µl of PCRI was used as template. For deletions at Jκ1/2, primers p195/p256/p258 (PCRI) and p196/p257 (PCRII) were used with the following conditions: PCRI, 95°C for 15 min; 30x (95°C for 45 s, 63°C for 45 s and 72°C for 25 s); 72°C for 5 min; PCRII, 95°C for 15 min; 30x (95°C for 45 s, 63°C for 45 s and 72°C for 10 s); 72°C for 5 min (Table S5). For deletions at Jκ4/5, primers p199/p205/p255 (PCRI) and p200/p206 (PCRII) were used with the same cycling conditions as for Jκ1/2 (Table S5). For deletions at Vκ3-1, primers p243/p244/p245 (PCRI) and p207/p210 (PCRII) were used with the

following conditions: PCRI, 95°C for 15 min; 30x (95°C for 45 s, 63°C for 45 s and 72°C for 50 s); 72°C for 5 min; PCRII, 95°C for 15 min; 30x (95°C for 45 s, 63°C for 45 s and 72°C for 20 s); 72°C for 5 min (Table S5). PCRII products were separated on 2% agarose gels stained with ethidium bromide. Fragments shorter than the expected size from the germline locus (Jκ1/2: <592 bp, Jκ4/5: <575 bp and Vκ3-1: <635 bp) were extracted (Macherey-Nagel) and sequenced (Genewiz). Deletion products were confirmed using Geneious (Kearse et al., 2012).

2.13. V(D)J PCR assays

Genomic DNAs of TC-Seq (RAG2^{core} and RAG2$^{-/-}$) cultures were used for V(D)J PCR assays. For RAG2^{full}, cells were cultured as for TC-Seq but infected with pMX-I-SceI-P2A-RAG2^{full}-EGFP. Duplicates originated from cell cultures with modified conditions: control was infected with pMX-EGFP; all cells were transferred onto irradiated S17 stroma cells after IL-7 washout on day 4 and depleted for 1.5 days. Semi-quantitative V(D)J PCRs were performed as previously described (Dudley et al., 2003; Schlissel et al., 1991) with modifications: 100, 50 or 25 ng of template DNA were amplified in 20µl reactions with HotStarTaq polymerase (Qiagen). For V(D)J PCRs, primers p58/p96 (Dh-Jh PCR), p96/p98 (VhQ52-DJh PCR) and p305/p306 (Vκ-Jκ PCR) were used with the following conditions: 95°C for 15 min; 32x (95°C for 45 s, 62°C for 45 s and 72°C for 2 min); 72°C for 5 min (Table S5). For control PCRs (*Myc*^{l}) primers p113/p114 were used with the following conditions: 95°C for 15 min; 30x (95°C for 45 s, 58°C for 45 s and 72°C for 20 s); 72°C for 5 min (Table S5). PCR products were separated on 1.5% agarose gels stained with ethidium bromide.

2.14. Accession numbers

The TC-Seq and IC-Seq sequencing data generated in this study can be accessed from the SRA database (SRP077983).

3. Results

3.1. Chromosomal rearrangements in pro-B cells

To examine RAG1/2-induced chromosomal rearrangements in developing B cells, I prepared TC-Seq libraries from cell cultures of primary murine pro-B cells deficient for RAG2 and harboring I-SceI sites at *c-myc* (*RAG2*$^{-/-}$*Myc*$^{I/I}$) that were infected with retroviruses expressing either I-SceI alone (RAG2$^{-/-}$ TC-Seq libraries) or I-SceI together with murine RAG2^{core} (RAG2^{core} TC-Seq libraries; Figures 12, S1 and see Materials and methods).

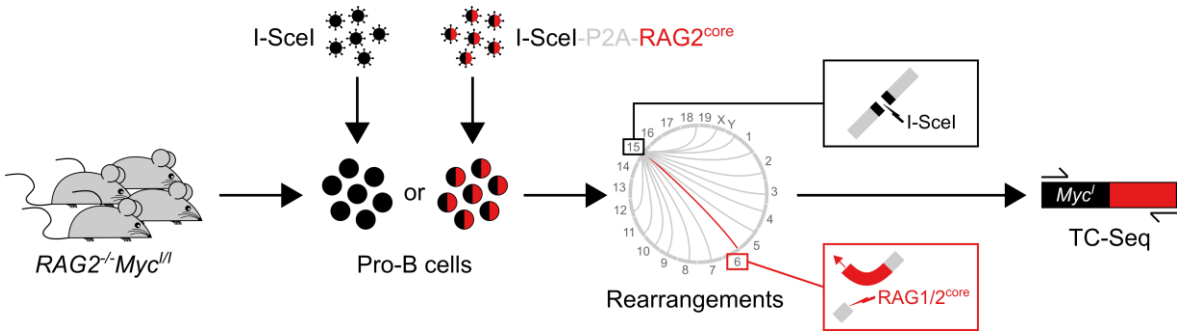

Figure 12: Detection of RAG1/2^{core}-induced chromosomal rearrangements by TC-Seq.

Primary *RAG2*$^{-/-}$*Myc*$^{I/I}$ pro-B cells are infected *ex vivo* with retroviruses that express either I-SceI alone (RAG2$^{-/-}$ TC-Seq libraries) or I-SceI together with murine RAG2^{core} (RAG2^{core} TC-Seq libraries) by using a "self-cleaving" P2A peptide. DNA breaks, such as those induced by RAG1/2^{core} at *Igκ* on chromosome 6 (red lightning), rearrange to the I-SceI break at *Myc*^{I} on chromosome 15 (black lightning) and are subsequently amplified by PCR, deep-sequenced and analyzed computationally. RAG2^{core} and RAG2$^{-/-}$ TC-Seq libraries were prepared in independent duplicates from infected pro-B cells of in total 180 mice.

RAG2^{core} was used since it promotes aberrant V(D)J recombination and causes genomic instability at *Tcr* loci in thymocytes (Deriano et al., 2011; Sekiguchi et al., 2001; Talukder et al., 2004; Curry and Schlissel, 2008). Moreover, mice expressing RAG2^{core} and deficient for either p53 alone or in combination with XRCC4-like factor (XLF) develop thymic or pro-B cell lymphomas, respectively, with translocations involving antigen receptor genes (Deriano et al., 2011; Lescale et al., 2016; Mijušković et al., 2015).

In agreement with previous TC-Seq studies in other cell types, chromosomal rearrangements in pro-B cells were especially abundant near the I-SceI cleavage site on chromosome 15 (Figures 13A and 13B; (Klein et al., 2011; Robbiani et al., 2015; Wang et al., 2014)). Moreover, rearrangements were enriched at genic regions, highly transcribed genes and early replication fragile sites (ERFSs), which define regions particularly susceptible to DNA damage during early replication (Figures 13C, 13D and 13E; (Barlow et al., 2013)).

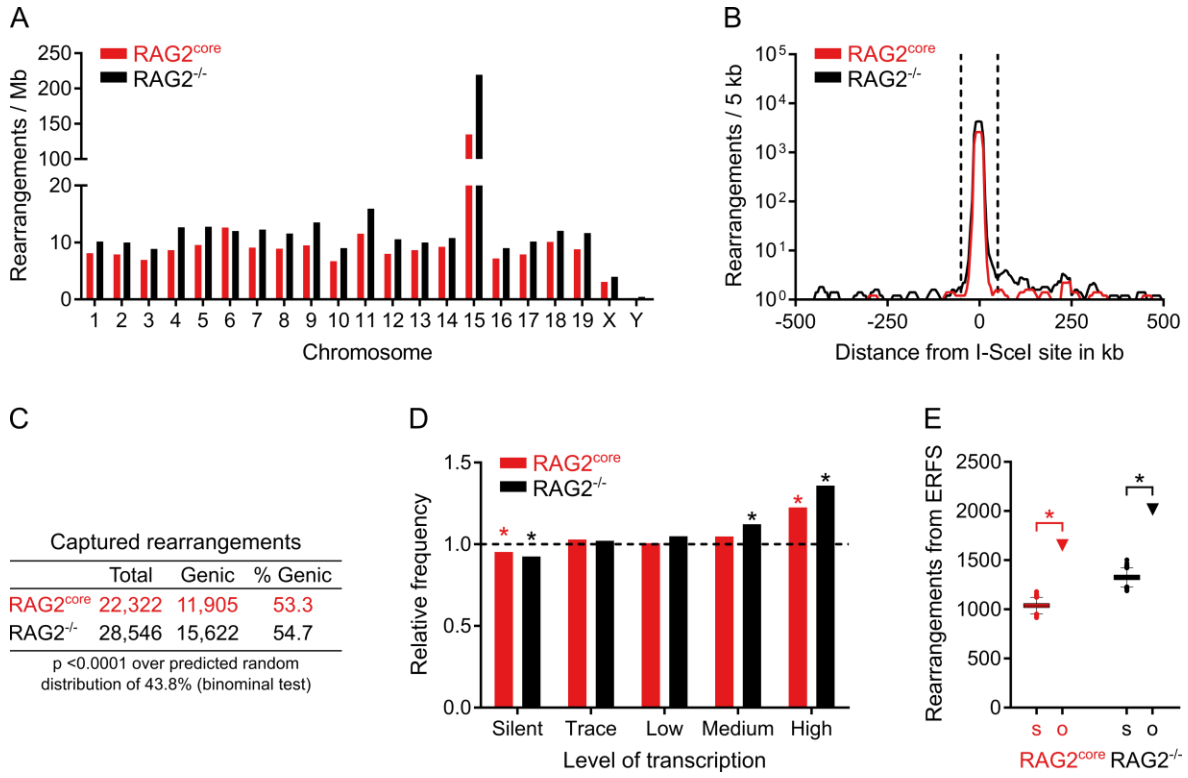

Figure 13: Landscape of chromosomal rearrangements in pro-B cells.

A- Chromosomal distribution of rearrangements. Events were normalized per Mb to account for different chromosome sizes.

B- Profile of rearrangements around the I-SceI site in 5 kb intervals. Dashed lines indicate the +/- 50 kb region excluded from the analysis for Figures 13C, 13D and 13E because of saturation.

C- Proportion of genic rearrangements.

D- Frequency of rearrangements derived from differentially transcribed genes compared to a random model (dashed line). Asterisks indicate values significantly different from random ($p < 0.01$, binominal test).

E- Observed number of rearrangements (o, triangle) originating from ERFS compared to the random Monte-Carlo simulation (s, boxplot). Asterisks indicate significant enrichment ($p < 0.0001$, binominal test).

For C to E, events from the saturated I-SceI region, cryptic I-SceI sites and other portions of the genome were excluded (see Materials and methods). Data analysis was performed with pooled RAG2^{core} and RAG2$^{-/-}$ TC-Seq libraries (2 independent experiments each).

3.2. DNA damage at physiologic and cryptic RSSs

To identify RAG1/2^{core}-dependent DNA damage, chromosomal rearrangements in RAG2^{core} and RAG2$^{-/-}$ TC-Seq libraries were compared. Briefly, genomic hotspots of rearrangement were identified, and those unique to RAG1/2^{core} were analyzed for the occurrence of breakpoint clusters (see Materials and methods). Overall, 33 RAG1/2^{core}-dependent rearrangement breakpoint clusters were detected throughout the genome (Table S1). In agreement with previous studies, limited recombination of the *Igh* locus by RAG1/2^{core} was observed and consequently only few disperse breakpoints were detected at Vh, Dh and Jh gene segments (Figure 14 and data not shown; (Akamatsu et al., 2003; Liang et al., 2002)).

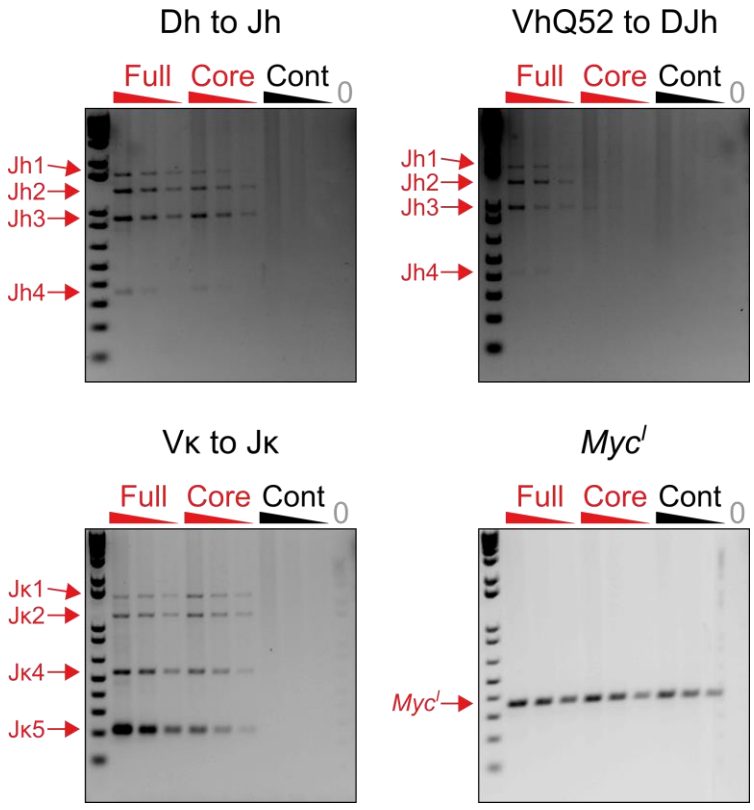

Figure 14: V(D)J recombination with RAG2-expressing retroviruses.

Ethidium bromide stained agarose gels with PCR amplicons from TC-Seq cultures infected with retroviruses expressing I-SceI-P2A-RAG2^{full} (Full), I-SceI-P2A-RAG2^{core} (Core) or I-SceI (Cont) and water control (0; see Figure S1 and Materials and methods). V(D)J recombinations at *Igh* and *Igκ* loci were amplified and PCR at *Myc^{l}* served as input control. PCRs were performed with serial dilutions of genomic DNA (20,000 cells/well, 10,000 cells/well and 5,000 cells/well; triangles, left to right). Red arrows point to V(D)J or *Myc^{l}* products. DNA ladder is shown alongside. All results were verified by at least 2 independent experiments.

In contrast, 24 RAG1/2^{core}-dependent breakpoint clusters were identified at *Igκ* (Figure 15 and Table S1).

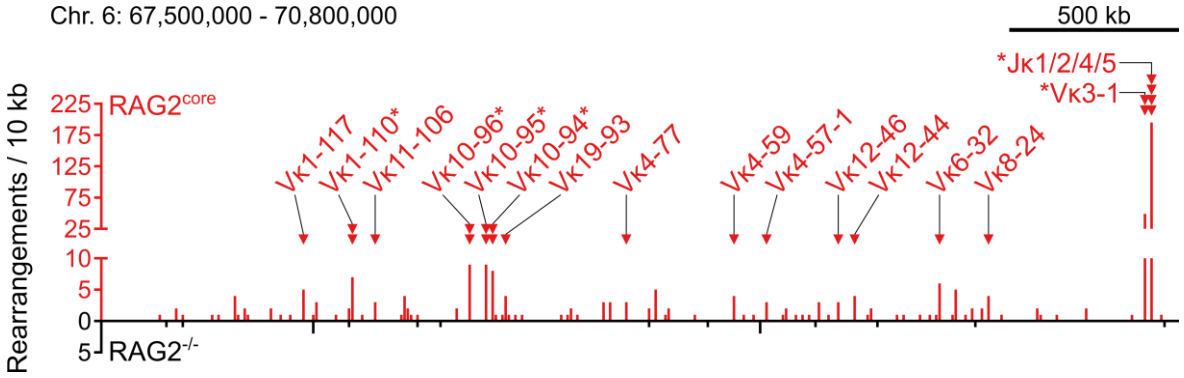

Figure 15: Overview of rearrangement breakpoints at the *Igκ* locus on chromosome 6.

Histogram of the number of breakpoints in the presence or absence of RAG2^{core} (red and black, respectively) in 10 kb intervals. RAG1/2^{core}-dependent rearrangement breakpoint clusters are indicated by red triangles and labeled with the corresponding Jκ or Vκ gene segment. Asterisks mark breakpoint clusters with biased rearrangements (see Figure 16). Chromosome coordinates and scale bar are indicated on top. Data analysis was performed with pooled RAG2^{core} and RAG2$^{-/-}$ TC-Seq libraries (2 independent experiments each).

Each functional Jκ segment (Jκ1, Jκ2, Jκ4, Jκ5) had a single cluster at its 23RSS cleavage site (Figure 16A). Surprisingly, DNA at these clusters recombined with the I-SceI break in a biased manner. Although in principle both DNA ends of a RAG1/2^{core}-induced break would have an equal probability of joining to the cleaved I-SceI site, most rearrangements occurred with only one of the two ends for any RSS. For example, rearrangements between the I-SceI break and RAG1/2^{core} breaks at Jκ1 exclusively involved the coding end (Figure 16A, rearrangements in grey), while those at the neighboring Jκ2 predominantly (86%) contained the signal end (Figure 16A, rearrangements in green). Moreover, rearrangements at Jκ1 did not extended beyond the 23RSS cleavage site of Jκ2, and *vice versa*. A similar phenomenon was observed for Jκ4/Jκ5 (Figure 16A).

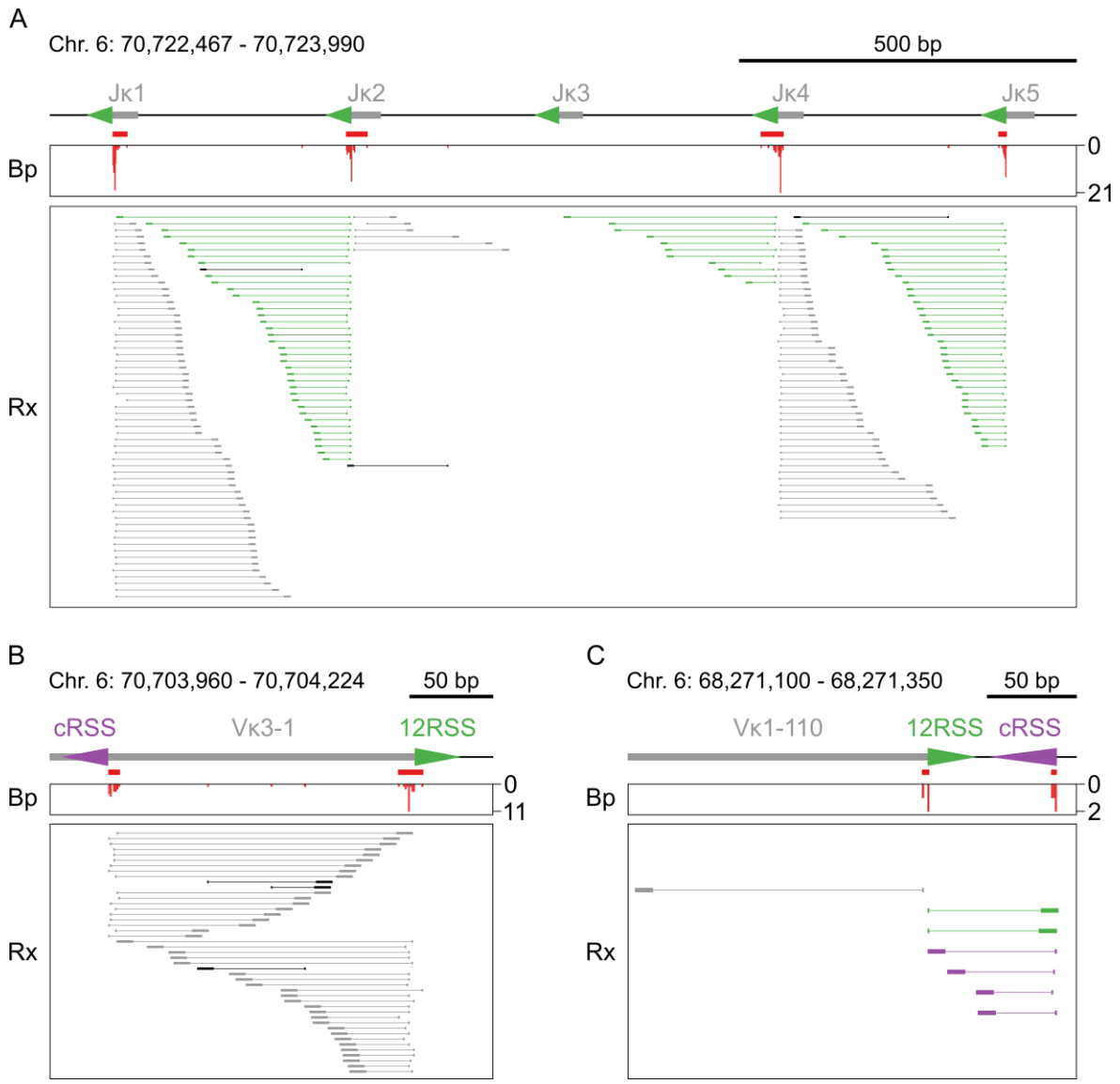

Figure 16: RAG1/2^{core}-dependent breakpoint clusters at Jκs and Vκs.

A to C- On top is a diagram of the region, with grey boxes representing *Ig* segments, triangles indicating 12/23RSSs (green) or cRSSs (purple), and red bars indicating the breakpoint clusters. In the middle, histogram showing the number and position of breakpoints (Bp, red). At the bottom, each horizontal line indicates a unique rearrangement (Rx), with its breakpoint represented by the vertical line, and its sheared end (which determines the uniqueness of the event) shown by the box. Color-coding indicates whether rearrangements contain RSSs/cRSSs (green/purple, signal ends) or not (grey, coding ends). Rearrangements in black are undefined. Chromosome coordinates and scale bar are indicated on top. Data analysis was performed with pooled RAG2^{core} and RAG2$^{-/-}$ TC-Seq libraries (2 independent experiments each).

In addition to Jκs, breakpoint clusters were also found at 15 Vκ gene segments. Strikingly, while 10 of these had a single cluster at their physiologic 12RSS cleavage sites, the other 5 (Vκ3-1, Vκ10-94, Vκ10-95, Vκ10-96 and Vκ1-110) revealed an additional cluster at a nearby cRSS (Figure 15 and Table S1). Overall, the heptamer sequences of these cRSSs were similar to the physiologic consensus and to those identified in previous studies (Figure 17A; (Hu et al., 2015)). However, none of the cRSSs were detectable by computational tools because of their low RSS information content (RIC) scores (Table S2; (Cowell et al., 2002; Merelli et al., 2010)). Similar to the biased recombination pattern observed at Jκs, Vκ rearrangements at neighboring 12RSS/cRSS clusters were biased for coding or signal ends and limited in length by both cleavage sites (Figures 16B and 16C).

The remaining breakpoint clusters (9) mapped to off-target regions outside of *Ig* loci (Table S1). Off targets were preferentially in transcribed genes (6) but not enriched in histone H3 lysine-4 trimethylation (H3K4me3), an active chromatin mark (data not shown). Off target clusters occurred near cRSS motifs that were similar to those identified at Vκ segments and also undetectable by computational tools (Figures 17A, 17B, 17C and Table S2).

I conclude that RAG1/2^{core} damages the B cell genome at physiologic and cryptic RSSs, and that some of the resulting DNA breaks at Jκs and Vκs recombine with the cleaved I-SceI site in a biased manner.

A

Heptamer sequences	
RSS	C A C A G T G
Cluster 52a	C A C A C A T
Vκ1-110	C A C A T A A
Cluster 642	C A C A T C C
Cluster 840	C A C A T C C
Vκ10-94	C A C T G C C
Vκ10-95	C A C T G C C
Vκ10-96	C A C T G C C
Vκ3-1	C A C T G C A
Cluster 373	C A C C A T G
Cluster 339b	C A C C T T T
Cluster 317	C A C C A C A
Cluster 299	C A G A G A T
Cluster 795	C A G A C T C
Cluster 766b	C A G A C A G

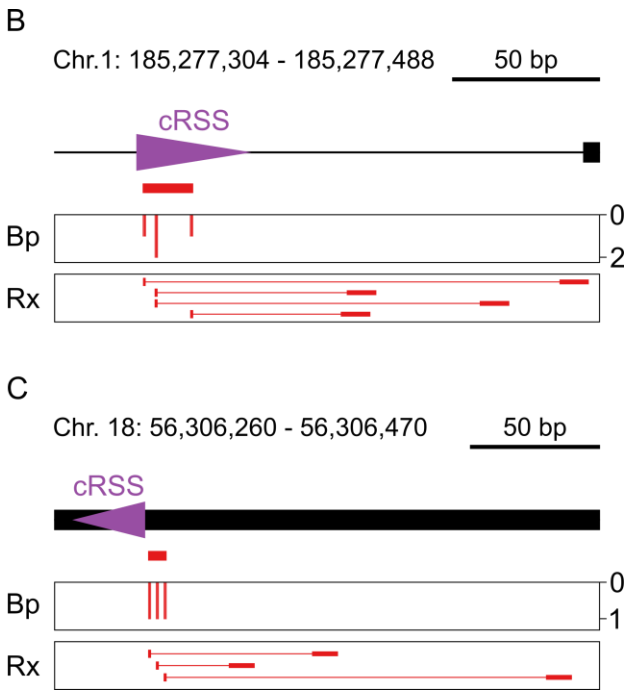

Figure 17: RAG1/2^{core}-dependent breakpoint clusters at cRSSs.

A- Heptamer sequences of cRSSs identified at RAG1/2^{core}-dependent breakpoint clusters. The canonical RSS heptamer is shown on top (green). Below are cRSS heptamer sequences at breakpoint clusters located at off targets and Vκ segments. Green nucleotides are shared with the canonical heptamer. Underlined nucleotides are identical to those of previously identified cRSS heptamers (Hu et al., 2015).

B and C- RAG1/2^{core}-dependent off-target breakpoint cluster 52a and 373, respectively. On top is a diagram of the region, with black boxes representing long terminal repeats, triangles indicating cRSSs (purple) and red bars indicating the breakpoint clusters. In the middle, histogram showing the number and position of breakpoints (Bp, red). At the bottom, each horizontal line indicates a unique rearrangement (Rx, red), with its breakpoint represented by the vertical line, and its sheared end (which determines the uniqueness of the event) shown by the box. Chromosome coordinates and scale bar are indicated on top. Cluster 52a is located within an intron of *Rab3gap2* and cluster 373 is intergenic.

Data analysis was performed with pooled RAG2^{core} and RAG2$^{-/-}$ TC-Seq libraries (2 independent experiments each).

3.3. Aberrant deletions at *Igκ*

The peculiar rearrangement pattern observed at Jκ and some of the Vκ clusters suggested that RAG1/2^{core} may mediate aberrant deletions by recombining neighboring RSSs and cRSSs at these sites. To examine this possibility, I developed high-sensitivity PCR assays based on the "poison primer" principle and searched for small aberrant V(D)J deletions (see Materials and methods; (Edgley et al., 2002)).

Strikingly, aberrant deletions mediated by either RAG1/2^{core} or endogenous wild type RAG1/2 were readily detected at Jκs, where the RSSs at Jκ1 and Jκ4 were joined to the neighboring Jκ2 and Jκ5 exons, respectively (Figures 18A and 18B). The resulting deletion junctions represent aberrant hybrid joints, which are defined as junctions formed by the joining of a RSS to its reactions partner's coding flank (Helmink and Sleckman, 2012). Moreover, the observed deletions occurred between two 23RSSs and thus violate the 12/23-rule of V(D)J recombination (Helmink and Sleckman, 2012).

In addition to those at Jκs, deletions mediated by either RAG1/2^{core} or RAG1/2 wild type were also identified at Vκ3-1, where joining of the 12RSS to the nearby cRSS described above generated aberrant signal joints (Figure 18C). I conclude that both RAG1/2^{core} and RAG1/2 wild type cause aberrant genomic deletions at Jκ and Vκ segments.

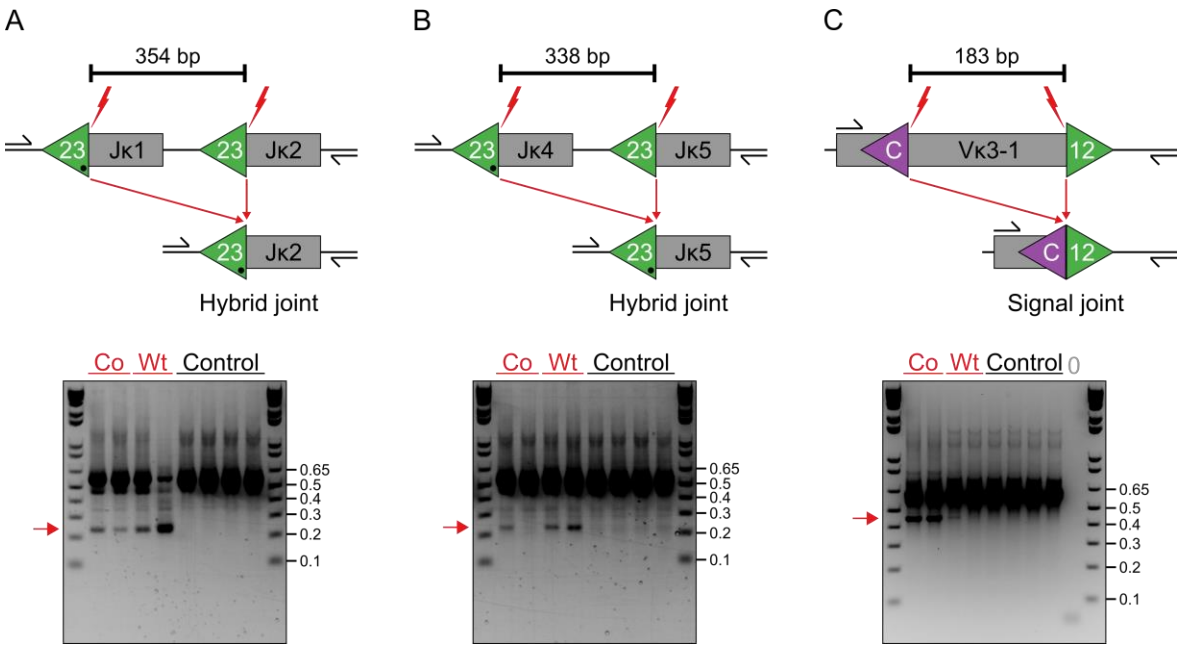

Figure 18: Aberrant deletions at *Igκ* mediated by RAG1/2^{core} and RAG1/2 wild type.

A to C- Deletions between neighboring RSSs or RSS/cRSS at Jκs and Vκs, respectively. Top, diagram of the locus before and after deletion by RAG1/2^{core} or RAG1/2 wild type. The predicted size of the deletion is shown above. Grey boxes represent *Ig* segments, triangles indicate 12/23RSSs (green) or cRSSs (purple), red lightning points to RAG1/2 cleavage sites and black arrows indicate the location of the internal primers used for the deletion PCR (see Materials and methods). Bottom, ethidium bromide stained agarose gel with deletion PCR amplicons from cultured RAG1/2^{core} (Co), RAG1/2 wild type (Wt) or RAG2$^{-/-}$ (Control) bone marrow B cells and water control (0). Red arrows point to amplified deletion junctions. DNA ladder is shown alongside. Deletions were captured from 20,000 cells/well (A and B) or 40,000 cells/well (C). Selected amplicons were extracted and confirmed by sequencing. The frequency of aberrant deletions was subsequently determined by dilutional PCR. For Co: 1 in 4,000 cells in (A), 1 in 20,000 cells in (B), 1 in 600 cells in (C). For Wt: 1 in 3,600 cells in (A), 1 in 2,700 cells in (B), 1 in 39,300 cells in (C). Frequency of deletions by RAG1/2^{core} were lower in the repeat experiment (see Materials and methods). All results were verified by at least 2 independent experiments.

3.4. Excised *Igκ* fragments insert into I-SceI breaks

Based on the co-localization of biased rearrangements and aberrant deletions, I hypothesized that Jκ/Vκ fragments might be aberrantly excised by RAG1/2^{core} and subsequently re-integrate at the I-SceI break (Figure 19 and see Discussion). To test this hypothesis, TC-Seq libraries were computationally screened for bona fide insertions, which would have been excluded from the initial bioinformatic analysis geared at identifying translocations. Briefly, insertions at the I-SceI site are flanked by *Myc^{I}* sequence on both ends, whereas translocations contain *Myc^{I}* sequence only on one end (Figure 19). Thus, all sequences with *Myc^{I}* on both ends were examined for intervening DNA originating from elsewhere in the genome (see Materials and methods).

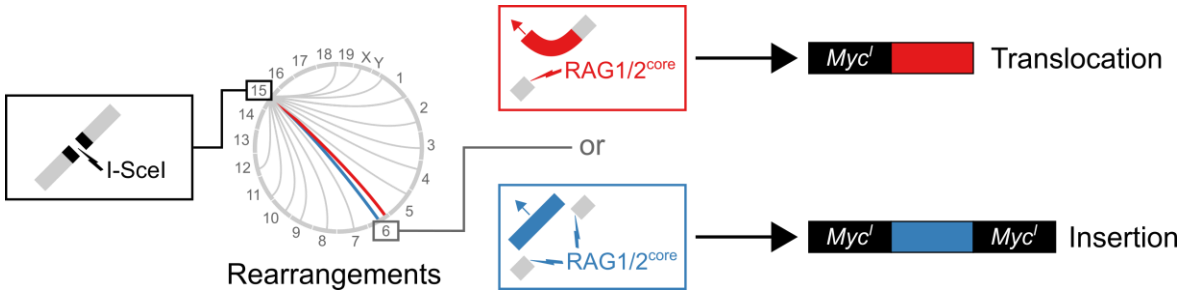

Figure 19: Cartoon diagram comparing RAG1/2^{core}-induced translocations and insertions.

In a translocation (red), RAG1/2^{core} introduces a single DNA break (red lightning) that recombines with the cleaved I-SceI site at *Myc^{I}* (black lightning) on chromosome 15. The resulting translocation contains *Myc^{I}* only on one side. In an insertion (blue), RAG1/2^{core} causes tandem DNA breaks (blue lightning) thereby excising a DNA fragment that subsequently re-integrates into the cleaved I-SceI site. The resulting insertion is flanked by *Myc^{I}* on both sides.

I-SceI insertions were detected in both RAG2^{core} and RAG2$^{-/-}$ TC-Seq libraries. Independent of RAG2^{core}-expression, inserted DNA fragments originated predominantly from a +/- 20kb region around the I-SceI cleavage site on chromosome 15, similar to the chromosomal rearrangements described above (Figures 13A, 13B, 20A and 20B). Overall, inserted DNA fragments ranged from 36 to 354 bp in RAG2^{core} and from 36 to 232 bp in RAG2$^{-/-}$ cells (36 bp being the lowest detection limit, see Materials and methods). Moreover, genic regions acted as preferred donors for insertions, particularly in RAG2^{core}-expressing cells (Figure 20C). In contrast, insertions originating from highly transcribed regions and ERFS were significantly enriched only in the absence of RAG2^{core}, indicating that its expression alters the insertion landscape (Figures 20D and 20E).

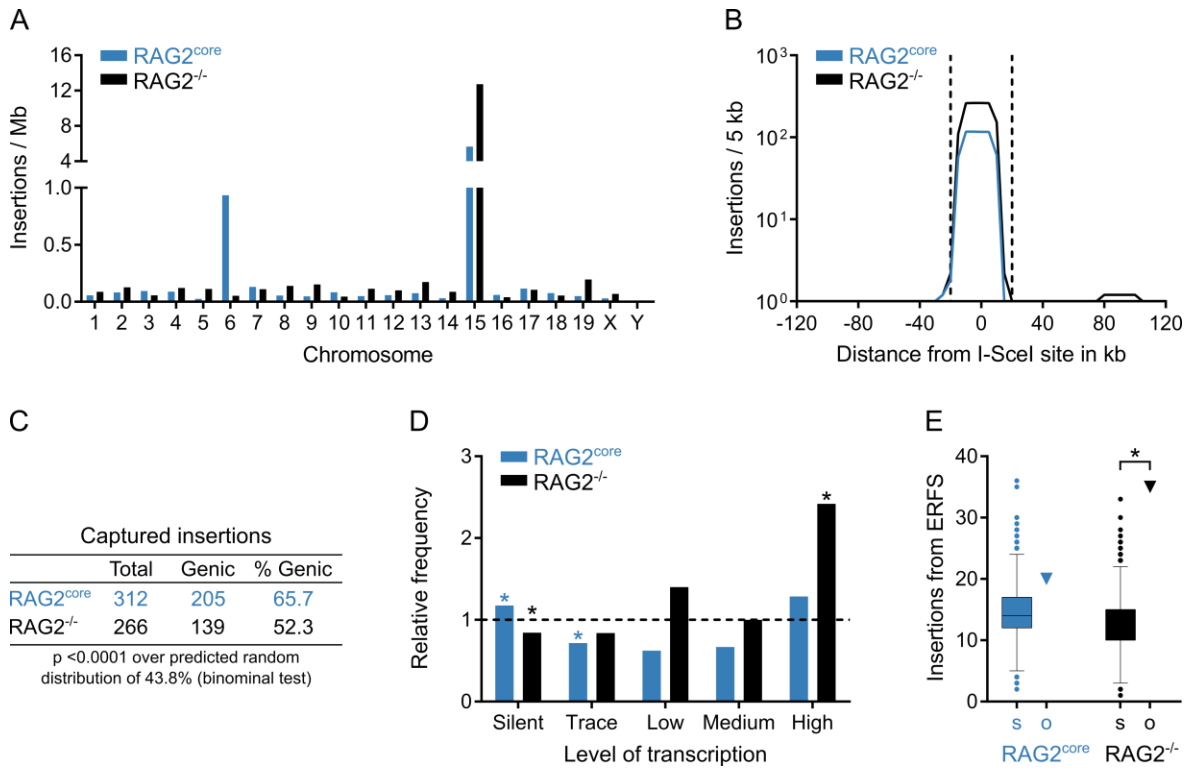

Figure 20: Landscape of insertions in primary pro-B cells by TC-Seq.

A- Origin of insertions by chromosome. Events were normalized per Mb to account for different chromosome sizes.

B- Profile of insertions near the I-SceI site in 5 kb intervals. Dashed lines indicate the +/- 20 kb region excluded from the analysis for Figures 20C, 20D and 20E because of saturation.

C- Proportion of insertions from genic regions.

D- Frequency of insertions derived from differentially transcribed genes compared to a random model (dashed line). Asterisks indicate values significantly different from random ($p < 0.01$, binominal test).

E- Observed number of insertions (o, triangle) originating from ERFS compared to the random Monte-Carlo simulation (s, boxplot). Asterisks indicate significant enrichment ($p < 0.0001$, binominal test).

For C to E, events from the saturated I-SceI region, cryptic I-SceI sites and other portions of the genome were excluded (see Materials and methods). Data analysis was performed with pooled RAG2^{core} and RAG2$^{-/-}$ TC-Seq libraries (2 independent experiments each).

In agreement with its influence on the insertion landscape, expression of RAG2^{core} correlated with a higher amount of insertions from chromosome 6 compared to RAG2$^{-/-}$ cells (140 vs. 8 events; Figure 20A), and nearly all of those (96%) originated from *Igκ*, while none derived from this locus in RAG2$^{-/-}$ cells (Figure 21 and Table S1).

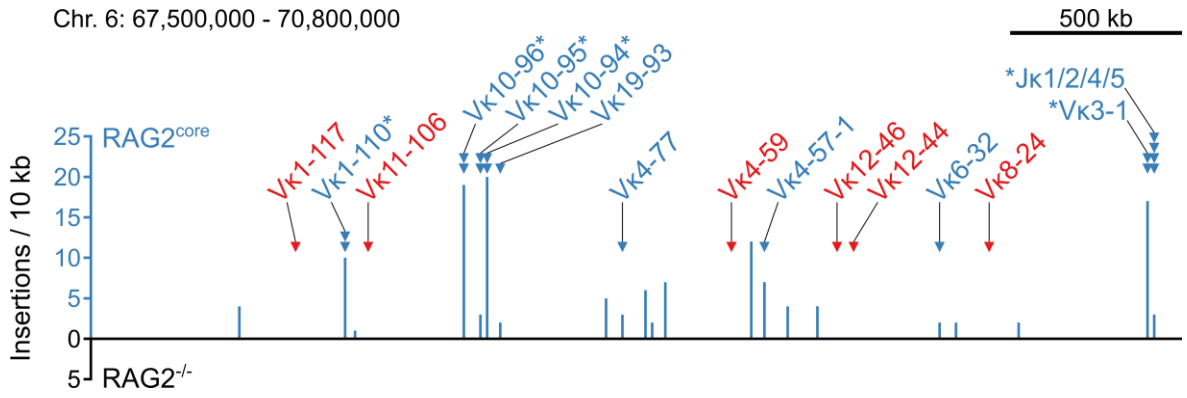

Figure 21: Overview of insertions originating from the *Igκ* locus on chromosome 6.

Histogram of the number of insertions derived from each site in the presence or absence of RAG2^{core} (blue and black, respectively) in 10 kb intervals. RAG1/2^{core}-dependent rearrangement breakpoint clusters at Jκs and Vκs (triangles, same as in Figure 15) are color-coded to indicate whether insertions from these sites are detected (blue) or not (red). Asterisks mark breakpoint clusters with biased rearrangements (see Figure 16). No insertions from *Igκ* were detected in RAG2$^{-/-}$ cells. Chromosome coordinates and scale bar are indicated on top.

Overall, *Igκ* insertions represented nearly half (43%) of all insertions in RAG2^{core} cells and exclusively originated from regions flanked by RSSs and/or cRSSs (Figures 22A, 22B and 22C). Interestingly, donor regions included all of the *Igκ* gene segments displaying biased breakpoint clusters, suggesting that DNA insertions from these sites are responsible for the observed recombination pattern (Figure 21, compare Figures 16 and 22 and see Discussion).

For 67% of *Igκ* insertions sequence information on both junctions was obtained, providing insight into the original deletion events (Table S3). Overall, *Igκ* insertions originated from DNA excision between pairs of divergent, convergent or head-to-tail RSSs, leading to insertions flanked by coding ends (coding-end insertions, 77), signal ends (signal-end insertions, 8) or both (hybrid-end insertions, 6), respectively (Figures 22A, 22B, 22C and Table S3). Most deletions (87 out of 91) occurred between RSS/cRSS pairs, three resulted from excisions between two cRSSs, and one derived from a deletion between two 23RSSs. I conclude that RAG1/2^{core} generates aberrant *Ig* fragments that are mobile and can be re-inserted into I-SceI breaks on a heterologous chromosome.

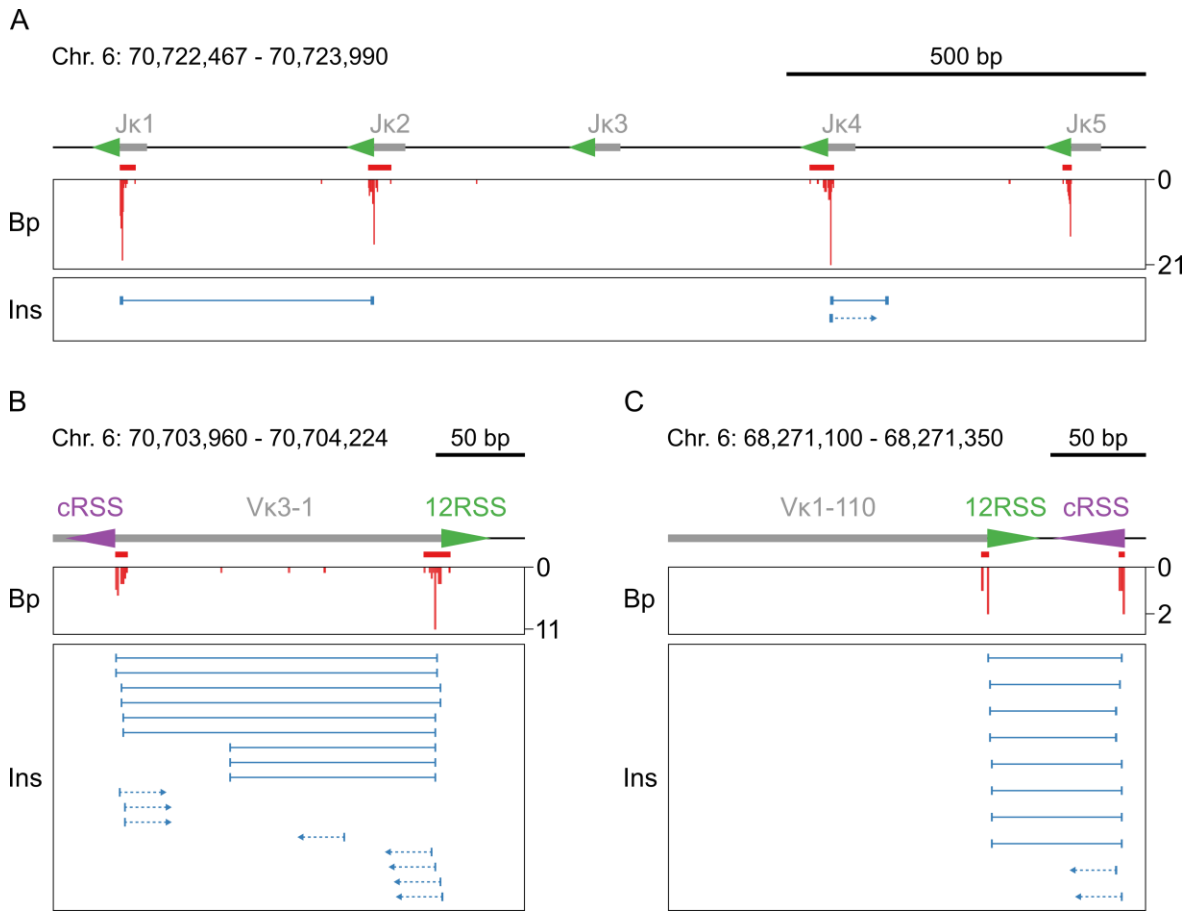

Figure 22: Insertions derived from RAG1/2^{core}-dependent breakpoint clusters at Jκs and Vκs.

A to C- On top is a diagram of the region, with grey boxes representing *Ig* segments, triangles indicating 12/23RSSs (green) or cRSSs (purple), and red bars indicating the rearrangement breakpoint clusters (same as in Figure 16). In the middle, histogram showing the number and position of breakpoints (Bp, red). At the bottom, each horizontal line indicates a unique insertion (Ins), with its breakpoints represented by the vertical lines at the ends. Arrows represent insertions for which only one of the two breakpoints could be identified. Chromosome coordinates and scale bar are indicated on top. Data analysis was performed with pooled RAG2^{core} and RAG2$^{-/-}$ TC-Seq libraries (2 independent experiments each).

3.5. Insertion of *Igκ* fragments excised by wild type RAG1/2

As demonstrated by my deletion PCR assays, RAG1/2 can produce aberrant *Igκ* deletions analogous to RAG1/2^{core}. Thus, mobilization and insertion of *Igκ* DNA could in principle also occur in wild type B cells. To test this possibility, I developed a next-generation insertion capture and sequencing method (IC-Seq), that qualitatively documents chromosomal insertions at an I-SceI site under physiologic conditions. IC-Seq libraries were prepared from primary bone marrow B cells expressing a Tamoxifen-inducible I-SceI transgene and bearing I-SceI cleavage sites (*ROSA*^{erISCEI}*Myc*$^{I/I}$*Igh*$^{I/I}$ and *ROSA*^{erISCEI}*Myc*$^{I/I}$*Igh*$^{I/I}$*AID*$^{-/-}$, see Materials and methods; (Robbiani et al., 2015)) that were treated *ex vivo* with Tamoxifen to induce I-SceI breaks in the presence of wild type RAG1/2. DNA insertions at the I-SceI site in *Myc*^{I} were amplified by PCR, deep-sequenced, and analyzed computationally (Figure 23 and see Materials and methods).

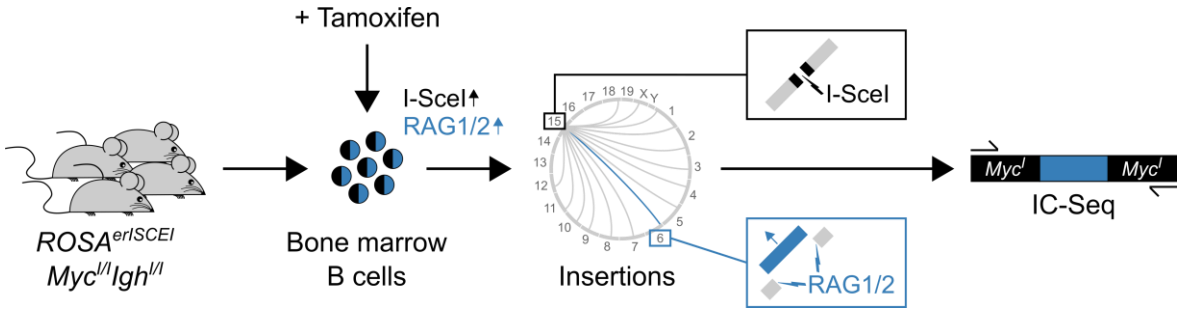

Figure 23: Detection of chromosomal insertions by IC-Seq.

ROSA^{erISCEI}*Myc*$^{I/I}$*Igh*$^{I/I}$ (and *ROSA*^{erISCEI}*Myc*$^{I/I}$*Igh*$^{I/I}$*AID*$^{-/-}$, see Materials and methods) bone marrow B cells are treated *ex vivo* with Tamoxifen to induce I-SceI breaks at *Myc*^{I} on chromosome 15 (black lightning). Mobilized DNA fragments, such as those excised by endogenous RAG1/2 from *Igκ* on chromosome 6 (blue lightning), insert into the cleaved I-SceI site and are subsequently amplified by PCR, deep-sequenced and analyzed computationally. Two RAG1/2 wild type IC-Seq libraries were independently prepared from Tamoxifen-treated bone marrow B cells of in total 12 mice.

Overall, I-SceI insertions from 7 different *Igκ* gene segments were detected (Jκ1, Jκ2, Jκ4, Jκ5, Vκ1-110, Vκ3-1 and Vκ4-69), of which 6 were also involved in the above described insertions mediated by RAG1/2^{core} (Table S3). Moreover, similar to RAG1/2^{core}, *Igκ* insertions in the presence of RAG1/2 originated exclusively from donor regions flanked by RSSs/and or cRSSs and were comprised of coding-, signal- and hybrid-end insertions (Figures 24A, 24B, 24C and Table

S3). I conclude that DNA insertions from *Igκ* are not limited to RAG1/2^{core} but also occur during physiologic V(D)J recombination by wild type RAG1/2.

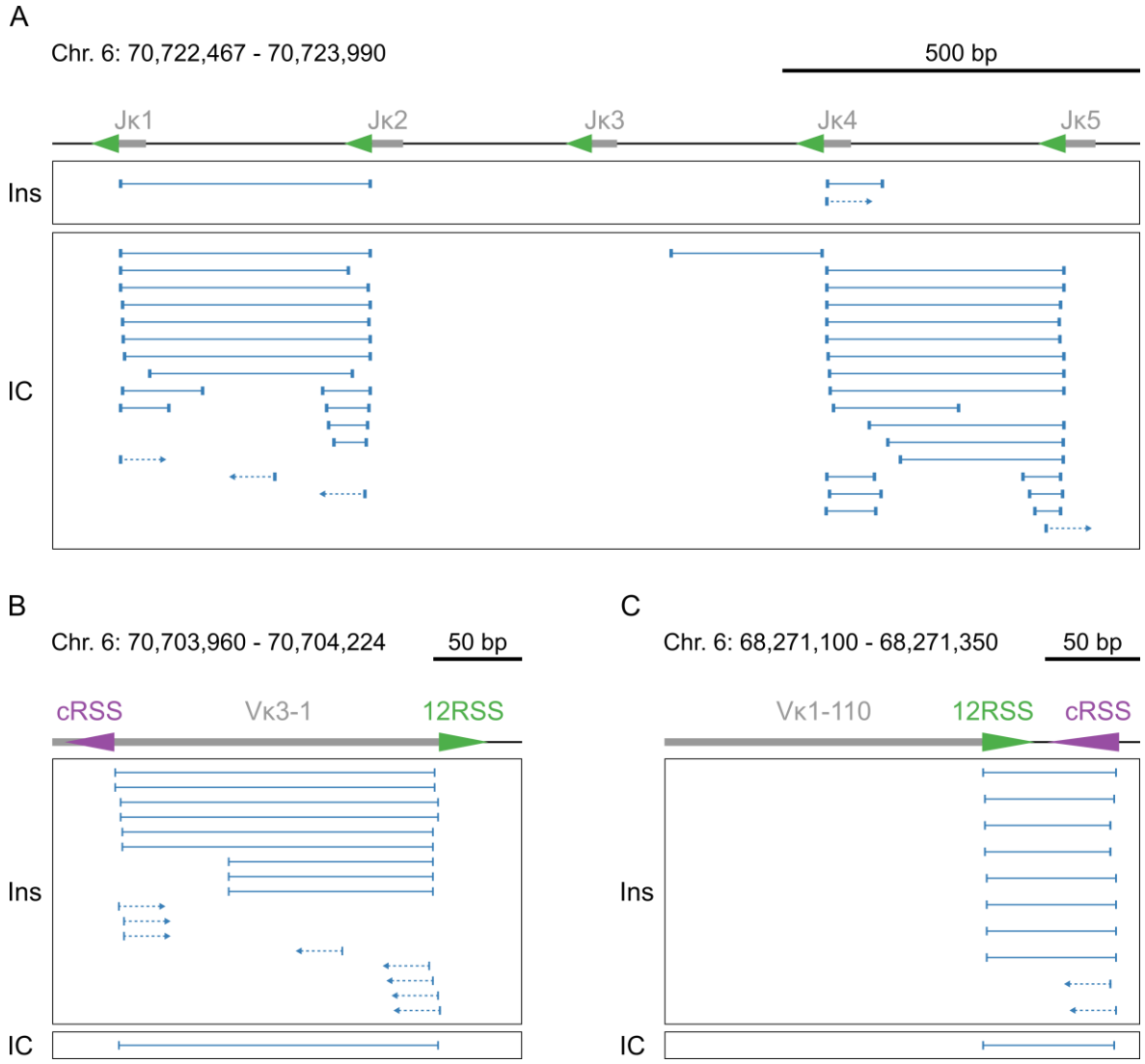

Figure 24: Qualitative comparison of insertions obtained by TC-Seq (RAG1/2^{core}) and IC-Seq (RAG1/2 wild type).

A to C- On top is a diagram of the region, with grey boxes representing *Ig* segments and triangles indicating 12/23RSSs (green) or cRSSs (purple). Below, insertions detected by TC-Seq (Ins, same as in Figure 22) and IC-Seq (IC). Each horizontal line indicates a unique insertion, with its breakpoints represented by the vertical lines at the ends. Arrows represent insertions for which only one of the two breakpoints could be identified. Chromosome coordinates and scale bar are indicated on top. Data analysis was performed with pooled IC-Seq libraries (two independent experiments).

3.6. Insertion of *IG* and *TCR* fragments at physiologic DNA breaks

To determine whether RAG1/2 causes insertions at physiologic DNA breaks *in vivo*, published whole genome sequences from ALL and FL patients were computationally screened for insertions deriving from *IG* and *TCR* loci (see Materials and methods). Overall, 5 out of 34 patients displayed genomic insertions of *IG* or *TCR* fragments at low frequency (Table S4). All insertions contained at least one RSS or cRSS motif and integrated near repetitive regions (Figures 25A, S2 and S3). Interestingly, DNA flanking one of the inserts was inverted to form a putative cRSS/cRSS signal joint and in another case a *TCR* fragment inserted at a translocation junction (Figures 25A and 25B). I conclude that RAG1/2 has the potential to destabilize the lymphocyte genome by mobilizing DNA that then re-inserts at RAG1/2-independent, physiologic DNA breaks *in vivo*.

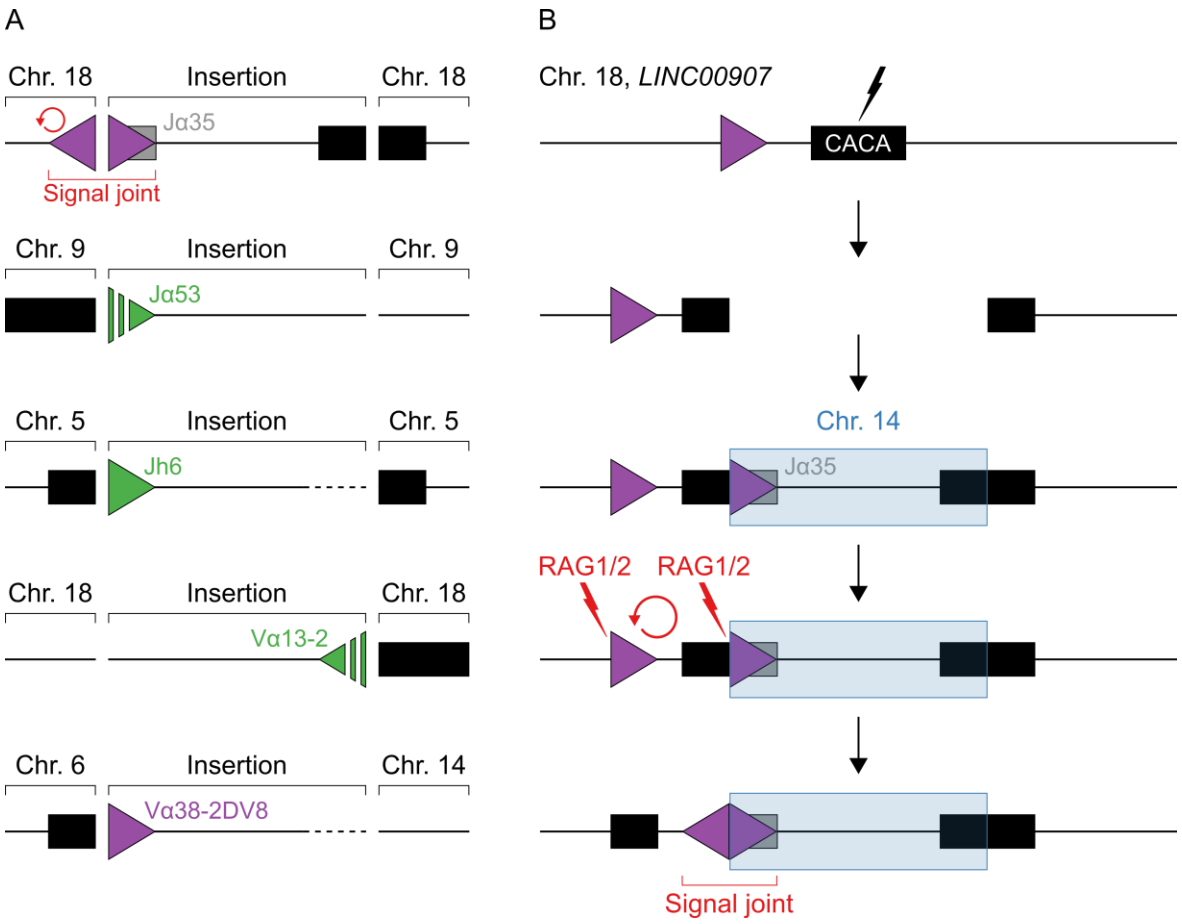

Figure 25: RAG1/2-induced insertions at physiologic DNA breaks *in vivo*.

A- Diagram representing *IG/TCR* insertions detected in human cancer. From top to bottom: hypodiploid ALL (first), early T-cell precursor ALL (second and third) and FL (fourth and fifth). Boxes indicate *IG*/*TCR* segments (grey) or repeat regions (black) and triangles represent 12/23RSSs (green) or cRSSs (purple). Inserted gene segments and RSSs/cRSSs are labeled with their corresponding *IG*/*TCR* segment of origin. Unresolved junctions are indicated by dashed lines/triangles. The insertion detected in hypodiploid ALL (top) is flanked by an upstream inversion (red arrow), forming a putative cRSS/cRSS signal joint. One of the insertions in FL (bottom) occurred at a translocation junction. Whole genome sequences from 34 cancer patients were analyzed.

B- Diagram illustrating a putative RAG1/2-mediated DNA inversion caused by the insertion of a cRSS. Boxes represent *IG* segments (grey) or repeat regions (black), triangles indicate cRSSs (purple), lightning indicates DNA cleavage induced by RAG1/2 (red) or unknown factors (black). From top to bottom: First, DNA is damaged at a simple CA-repeat region within *LINC00907* on chromosome 18. Second, the locus is opened at the break. Third, RAG1/2 excises a DNA fragment containing a cRSS from the *TRA* locus on chromosome 14 which subsequently re-inserts into the break (blue). Fourth, RAG1/2 cleaves and inverts DNA between the cRSS in the insert and a cRSS near the insertion site (red arrow). Finally, the DNA inversion generates a putative cRSS/cRSS signal joint as observed in Figure 25A (top).

4. Discussion

4.1. RAG1/2 damages the pro-B cell genome at physiologic and cryptic RSSs

I used TC-Seq to examine chromosomal rearrangements in the pro-B cell genome and identified 33 RAG1/2^{core}-dependent breakpoint clusters, of which 19 occurred at physiologic RSS cleavage sites. Consistent with this finding, a previous study in ATM-deficient pro-B cell lines detected chromosomal rearrangements between I-SceI breaks at *c-myc* and RAG1/2-induced breaks at antigen receptor loci including *Igκ* (Zhang et al., 2012). Interestingly, off-target clusters at cRSSs were not detected in those experiments. In contrast, 14 of the 33 RAG1/2^{core}-dependent breakpoint clusters identified herein were located near cRSS motifs at Vκs and off-target regions outside *Ig* loci. Off targets were not enriched in H3K4me3, an active chromatin mark that has been shown to co-localize with RAG1/2 binding and cleavage in developing B cells (data not shown, (Hu et al., 2015; Ji et al., 2010; Teng et al., 2015)). Its absence might result from the fact that RAG2^{core} lacks the C-terminal plant homeo domain, which normally mediates RAG1/2 binding to H3K4me3 (Liu et al., 2007; Matthews et al., 2007; Ramón-Maiques et al., 2007; West et al., 2005).

My results demonstrate that RAG1/2^{core}-mediated cleavage of cRSSs enables chromosomal rearrangements by producing cleaved ends that can recombine with RAG1/2-independent DNA breaks. Moreover, my data confirms that neighboring RSSs and cRSSs are substrates for aberrant genomic deletions, in agreement with previous studies using engineered RSSs (Hu et al., 2015; Mahowald et al., 2009). I speculate that the cRSSs at Vκs identified herein might also serve as beneficial substrates for secondary V-J rearrangements during V-gene replacement, similar to those described at Vhs (Rahman et al., 2006).

4.2. Aberrantly excised *Igκ* DNA re-inserts at I-SceI breaks

I hypothesized that some of the observed rearrangements resembling translocations may actually represent insertions of deleted DNA into the I-SceI break. Because DNA is sonicated during preparation of TC-Seq libraries, a fraction of insertions would be randomly truncated and appear as translocations in the analysis (see comparison between translocation and insertion in Figure 19). In agreement with this prediction, bona fide insertions originating from all RAG1/2^{core} breakpoint clusters with biased rearrangements were identified. Furthermore, by using a novel next-generation sequencing method (IC-Seq) I confirmed *Igκ* insertions at I-SceI breaks in the presence of wild type RAG1/2. *Igκ* insertions mediated by RAG1/2^{core} and RAG1/2 wild type were similar in that they both originated from donor regions with RSSs/cRSSs and were comprised of all three insertion species (signal-end, coding-end, and hybrid-end).

Overall, insertions detected by both TC-Seq and IC-Seq were short (354 bp or less). The absence of larger insertions is likely due to technical limitations. During TC-Seq, which was originally designed to detect chromosomal translocations, the size of insertions is mainly limited by the sonication of genomic DNA (see Materials and methods). I therefore expect long insertions to be truncated and appear as “translocations” in the computational analysis. In this regard, some of the apparent translocations at Vκs and Jκs could in principle result from the insertion of large physiologic excision fragments (10s-100s kb). Moreover, since even short insertions can be truncated, it is possible that TC-Seq considerably underestimates the actual frequency of insertions. During IC-Seq, which omits DNA sonication, the major factor limiting the detection of large insertions is PCR amplification. DNA templates with large insertions are likely outcompeted by those with small or no insertions. Finally, both TC-Seq and IC-Seq utilize high throughput sequencing (see Materials and methods), which is inefficient for DNA fragments above 1.5 kb.

4.3. *Igκ* insertions at I-SceI breaks are not mediated by DNA transposition or trans-V(D)J recombination

Three distinct insertion species from *Igκ* were observed in this study: those flanked by RSS/cRSS pairs (signal-end insertions), those lacking RSSs altogether (coding-end insertions), and those bearing only one RSS or cRSS (hybrid-end insertions).

Signal-end insertions derive from DNA deletion between convergent RSSs, which are normally joined to form episomal signal joints. There is some *in vivo* evidence that RAG1/2 can induce genomic insertions by re-cleaving and subsequently re-integrating episomal signal joints through either trans-V(D)J recombination or DNA transposition (Curry et al., 2007; Messier et al., 2003; Vanura et al., 2007). However, the observed signal-end insertions are not compatible with these two pathways because they occur at RAG1/2-independent DNA breaks generated by I-SceI. In contrast, during trans-V(D)J recombination it is RAG1/2 that cleaves the RSS/cRSS at the insertion site, and in DNA transposition RAG1/2 is responsible for catalyzing the nucleophilic attack required for insertion. Thus, the RAG1/2-induced signal-end insertions observed in my study are mediated by a pathway distinct from these previously described mechanisms.

Coding-end insertions do not fit previously proposed RAG1/2 insertion mechanisms either, since both trans-V(D)J recombination and DNA transposition require RSSs-containing donor fragments (Agrawal et al., 1998; Curry et al., 2007; Hiom et al., 1998; Vanura et al., 2007). Coding-end insertions originate from DNA deletions between divergent RSSs, whose products are predicted to circularize into episomal coding joints. Since these cannot be re-cleaved by RAG1/2, coding-end insertions likely originate from non-circularized, linear deletion products.

Hybrid-end insertions derive from deletions between head-to-tail RSSs. In principle, such deletions produce episomal hybrid joints that contain a single RSS or cRSS. Although *in vitro* assays have shown that RAG1/2 can induce breaks at single RSSs, the extent to which this occurs *in vivo* is unclear (Eastman and Schatz, 1997; McBlane et al., 1995; Rahman et al., 2006; Yu and

Lieber, 2000). Hence, similar to coding-end insertions, those with hybrid ends likely derive from linear deletion products.

Although *Igκ* insertions originate from distinct types of RAG1/2 deletions, I propose a model in which they all share a common intermediate: excised linear DNA fragments that escaped from the post-cleavage complex prior to end joining (Figure 26). This model agrees with biochemical experiments and studies with reporter cell lines showing that cleaved ends can prematurely escape the post-cleavage complex upon destabilization by either RAG2^{core}, non-consensus RSS heptamers or absence of the DNA damage response kinase ATM (Arnal et al., 2010; Bredemeyer et al., 2006; Coussens et al., 2013; Deriano et al., 2011). My data support this model in two ways. First, the occurrence of coding- and hybrid-end insertions speaks against DNA circularization, and points to the existence of stable, linear DNA deletion products. Second, since DNA integration is independent of RAG1/2, neither donor fragments nor insertion sites require RSSs/cRSSs for the insertion process. In agreement with my findings, previous studies in reporter cell lines detected a few insertions of RSS-flanked donor substrates which were not mediated by DNA transposition or by trans-V(D)J recombination (Chatterji et al., 2006; Reddy et al., 2006). Similarly, a study in primary T cells reported a few cases in which the insertion of a specific RSS-flanked *Tcrβ* fragment occurred independent of both pathways (Curry et al., 2007). I conclude that RAG1/2 likely mobilizes linear deletion products, which are stable and have the capacity to re-insert back into the genome at independently generated DNA breaks on heterologous chromosomes. Thus, my findings reveal a novel RAG1/2-mediated insertion pathway distinct from DNA transposition and trans-V(D)J recombination.

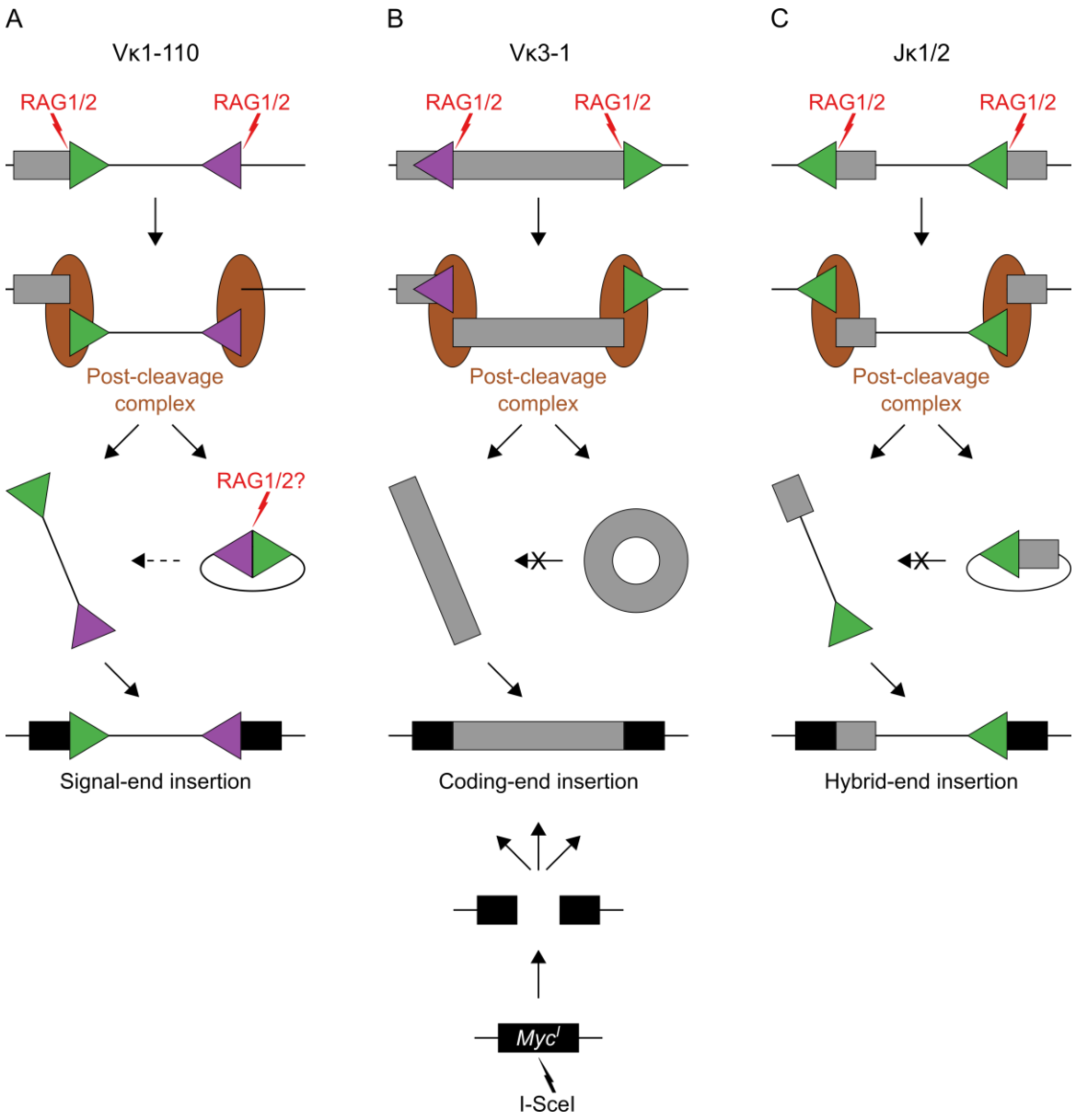

Figure 26: RAG1/2 mobilizes DNA from antibody gene segments into RAG1/2-independent DNA breaks.

A to C- Cartoon diagram to illustrate the pathways leading to insertion of RAG1/2 mobilized DNA into I-SceI breaks. Aberrant RAG1/2-mediated DNA excision at Vκ1-110 (A), Vκ3-1 (B) and Jκ1/2 (C) generates signal-end, coding-end and hybrid-end insertions, respectively. Boxes represent *Ig* segments (grey) or *Myc*^{I} (black), triangles indicate RSSs (green) or cRSSs (purple), red lightning points to RAG1/2 cleavage sites and brown ellipses represent the post-cleavage complex. From top to bottom: First, RAG1/2 induces DNA breaks at paired RSSs/cRSSs. Second, DNA is aberrantly excised and cleaved ends remain bound to the post-cleavage complex to support their repair by the NHEJ machinery. Third, excised DNA is either circularized and released from the post-cleavage complex as episomal joint (right arrow) or it escapes prior to end joining as linear fragment (left arrow). For signal-end insertions (A), linear DNA fragments might also originate from re-cleavage of episomal signal joints by RAG1/2 (dotted arrow). For coding-end and hybrid-end insertions (B and C), re-cleavage of episomal joints is unlikely due to the absence of paired RSSs (crossed arrows). Finally, mobilized linear DNA fragments re-insert into the genome at the I-SceI break.

4.4. Insertions originating from *Igh*

In theory, RAG1/2 could also mobilize DNA at *Igh*. However, insertions derived from this locus were neither detected in the presence of RAG1/2^{core} (TC-Seq) nor of RAG1/2 wild type (IC-Seq). This might be expected with RAG1/2^{core}, since *Igh* recombination is limited in the absence of the RAG2 C-terminus (Akamatsu et al., 2003; Liang et al., 2002). However, *Igh* recombination is not impaired in the presence of wild type RAG1/2 and thus *Igh* could serve as insertion donor during IC-Seq. I hypothesize that the absence of insertions from *Igh* is caused by the presence of two I-SceI sites in the utilized bone marrow B cells, one at *c-myc* and the other one at *Igh* (*Myc$^{I/I}$Igh$^{I/I}$*, see Materials and methods). Thus, if RAG1/2 releases deletion products from *Igh^{I}*, those fragments will have a higher probability to re-insert at the proximal I-SceI break at *Igh^{I}* than at the distal break at *Myc^{I}*. In agreement with this, a clear preference for proximal DNA insertion is observed at *Myc^{I}*. I therefore speculate that RAG1/2 might mobilize DNA at *Igh^{I}* but those fragments are likely captured *in cis* by the nearby I-SceI break and thus do not re-insert at detectable levels at *Myc^{I}*.

4.5. Insertions derived from non-*Ig* loci

Although RAG2^{core}-expression significantly alters the landscape of chromosomal insertions at I-SceI breaks, the majority of events originates from outside the *Igκ* locus in both RAG2^{core} and RAG2$^{-/-}$ pro-B cells (57% and 100% of total, respectively). Those insertions possibly derive from regions prone to genomic instability caused by DNA transcription, replication or other sources of DNA damage. Consistent with this possibility, chromosomal insertions in RAG2$^{-/-}$ cells preferentially originate from highly transcribed genes and ERFSs. Alternatively or in addition, non-*Igκ* insertions may represent “templated-sequence insertions” which derive from reverse-transcribed RNA (Onozawa et al., 2014). Finally, I cannot exclude that some insertions originate from RAG1/2-mediated deletions at off-target sites. In this context, it is intriguing that insertions of non-*Ig* DNA into antibody receptor genes were recently shown to contribute to antibody diversification (Tan et al., 2016).

4.6. RAG1/2 causes insertions at independent, physiologic DNA breaks

As demonstrated by the computational analysis of human cancers in this study, RAG1/2-induced DNA insertions are not limited to I-SceI breaks but also occur at physiologic DNA breaks *in vivo*. The low number of *IG*/*TCR* insertions detected in the tumor analysis is likely due to limitations of currently available datasets as well as general limitations of whole genome sequencing techniques. Many of the publicly available tumor datasets either do not have a sufficient coverage or are not sequenced using long enough reads (e.g. 100 bp and above) to allow for robust detection of insertion junctions. Moreover, the preparation of genomic libraries generally involves DNA fragmentation, which inevitably truncates existing insertions thereby causing them to appear as "translocations" in the computational analysis.

Nevertheless, the detection of RAG1/2-induced insertions is particularly important since they pose a threat to genomic stability in at least two ways. First, they provide functional RSS and/or cRSS substrates for secondary rearrangements. In fact, introducing a RSS outside of *Ig* loci has been shown to cause aberrant RAG1/2-mediated deletions and inversions (Hu et al., 2015; Mahowald et al., 2009). Consistent with this, one of the tumor-associated insertions was accompanied by the formation of a putative cRSS/cRSS signal joint, which likely originated from a secondary RAG1/2-mediated DNA inversion between the cRSS in the insert and a nearby cRSS. These and other downstream recombinations (e.g. deletions and translocations) might also render RAG1/2-induced insertions especially difficult to detect. Second, although none of the insertions in the analyzed patients are cancer drivers, the oncogenic insertion of an excised *TCR* fragment was recently described (Navarro et al., 2015). In the reported T-ALL patient, a DNA fragment flanked by two RSSs was excised from the *TRB* locus and re-inserted upstream of the *TAL1* oncogene, causing its activation. Notably, the *TRB* fragment inserted at a RAG1/2-independent DNA break, analogous to the insertions detected in my study. Furthermore, the oncogenic insertion of an *IGH* fragment was described in a patient with diffuse large B-cell lymphoma (Chaganti et al., 1998). In the reported patient, a rearranged DJ fragment inserted into a translocation junction involving the *BCL6* oncogene which led to the expression of an

aberrant *BCL6-IGH* fusion transcript. Similarly, an inserted *TCR* fragment at a translocation junction was detected in this study. Thus, RAG1/2 has the capacity to destabilize the lymphocyte genome by producing cancer-associated DNA insertions.

5. Outlook

My findings reveal a novel RAG1/2-mediated insertion pathway which destabilizes the genome and shares features with reported oncogenic DNA insertions. Three consecutive steps contribute to this pathway (see Figure 26): First, DNA is aberrantly excised from V(D)J loci by RAG1/2. Second, excised DNA is released from the post-cleavage complex as linear fragments. Third, released fragments re-integrate at RAG1/2-independent DNA breaks in the genome.

Additional studies are required to further investigate this novel pathway, particularly the contribution of post-cleavage complex destabilization to the release of linear DNA fragments and the precise mechanisms of DNA re-integration at genomic breaks. Moreover, complementary studies are necessary to investigate if other loci beyond *Ig*/*Tcr* serve as donors for RAG1/2-mediated DNA insertions. Since events from such sites likely occur at very low frequency, their experimental validation will be particularly challenging and might require further improvements of current assays regarding their sensitivity and specificity. Finally, an in-depth analysis of RAG1/2-mediated DNA insertions in human cancer is required. In this context, novel sequencing technologies with longer reading lengths will significantly reduce the current challenges in the bioinformatic detection of insertions (Goodwin et al., 2016). Altogether, future studies of RAG1/2-mediated DNA insertions will provide new insights into the genome destabilization in lymphocytes and thereby improve our mechanistic understanding of oncogenesis.

6. References

Agrawal, A., Q.M. Eastman, and D.G. Schatz. 1998. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature*. 394:744–751. doi:10.1038/29457.

Akamatsu, Y., R. Monroe, D.D. Dudley, S.K. Elkin, F. Gärtner, S.R. Talukder, Y. Takahama, F.W. Alt, C.H. Bassing, and M.A. Oettinger. 2003. Deletion of the RAG2 C terminus leads to impaired lymphoid development in mice. *Proc. Natl. Acad. Sci.* 100:1209–1214. doi:10.1073/pnas.0237043100.

Alt, F.W., Y. Zhang, F.-L. Meng, C. Guo, and B. Schwer. 2013. Mechanisms of Programmed DNA Lesions and Genomic Instability in the Immune System. *Cell*. 152:417–429. doi:10.1016/j.cell.2013.01.007.

Arnal, S.M., A.J. Holub, S.S. Salus, and D.B. Roth. 2010. Non-consensus heptamer sequences destabilize the RAG post-cleavage complex, making ends available to alternative DNA repair pathways. *Nucleic Acids Res.* 38:2944–2954. doi:10.1093/nar/gkp1252.

Barlow, J.H., R.B. Faryabi, E. Callén, N. Wong, A. Malhowski, H.T. Chen, G. Gutierrez-Cruz, H.-W. Sun, P. McKinnon, G. Wright, R. Casellas, D.F. Robbiani, L. Staudt, O. Fernandez-Capetillo, and A. Nussenzweig. 2013. Identification of early replicating fragile sites that contribute to genome instability. *Cell*. 152:620–32. doi:10.1016/j.cell.2013.01.006.

Bredemeyer, A.L., G.G. Sharma, C.-Y. Huang, B. a Helmink, L.M. Walker, K.C. Khor, B. Nuskey, K.E. Sullivan, T.K. Pandita, C.H. Bassing, and B.P. Sleckman. 2006. ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. *Nature*. 442:466–70. doi:10.1038/nature04866.

Chaganti, S.R., P.H. Rao, W. Chen, V. Dyomin, S.C. Jhanwar, N.Z. Parsa, R. Dalla-Favera, and R.S.K. Chaganti. 1998. Deregulation of BCL6 in non-Hodgkin lymphoma by insertion of IGH sequences in complex translocations involving band 3q27. *Genes. Chromosomes*

Cancer. 23:328–336. doi:10.1002/(SICI)1098-2264(199812)23:4<328::AID-GCC8>3.0.CO;2-M.

Chatterji, M., C. Tsai, and D.G. Schatz. 2006. Mobilization of RAG-generated signal ends by transposition and insertion in vivo. *Mol. Cell. Biol.* 26:1558–68. doi:10.1128/MCB.26.4.1558-1568.2006.

Clatworthy, A.E., M.A. Valencia, J.E. Haber, and M.A. Oettinger. 2003. V(D)J Recombination and RAG-Mediated Transposition in Yeast. *Mol. Cell*. 12:489–499. doi:10.1016/S1097-2765(03)00305-8.

Coussens, M.A., R.L. Wendland, L. Deriano, C.R. Lindsay, S.M. Arnal, and D.B. Roth. 2013. RAG2's Acidic Hinge Restricts Repair-Pathway Choice and Promotes Genomic Stability. *Cell Rep.* 4:870–878. doi:10.1016/j.celrep.2013.07.041.

Cowell, L.G., M. Davila, T.B. Kepler, and G. Kelsoe. 2002. Identification and utilization of arbitrary correlations in models of recombination signal sequences. *Genome Biol.* 3:research0072.1-research0072.20.

Curry, J.D., and M.S. Schlissel. 2008. RAG2's non-core domain contributes to the ordered regulation of V(D)J recombination. *Nucleic Acids Res.* 36:5750–5762. doi:10.1093/nar/gkn553.

Curry, J.D., D. Schulz, C.J. Guidos, J.S. Danska, L. Nutter, A. Nussenzweig, and M.S. Schlissel. 2007. Chromosomal reinsertion of broken RSS ends during T cell development. *J. Exp. Med.* 204:2293–2303. doi:10.1084/jem.20070583.

Deriano, L., J. Chaumeil, M. Coussens, A. Multani, Y. Chou, A.V. Alekseyenko, S. Chang, J.A. Skok, and D.B. Roth. 2011. The RAG2 C terminus suppresses genomic instability and lymphomagenesis. *Nature*. 471:119–123. doi:10.1038/nature09755.

Dobin, A., C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T.R. Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 29:15–21. doi:10.1093/bioinformatics/bts635.

Dudley, D.D., J. Sekiguchi, C. Zhu, M.J. Sadofsky, S. Whitlow, J. DeVido, R.J. Monroe, C.H. Bassing, and F.W. Alt. 2003. Impaired V(D)J Recombination and Lymphocyte Development in Core RAG1-expressing Mice. *J. Exp. Med.* 198:1439–1450. doi:10.1084/jem.20030627.

Eastman, Q.M., and D.G. Schatz. 1997. Nicking is asynchronous and stimulated by synapsis in 12/23 rule-regulated V(D)J cleavage. *Nucleic Acids Res.* 25:4370–4378.

Edgley, M., A. D'Souza, G. Moulder, S. McKay, B. Shen, E. Gilchrist, D. Moerman, and R. Barstead. 2002. Improved detection of small deletions in complex pools of DNA. *Nucleic Acids Res.* 30:e52.

Elkin, S.K., A.G. Matthews, and M. a Oettinger. 2003. The C-terminal portion of RAG2 protects against transposition in vitro. *EMBO J.* 22:1931–8. doi:10.1093/emboj/cdg184.

Goodwin, S., J.D. McPherson, and W.R. McCombie. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17:333–351. doi:10.1038/nrg.2016.49.

Helmink, B. a, and B.P. Sleckman. 2012. The response to and repair of RAG-mediated DNA double-strand breaks. *Annu. Rev. Immunol.* 30:175–202. doi:10.1146/annurev-immunol-030409-101320.

Hiom, K., M. Melek, and M. Gellert. 1998. DNA Transposition by the RAG1 and RAG2 Proteins: A Possible Source of Oncogenic Translocations. *Cell*. 94:463–470. doi:10.1016/S0092-8674(00)81587-1.

Holmfeldt, L., L. Wei, E. Diaz-Flores, M. Walsh, J. Zhang, L. Ding, D. Payne-Turner, M. Churchman, A. Andersson, S.-C. Chen, K. McCastlain, J. Becksfort, J. Ma, G. Wu, S.N. Patel, S.L. Heatley, L.A. Phillips, G. Song, J. Easton, M. Parker, X. Chen, M. Rusch, K. Boggs, B. Vadodaria, E. Hedlund, C. Drenberg, S. Baker, D. Pei, C. Cheng, R. Huether, C. Lu, R.S. Fulton, L.L. Fulton, Y. Tabib, D.J. Dooling, K. Ochoa, M. Minden, I.D. Lewis, L.B. To, P. Marlton, A.W. Roberts, G. Raca, W. Stock, G. Neale, H.G. Drexler, R.A. Dickins, D.W.

Ellison, S.A. Shurtleff, C.-H. Pui, R.C. Ribeiro, M. Devidas, A.J. Carroll, N.A. Heerema, B. Wood, M.J. Borowitz, J.M. Gastier-Foster, S.C. Raimondi, E.R. Mardis, R.K. Wilson, J.R. Downing, S.P. Hunger, M.L. Loh, and C.G. Mullighan. 2013. THE GENOMIC LANDSCAPE OF HYPODIPLOID ACUTE LYMPHOBLASTIC LEUKEMIA. *Nat. Genet.* 45:242–252. doi:10.1038/ng.2532.

Hu, J., Y. Zhang, L. Zhao, R.L. Frock, Z. Du, R.M. Meyers, F. Meng, D.G. Schatz, and F.W. Alt. 2015. Chromosomal Loop Domains Direct the Recombination of Antigen Receptor Genes. *Cell*. 163:947–959. doi:10.1016/j.cell.2015.10.016.

Ji, Y., W. Resch, E. Corbett, A. Yamane, R. Casellas, and D.G. Schatz. 2010. The In Vivo Pattern of Binding of RAG1 and RAG2 to Antigen Receptor Loci. *Cell*. 141:419–431. doi:10.1016/j.cell.2010.03.010.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, and A. Drummond. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 28:1647–1649. doi:10.1093/bioinformatics/bts199.

Kim, M.-S., M. Lapkouski, W. Yang, and M. Gellert. 2015. Crystal structure of the V(D)J recombinase RAG1-RAG2. *Nature*. 518:507–511. doi:10.1038/nature14174.

Kitamura, T., Y. Koshino, F. Shibata, T. Oki, H. Nakajima, T. Nosaka, and H. Kumagai. 2003. Retrovirus-mediated gene transfer and expression cloning: powerful tools in functional genomics. *Exp. Hematol.* 31:1007–14.

Klein, I.A., W. Resch, M. Jankovic, T. Oliveira, A. Yamane, H. Nakahashi, M. Di Virgilio, A. Bothmer, A. Nussenzweig, D.F. Robbiani, R. Casellas, and M.C. Nussenzweig. 2011. Translocation-Capture Sequencing Reveals the Extent and Nature of Chromosomal Rearrangements in B Lymphocytes. *Cell*. 147:95–106. doi:10.1016/j.cell.2011.07.048.

Küppers, R., and R. Dalla-Favera. 2001. Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene*. 20:5580–94. doi:10.1038/sj.onc.1204640.

Lee, G.S., M.B. Neiditch, R.R. Sinden, and D.B. Roth. 2002. Targeted transposition by the V(D)J recombinase. *Mol. Cell. Biol.* 22:2068–77. doi:10.1128/MCB.22.7.2068.

Lescale, C., V. Abramowski, M. Bedora-Faure, V. Murigneux, G. Vera, D.B. Roth, P. Revy, J.-P. de Villartay, and L. Deriano. 2016. RAG2 and XLF/Cernunnos interplay reveals a novel role for the RAG complex in DNA repair. *Nat. Commun.* 7:10529. doi:10.1038/ncomms10529.

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

Lewis, S.M., and G.E. Wu. 2000. The Old and the Restless. *J. Exp. Med.* 191:1631–1636. doi:10.1084/jem.191.10.1631.

Liang, H.-E., L.-Y. Hsu, D. Cado, L.G. Cowell, G. Kelsoe, and M.S. Schlissel. 2002. The "Dispensable" Portion of RAG2 Is Necessary for Efficient V-to-DJ Rearrangement during B and T Cell Development. *Immunity*. 17:639–651. doi:10.1016/S1074-7613(02)00448-X.

Lieber, M.R. 2016. Mechanisms of human lymphoid chromosomal translocations. *Nat. Rev. Cancer*. 16:387–398. doi:10.1038/nrc.2016.40.

Liu, Y., R. Subrahmanyam, T. Chakraborty, R. Sen, and S. Desiderio. 2007. A PHD FINGER DOMAIN IN RAG-2 THAT BINDS HYPERMETHYLATED LYSINE 4 OF HISTONE H3 IS NECESSARY FOR EFFICIENT V(D)J REARRANGEMENT. *Immunity*. 27:561–571. doi:10.1016/j.immuni.2007.09.005.

Mahowald, G.K., J.M. Baron, M.A. Mahowald, S. Kulkarni, A.L. Bredemeyer, C.H. Bassing, and B.P. Sleckman. 2009. Aberrantly resolved RAG-mediated DNA breaks in Atm-deficient lymphocytes target chromosomal breakpoints in cis. *Proc. Natl. Acad. Sci.* 106:18339–18344. doi:10.1073/pnas.0902545106.

Matthews, A.G.W., A.J. Kuo, S. Ramón-Maiques, S. Han, K.S. Champagne, D. Ivanov, M. Gallardo, D. Carney, P. Cheung, D.N. Ciccone, K.L. Walter, P.J. Utz, Y. Shi, T.G. Kutateladze, W. Yang, O. Gozani, and M.A. Oettinger. 2007. RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature*. 450:1106–1110. doi:10.1038/nature06431.

McBlane, J.F., D.C. van Gent, D.A. Ramsden, C. Romeo, C.A. Cuomo, M. Gellert, and M.A. Oettinger. 1995. Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell*. 83:387–395. doi:10.1016/0092-8674(95)90116-7.

Merelli, I., A. Guffanti, M. Fabbri, A. Cocito, L. Furia, U. Grazini, R.J. Bonnal, L. Milanesi, and F. McBlane. 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:W262–W267. doi:10.1093/nar/gkq391.

Messier, T.L., J.P. O'Neill, S.-M. Hou, J. a Nicklas, and B. a Finette. 2003. In vivo transposition mediated by V(D)J recombinase in human T lymphocytes. *EMBO J.* 22:1381–8. doi:10.1093/emboj/cdg137.

Mijušković, M., Y.-F. Chou, V. Gigi, C.R. Lindsay, O. Shestova, S.M. Lewis, and D.B. Roth. 2015. Off-Target V(D)J Recombination Drives Lymphomagenesis and Is Escalated by Loss of the Rag2 C Terminus. *Cell Rep.* 12:1842–1852. doi:10.1016/j.celrep.2015.08.034.

Mombaerts, P., J. Iacomini, R.S. Johnson, K. Herrup, S. Tonegawa, and V.E. Papaioannou. 1992. RAG-1-deficient mice have no mature B and T lymphocytes. *Cell*. 68:869–877. doi:10.1016/0092-8674(92)90030-G.

Murphy, K. 2012. Janeway's immunobiology. 8th ed. Garland Science, New York. 868 pp.

Navarro, J.-M., A. Touzart, L.C. Pradel, M. Loosveld, M. Koubi, R. Fenouil, S. Le Noir, M.A. Maqbool, E. Morgado, C. Gregoire, S. Jaeger, E. Mamessier, C. Pignon, S. Hacein-Bey-Abina, B. Malissen, M. Gut, I.G. Gut, H. Dombret, E.A. Macintyre, S.J. Howe, H.B. Gaspar, A.J. Thrasher, N. Ifrah, D. Payet-Bornet, E. Duprez, J.-C. Andrau, V. Asnafi, and B. Nadel.

2015. Site- and allele-specific polycomb dysregulation in T-cell leukaemia. *Nat. Commun.* 6:6094. doi:10.1038/ncomms7094.

Neiditch, M.B., G.S. Lee, L.E. Huye, V.L. Brandt, and D.B. Roth. 2002. The V(D)J Recombinase Efficiently Cleaves and Transposes Signal Joints. *Mol. Cell*. 9:871–878. doi:10.1016/S1097-2765(02)00494-X.

Nussenzweig, A., and M.C. Nussenzweig. 2010. Origin of chromosomal translocations in lymphoid cancer. *Cell*. 141:27–38. doi:10.1016/j.cell.2010.03.016.

Okosun, J., C. Bödör, J. Wang, S. Araf, C.-Y. Yang, C. Pan, S. Boller, D. Cittaro, M. Bozek, S. Iqbal, J. Matthews, D. Wrench, J. Marzec, K. Tawana, N. Popov, C. O'Riain, D. O'Shea, E. Carlotti, A. Davies, C.H. Lawrie, A. Matolcsy, M. Calaminici, A. Norton, R.J. Byers, C. Mein, E. Stupka, T.A. Lister, G. Lenz, S. Montoto, J.G. Gribben, Y. Fan, R. Grosschedl, C. Chelala, and J. Fitzgibbon. 2014. Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. *Nat. Genet.* 46:176–181. doi:10.1038/ng.2856.

Oliveira, T.Y., W. Resch, M. Jankovic, R. Casellas, M.C. Nussenzweig, and I. a Klein. 2012. Translocation capture sequencing: a method for high throughput mapping of chromosomal rearrangements. *J. Immunol. Methods*. 375:176–81. doi:10.1016/j.jim.2011.10.007.

Onozawa, M., Z. Zhang, Y.J. Kim, L. Goldberg, T. Varga, P.L. Bergsagel, W.M. Kuehl, and P.D. Aplan. 2014. Repair of DNA double-strand breaks by templated nucleotide sequence insertions derived from distant regions of the genome. *Proc. Natl. Acad. Sci.* 111:7729–7734. doi:10.1073/pnas.1321889111.

Posey, J.E., M.J. Pytlos, R.R. Sinden, and D.B. Roth. 2006. Target DNA structure plays a critical role in RAG transposition. *PLoS Biol.* 4:e350. doi:10.1371/journal.pbio.0040350.

Rahman, N.S., L.J. Godderz, S.J. Stray, J.D. Capra, and K.K. Rodgers. 2006. DNA Cleavage of a Cryptic Recombination Signal Sequence by RAG1 and RAG2 IMPLICATIONS FOR PARTIAL VH GENE REPLACEMENT. *J. Biol. Chem.* 281:12370–12380. doi:10.1074/jbc.M507906200.

Ramón-Maiques, S., A.J. Kuo, D. Carney, A.G.W. Matthews, M.A. Oettinger, O. Gozani, and W. Yang. 2007. The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. *Proc. Natl. Acad. Sci. U. S. A.* 104:18993–18998. doi:10.1073/pnas.0709170104.

Reddy, Y.V.R., E.J. Perkins, and D.A. Ramsden. 2006. Genomic instability due to V(D)J recombination-associated transposition. *Genes Dev.* 20:1575–82. doi:10.1101/gad.1432706.

Revilla-i-Domingo, R., I. Bilic, B. Vilagos, H. Tagoh, A. Ebert, I.M. Tamir, L. Smeenk, J. Trupke, A. Sommer, M. Jaritz, and M. Busslinger. 2012. The B-cell identity factor Pax5 regulates distinct transcriptional programmes in early and late B lymphopoiesis. *EMBO J.* 31:3130–3146. doi:10.1038/emboj.2012.155.

Robbiani, D.F., A. Bothmer, E. Callen, B. Reina-San-Martin, Y. Dorsett, S. Difilippantonio, D.J. Bolland, H.T. Chen, A.E. Corcoran, A. Nussenzweig, and M.C. Nussenzweig. 2008. AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. *Cell*. 135:1028–38. doi:10.1016/j.cell.2008.09.062.

Robbiani, D.F., S. Deroubaix, N. Feldhahn, T.Y. Oliveira, E. Callen, Q. Wang, M. Jankovic, I.T. Silva, P.C. Rommel, D. Bosque, T. Eisenreich, A. Nussenzweig, and M.C. Nussenzweig. 2015. Plasmodium Infection Promotes Genomic Instability and AID Dependent B Cell Lymphoma. *Cell*. 162:727–737. doi:10.1016/j.cell.2015.07.019.

Roth, D.B. 2003. Restraining the V(D)J recombinase. *Nat. Rev. Immunol.* 3:656–66. doi:10.1038/nri1152.

Ru, H., M.G. Chambers, T.-M. Fu, A.B. Tong, M. Liao, and H. Wu. 2015. Molecular Mechanism of V(D)J Recombination from Synaptic RAG1-RAG2 Complex Structures. *Cell*. 163:1138–1152. doi:10.1016/j.cell.2015.10.055.

Schatz, D.G., and Y. Ji. 2011. Recombination centres and the orchestration of V(D)J recombination. *Nat. Rev. Immunol.* 11:251–263. doi:10.1038/nri2941.

Schatz, D.G., and P.C. Swanson. 2011. V(D)J Recombination: Mechanisms of Initiation. *Annu. Rev. Genet.* 45:167–202. doi:10.1146/annurev-genet-110410-132552.

Schlissel, M.S., L.M. Corcoran, and D. Baltimore. 1991. Virus-transformed pre-B cells show ordered activation but not inactivation of immunoglobulin gene rearrangement and transcription. *J. Exp. Med.* 173:711–720. doi:10.1084/jem.173.3.711.

Sekiguchi, J.A., S. Whitlow, and F.W. Alt. 2001. Increased accumulation of hybrid V(D)J joins in cells expressing truncated versus full-length RAGs. *Mol. Cell*. 8:1383–90.

Shinkai, Y., G. Rathbun, K.-P. Lam, E.M. Oltz, V. Stewart, M. Mendelsohn, J. Charron, M. Datta, F. Young, A.M. Stall, and F.W. Alt. 1992. RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell*. 68:855–867. doi:10.1016/0092-8674(92)90029-C.

Szymczak-Workman, A.L., K.M. Vignali, and D. a a Vignali. 2012. Design and construction of 2A peptide-linked multicistronic vectors. *Cold Spring Harb. Protoc.* 2012:199–204. doi:10.1101/pdb.ip067876.

Talukder, S.R., D.D. Dudley, F.W. Alt, Y. Takahama, and Y. Akamatsu. 2004. Increased frequency of aberrant V(D)J recombination products in core RAG-expressing mice. *Nucleic Acids Res.* 32:4539–4549. doi:10.1093/nar/gkh778.

Tan, J., K. Pieper, L. Piccoli, A. Abdi, M. Foglierini, R. Geiger, C. Maria Tully, D. Jarrossay, F. Maina Ndungu, J. Wambua, P. Bejon, C. Silacci Fregni, B. Fernandez-Rodriguez, S. Barbieri, S. Bianchi, K. Marsh, V. Thathy, D. Corti, F. Sallusto, P. Bull, and A. Lanzavecchia. 2016. A

LAIR1 insertion generates broadly reactive antibodies against malaria variant antigens. *Nature*. 529:105–109. doi:10.1038/nature16450.

Teng, G., Y. Maman, W. Resch, M. Kim, A. Yamane, J. Qian, K.-R. Kieffer-Kwon, M. Mandal, Y. Ji, E. Meffre, M.R. Clark, L.G. Cowell, R. Casellas, and D.G. Schatz. 2015. RAG Represents a Widespread Threat to the Lymphocyte Genome. *Cell*. 162:751–765. doi:10.1016/j.cell.2015.07.009.

Trapnell, C., D.G. Hendrickson, M. Sauvageau, L. Goff, J.L. Rinn, and L. Pachter. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* 31:46–53. doi:10.1038/nbt.2450.

Tsai, C.-L., M. Chatterji, and D.G. Schatz. 2003. DNA mismatches and GC-rich motifs target transposition by the RAG1/RAG2 transposase. *Nucleic Acids Res.* 31:6180–6190. doi:10.1093/nar/gkg819.

Vanura, K., B. Montpellier, T. Le, S. Spicuglia, J.-M. Navarro, O. Cabaud, S. Roulland, E. Vachez, I. Prinz, P. Ferrier, R. Marculescu, U. Jäger, and B. Nadel. 2007. In Vivo Reinsertion of Excised Episomes by the V(D)J Recombinase: A Potential Threat to Genomic Stability. *PLoS Biol*. 5:e43. doi:10.1371/journal.pbio.0050043.

Wang, Q., T. Oliveira, M. Jankovic, I.T. Silva, O. Hakim, K. Yao, A. Gazumyan, C.T. Mayer, R. Pavri, R. Casellas, M.C. Nussenzweig, and D.F. Robbiani. 2014. Epigenetic targeting of activation-induced cytidine deaminase. *Proc. Natl. Acad. Sci. U. S. A.* 111:18667–18672. doi:10.1073/pnas.1420575111.

West, K.L., N.C. Singha, P. De Ioannes, L. Lacomis, H. Erdjument-Bromage, P. Tempst, and P. Cortes. 2005. A Direct Interaction between the RAG2 C Terminus and the Core Histones Is Required for Efficient V(D)J Recombination. *Immunity*. 23:203–212. doi:10.1016/j.immuni.2005.07.004.

Yu, K., and M.R. Lieber. 2000. The Nicking Step in V(D)J Recombination Is Independent of Synapsis: Implications for the Immune Repertoire. *Mol. Cell. Biol.* 20:7914–7921.

Zhang, Y., R.P. McCord, Y.-J. Ho, B.R. Lajoie, D.G. Hildebrand, A.C. Simon, M.S. Becker, F.W. Alt, and J. Dekker. 2012. Spatial Organization of the Mouse Genome and Its Role in Recurrent Chromosomal Translocations. *Cell*. 148:908–921. doi:10.1016/j.cell.2012.02.002.

7. Appendices

7.1. Supplemental figures

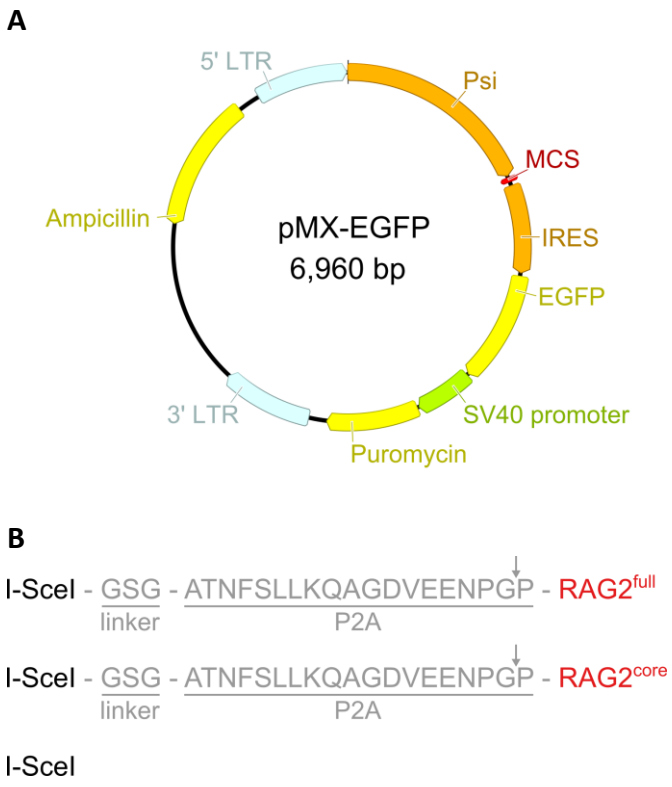

Figure S1: Overview of retroviral plasmids.

A- Schematic overview of the retroviral expression vector pMX-EGFP. The plasmid contains a retroviral 5' long terminal repeat (5' LTR, light blue), an extended retroviral packaging signal (Psi, orange), a multi-cloning site (MCS, red), an internal ribosomal entry site (IRES, orange), the coding sequence of the enhanced green fluorescent protein (EGFP, yellow), a SV40 promoter (SV40, green), the coding sequence of the puromycin-resistance gene (Puromycin, yellow), a retroviral 3' long terminal repeat (3' LTR, light blue) and the coding sequence of the ampicillin-resistance gene (Ampicillin, yellow). LTRs and Psi are derived from the Moloney murine leukemia virus, IRES is derived from the encephalomyocarditis virus and SV40 is derived from the simian virus 40. pMX-EGFP is adapted from (Kitamura et al., 2003).

B- Overview of I-SceI/RAG2-expression constructs. Top and middle: I-SceI (black) is fused to RAG2^{full} or RAG2^{core} (both in red) through a GSG-linker and a P2A-peptide sequence (both in grey). During translation, P2A "self-cleaves" by ribosomal skipping (arrow) allowing co-expression of both I-SceI and RAG2^{full} or RAG2^{core}. The GSG-linker enhances the "self-cleavage" efficiency of P2A (Szymczak-Workman et al., 2012). Both constructs were cloned into the MCS of pMX-EGFP to generate pMX-I-SceI-P2A-RAG2^{full}-EGFP and pMX-I-SceI-P2A-RAG2^{core}-EGFP, respectively (see Materials and methods). Bottom: An I-SceI-expression construct that was previously cloned into pMX-EGFP (pMX-I-SceI-EGFP; (Robbiani et al., 2008)). I-SceI is preceded by a nuclear localization sequence and a human influenza hemagglutinin tag in all constructs (not shown).

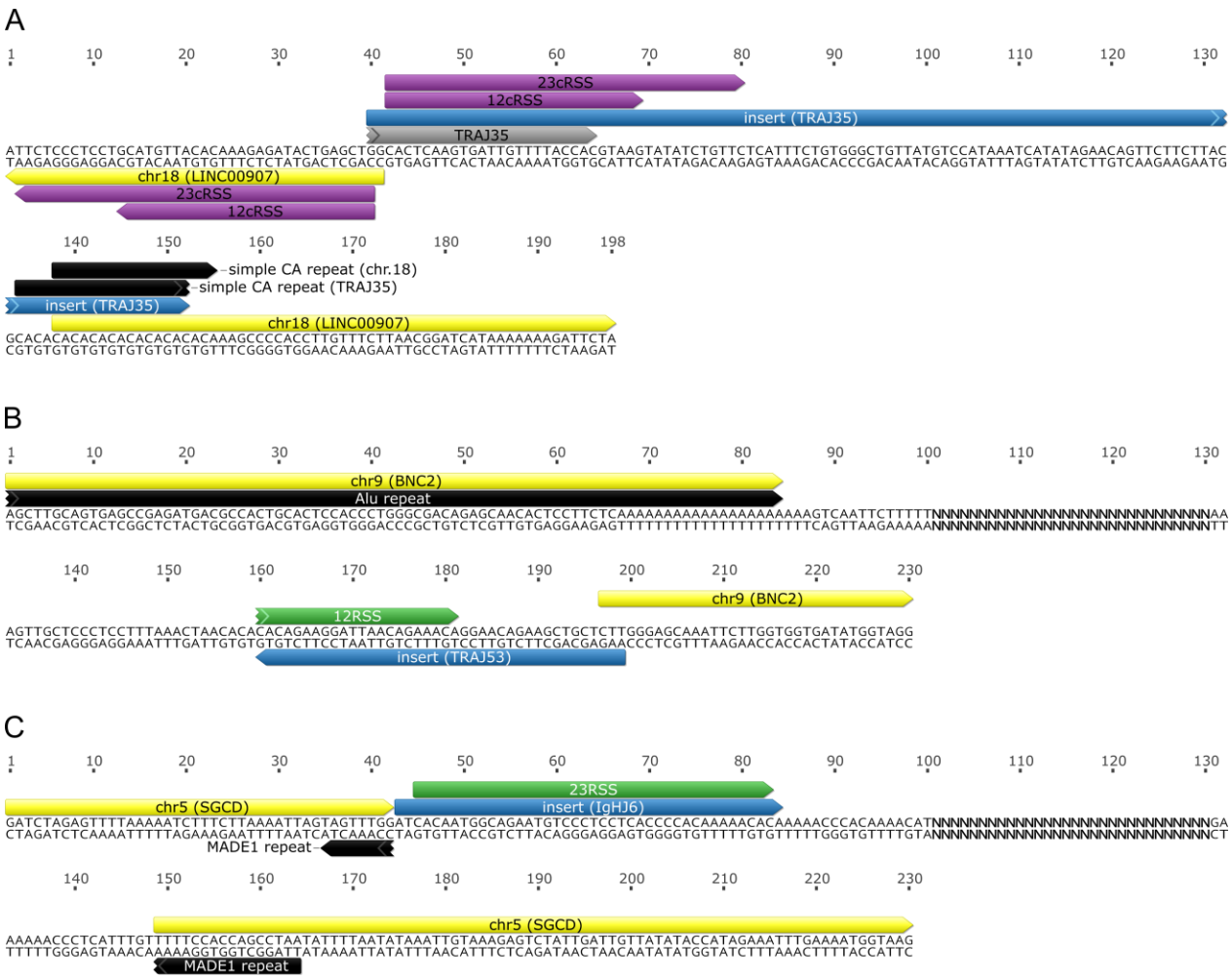

Figure S2: Sequences of inserted *IG*/*TCR* fragments detected in ALL.

A to C- Annotated sequences of insertions detected in hypodiploid ALL (A) and early T-cell precursor ALL (B and C). Annotations are color-coded to indicate insertion sites (yellow), inserted fragments (blue), *IG*/*TCR* segments (grey), repeat regions (black), RSSs (green) and cRSSs (purple). Sequences are annotated on top or below depending on strand orientation (positive or negative, respectively). Insertion sites are labeled with their corresponding chromosome and gene (all are located in introns). Inserted fragments are labeled with their corresponding *IG*/*TCR* segment of origin. Unresolved junctions are indicated by a stretch of "N". Bp positions are indicated on top. Whole genome sequences from 28 cancer patients were analyzed.

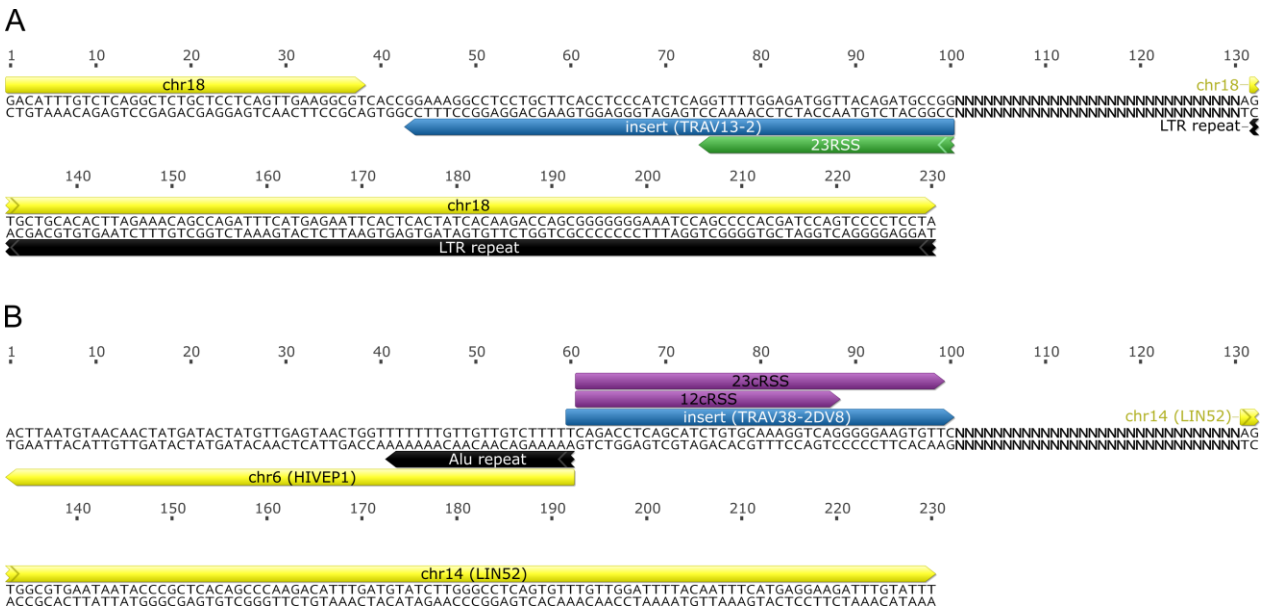

Figure S3: Sequences of inserted *IG*/*TCR* fragments detected in FL.

A to B- Annotated sequences of insertions detected in FL. Annotations are color-coded to indicate insertion sites (yellow), inserted fragments (blue), repeat regions (black), RSSs (green) and cRSSs (purple). Sequences are annotated on top or below depending on strand orientation (positive or negative, respectively). Insertion sites are labeled with their corresponding chromosome and gene, if applicable. The genic insertion site in B is located in an intron. Inserted fragments are labeled with their corresponding *IG*/*TCR* segment of origin. Unresolved junctions are indicated by a stretch of "N". Bp positions are indicated on top. Whole genome sequences from 6 cancer patients were analyzed.

7.2. Supplemental tables

Table S1: Overview of RAG1/2^{core}-dependent rearrangement breakpoint clusters

Breakpoint cluster ID	Chromosome	Start position	End position	Size [bps]	Number of rearrangements	Overlap with repetitive sequences [%]	Type of repetitive sequence	Distance to closest gene [bps]	Closest gene	Insertions originating from cluster
cluster 666	chr6	68121826	68121829	3	3	0		0	Igκv1-117	
cluster 668a	chr6	68271264	68271268	4	3	0		0	Igκv1-110	yes
cluster 669	chr6	68339988	68340002	14	3	0		0	Igκv11-106	
cluster 671a	chr6	68631962	68631969	7	4	0		0	Igκv10-96	yes
cluster 672b	chr6	68680840	68680854	14	5	0		0	Igκv10-95	yes
cluster 673a	chr6	68704509	68704515	6	3	0		0	Igκv10-94	yes
cluster 674a	chr6	68736285	68736302	17	4	0		0	Igκv19-93	yes
cluster 676	chr6	69110898	69110907	9	3	0		0	Igκv4-77	yes
cluster 678	chr6	69438203	69438223	20	4	0		0	Igκv4-59	
cluster 679	chr6	69544356	69544372	16	3	0		0	Igκv4-57-1	yes
cluster 680	chr6	69764524	69764527	3	3	0		0	Igκv12-46	
cluster 681a	chr6	69814625	69814637	12	3	0		0	Igκv12-44	
cluster 682	chr6	70074015	70074035	20	5	0		0	Igκv6-32	yes
cluster 684a	chr6	70216860	70216862	2	3	0		0	Igκv8-24	
cluster 686d	chr6	70704167	70704182	15	24	0		0	Igκv3-1	yes
cluster 687a	chr6	70722561	70722582	21	58	0		0	Igκj1	yes
cluster 687c	chr6	70722907	70722938	31	43	0		0	Igκj2	yes
cluster 688a	chr6	70723521	70723555	34	56	0		0	Igκj4	yes
cluster 688c	chr6	70723874	70723886	12	35	0		0	Igκj5	yes
cluster 668b	chr6	68271336	68271339	3	4	0		-70	Igκv1-110	yes
cluster 671b	chr6	68632044	68632052	8	4	0		0	Igκv10-96	yes
cluster 672a	chr6	68680755	68680763	8	3	0		0	Igκv10-95	yes
cluster 673b	chr6	68704593	68704599	6	5	0		0	Igκv10-94	yes
cluster 686a	chr6	70703994	70704001	7	18	0		0	Igκv3-1	yes
cluster 299	chr16	11470965	11470979	14	3	0		0	Snx29	
cluster 317	chr16	60126546	60126553	7	3	0		0	Epha6	
cluster 339b	chr17	51880737	51880747	10	3	0		-17521	Gm37176	
cluster 373	chr18	56306296	56306303	7	3	100	LTR	-84279	Gm25476	
cluster 52a	chr1	185277333	185277350	17	4	0		0	Rab3gap2	
cluster 642	chr5	140311229	140311254	25	4	100	LTR	0	Mad1l1	
cluster 766b	chr7	118655340	118655355	15	3	0		0	Tmc5	
cluster 795	chr8	14261740	14261741	1	3	0		0	Dlgap2	
cluster 840	chr8	121550980	121551005	25	3	0		0	Fbxo31,Gm20388	

Igκ coordinates: chromosome 6, 67,555,636 - 70,726,966
color-coding indicates if breakpoint clusters occur at physiologic RSSs (green) or at cRSSs (purple)

Table S2: RIC scores of cRSSs detected at Vκ and off-target clusters

	12cRSS configuration		
Breakpoint cluster ID	**Sequence**	**RIC score**	**RIC pass/fail**
12RSS Vκ3-1	cacagtgctccagggctgaacaaaaacc	-18.69136832	Pass
Cluster 52a	cacacatgcaaaaccctcccccacatcc	-42.24220581	Fail
Cluster 299	cagagatgtgacctcccagatgttctgc	-60.19876497	Fail
Cluster 317	caccacacaggcaaaccattcagctcca	-60.2941152	Fail
Cluster 339b	caccttttctaacactgctgctctttct	-62.57822193	Fail
Cluster 373	caccatgcttcctgccatgacgataatg	-39.9072581	Fail
Cluster 642	cacatcccatccacaccagggagagagg	-56.88263022	Fail
Cluster 766b	cagacaggtagctcattgcatggtcaca	-50.29668766	Fail
Cluster 795	cagactcatgtggatgagggatggtgat	-60.57313112	Fail
Cluster 840	cacatccgccactgtccaaagcttctca	-57.70119555	Fail
Vκ1-110	cacataaataacatatttagcagctggg	-59.05145761	Fail
Vκ3-1	cactgcattaaacttgtgccataatatt	-46.79598368	Fail
Vκ10-94/95/96	cactgccactgaaccttgatgggactcc	-48.78786612	Fail

	23cRSS configuration		
Breakpoint cluster ID	**Sequence**	**RIC score**	**RIC pass/fail**
23RSS Jκ1	cacagtggtagtactccactgtctggctgtacaaaaacc	-26.69788349	Pass
Cluster 52a	cacacatgcaaaaccctcccccacatccttgctcagtct	-66.3578573	Fail
Cluster 299	cagagatgtgacctcccagatgttctgctggagtatttt	-69.55445364	Fail
Cluster 317	caccacacaggcaaaccattcagctccacccgtcggctg	-80.48034728	Fail
Cluster 339b	caccttttctaacactgctgctctttctgccacaacttt	-62.073966	Fail
Cluster 373	caccatgcttcctgccatgacgataatggactaaacctc	-70.44054501	Fail
Cluster 642	cacatcccatccacaccagggagagagggatgagtgtat	-77.97159185	Fail
Cluster 766b	cagacaggtagctcattgcatggtcacatcctaaccctg	-80.55726856	Fail
Cluster 795	cagactcatgtggatgagggatggtgatgctgttgtggg	-70.71142104	Fail
Cluster 840	cacatccgccactgtccaaagcttctcaggactaacaga	-68.39744464	Fail
Vκ1-110	cacataaataacatatttagcagctgggatacccaaagt	-61.92512645	Fail
Vκ3-1	cactgcattaaacttgtgccataatattcaacactttca	-66.49278208	Fail
Vκ10-94	cactgccactgaaccttgatgggactcctgagtgtaaac	-63.43022855	Fail
Vκ10-95/96	cactgccactgaaccttgatgggactcctgagtgtaatc	-65.62745313	Fail

RIC scores of physiologic RSSs (green) are shown as control for analyzed cRSSs
pass/fail thresholds: 12RSS pass with RIC > -38.81, 23RSS pass with RIC > -58.45 (Cowell et al., 2002)

Table S3: Overview of RAG1/2^{core}-induced *Igκ* insertions at *Myc*^{I}

Method	Insertion ID	Chromosome	Start position	End position	Size [bps]	Overlap with repetitive sequences [%]	Type of repetitive sequence	Distance to closest gene [bps]	Closest gene	Insertion species
IC-Seq	Lib6_1002	chr6	68271267	68271337	70	0		-1	Igkv1-110	signal-end
IC-Seq	Lib6_1003	chr6	68782431	68782432	single junction	100	LINE	-13329	Igkv4-91	undefined
IC-Seq	Lib5_1613	chr6	69283999	69284000	single junction	0		0	Igkv4-69	single junction
IC-Seq	Lib5_1616	chr6	69866175	69866247	72	93	SINE	2380	Igkv5-40-1	undefined
IC-Seq	Lib6_1004	chr6	70703996	70704177	181	0		0	Igkv3-1	coding-end
IC-Seq	Lib6_1006	chr6	70722561	70722633	72	0		0	Igkj1	undefined
IC-Seq	Lib6_1005	chr6	70722561	70722562	single junction	0		0	Igkj1	single junction
IC-Seq	Lib5_1617	chr6	70722561	70722884	323	0		0	Igkj1	hybrid-end
IC-Seq	Lib5_1618	chr6	70722561	70722915	354	0		0	Igkj1	hybrid-end
IC-Seq	Lib6_1007	chr6	70722561	70722912	351	0		0	Igkj1	hybrid-end
IC-Seq	Lib5_1620	chr6	70722563	70722915	352	0		0	Igkj1	hybrid-end
IC-Seq	Lib5_1619	chr6	70722563	70722913	350	0		0	Igkj1	hybrid-end
IC-Seq	Lib6_1008	chr6	70722563	70722680	117	0		0	Igkj1	undefined
IC-Seq	Lib5_1621	chr6	70722564	70722915	351	0		0	Igkj1	hybrid-end
IC-Seq	Lib5_1622	chr6	70722566	70722915	349	0		0	Igkj1	hybrid-end
IC-Seq	Lib5_1623	chr6	70722601	70722890	289	0		-3	Igkj1	undefined
IC-Seq	Lib6_1009	chr6	70722780	70722781	single junction	0		135	Igkj2	single junction
IC-Seq	Lib5_1624	chr6	70722843	70722915	72	0		1	Igkj2	undefined
IC-Seq	Lib6_1010	chr6	70722849	70722913	64	0		3	Igkj2	undefined
IC-Seq	Lib5_1625	chr6	70722852	70722911	59	0		5	Igkj2	undefined
IC-Seq	Lib5_1626	chr6	70722859	70722909	50	0		7	Igkj2	undefined
IC-Seq	Lib6_1011	chr6	70722906	70722907	single junction	0		9	Igkj2	single junction
IC-Seq	Lib6_1012	chr6	70723331	70723547	216	0		1	Igkj4	undefined
IC-Seq	Lib6_1013	chr6	70723548	70723622	74	0		0	Igkj4	undefined
IC-Seq	Lib6_1014	chr6	70723549	70723880	331	0		0	Igkj4	hybrid-end
IC-Seq	Lib5_1631	chr6	70723549	70723885	336	0		0	Igkj4	hybrid-end
IC-Seq	Lib5_1630	chr6	70723549	70723885	336	0		0	Igkj4	hybrid-end
IC-Seq	Lib5_1629	chr6	70723549	70723881	332	0		0	Igkj4	hybrid-end
IC-Seq	Lib5_1627	chr6	70723549	70723620	71	0		0	Igkj4	undefined
IC-Seq	Lib5_1628	chr6	70723549	70723879	330	0		0	Igkj4	hybrid-end
IC-Seq	Lib5_1632	chr6	70723551	70723885	334	0		0	Igkj4	hybrid-end
IC-Seq	Lib6_1015	chr6	70723552	70723629	77	0		0	Igkj4	undefined
IC-Seq	Lib5_1633	chr6	70723552	70723885	333	0		0	Igkj4	hybrid-end
IC-Seq	Lib5_1634	chr6	70723553	70723885	332	0		0	Igkj4	hybrid-end
IC-Seq	Lib6_1016	chr6	70723558	70723738	180	0		0	Igkj4	undefined
IC-Seq	Lib5_1635	chr6	70723608	70723885	277	0		1	Igkj5	undefined
IC-Seq	Lib6_1017	chr6	70723634	70723884	250	0		2	Igkj5	undefined
IC-Seq	Lib6_1018	chr6	70723652	70723884	232	0		2	Igkj5	undefined
IC-Seq	Lib6_1019	chr6	70723823	70723881	58	0		5	Igkj5	undefined
IC-Seq	Lib6_1020	chr6	70723832	70723883	51	0		3	Igkj5	undefined
IC-Seq	Lib6_1021	chr6	70723840	70723881	41	0		5	Igkj5	undefined
IC-Seq	Lib5_1636	chr6	70723856	70723857	single junction	0		29	Igkj5	single junction
TC-Seq	left_f6d104b28b8b60e5c931f8e8cbe4ea6309c2917588df2655fe719796a20256d54_2019	chr6	67954228	67954317	89	0		0	Igkv9-123	coding-end
TC-Seq	left_10ae8720aad403a7187a75197d254f1853a5ea82f2759e8d7660ba1eb7747726_2127	chr6	67954228	67954317	89	0		0	Igkv9-123	coding-end
TC-Seq	left_7dd5387891671746bdb6fbe462be0772e6f42651c643f5028ce6fcf44e95f5e80_1996	chr6	67954230	67954316	86	0		0	Igkv9-123	coding-end
TC-Seq	left_d88850cfc0b7f217da9617ff257612e7ae4204faab47415e9f1b3939976c73087_1989	chr6	67954230	67954316	86	0		0	Igkv9-123	coding-end
TC-Seq	left_4af47d0361cb78cb0942793b73aedcd3c2925eecce8945105aea4960fabd71a8_1994	chr6	68271267	68271338	71	0		-1	Igkv1-110	signal-end
TC-Seq	left_11a4560134bbd11cb777999dc3a1bb010a21dacc3f68e20b4a4ae83c2271c4ff_2123	chr6	68271268	68271335	67	0		-2	Igkv1-110	signal-end
TC-Seq	right_d47152e58d8e62b302b2582923e3a9961e9f85caadd0b4a9991aacad585884ad_1806	chr6	68271268	68271337	69	0		-2	Igkv1-110	signal-end
TC-Seq	left_08888c4c1e033bf6f92d902e47208f480be092f755e644d0ee4d3b436d730ebe_2124	chr6	68271268	68271335	67	0		-2	Igkv1-110	signal-end
TC-Seq	left_8ed9063f7b26de430e229baa3d57e46f26155769f9c987cd64da7bb94e3aea6c_2032	chr6	68271269	68271338	69	0		-3	Igkv1-110	signal-end
TC-Seq	right_288d67832891326 59a94bb27f00b4dbdad2c7c1e4fa3c139f673117ee7bb37f0_1811	chr6	68271269	68271338	69	0		-3	Igkv1-110	signal-end
TC-Seq	left_d6a6e32eece1211acaea3d36ef89c88b068b64870c6f1a23d27f247 21afd0e34_2007	chr6	68271269	68271338	69	0		-3	Igkv1-110	signal-end

TC-Seq	left_9a468d702a6198c73f9d2dfff52706ff02a411427cff22f530991bd402dfbc88_2007	chr6	68271269	68271338	69	0		-3	Igκv1-110	signal-end
TC-Seq	left_bd238c0ebf51bf49b364817041a597c931b1de53da14b857e299965e12b4c627_2124	chr6	68271334	68271335	single junction	0		-68	Igκv1-110	single junction
TC-Seq	left_37cb358c4785dda1354c6e5797f6f30190c5430ac0f30b38eaf45a9eff22d114_2115	chr6	68271337	68271338	single junction	0		-71	Igκv1-110	single junction
TC-Seq	right_2994a19e4d844572d6d83c5844c8968350cbce8e2145c562e33005fe0b2c5474_1815	chr6	68303074	68303155	81	0		0	Igκv2-109	coding-end
TC-Seq	left_65a73c5ff31f738cf6fc20e27bc61bf7a6ccf8dda4cce6922b3c035fb24f2d6a_1994	chr6	68631962	68632053	91	0		0	Igκv10-96	coding-end
TC-Seq	right_723b5055aa958fefd36c2e3914e47bc442baea32fd663a593814dadd94b2d31d_1817	chr6	68631963	68632041	78	0		0	Igκv10-96	coding-end
TC-Seq	right_10c1b9a5b85a7b3468f6523309769a08dc6b9e059d712d8e23c2c022509f437d_1854	chr6	68631963	68632051	88	0		0	Igκv10-96	coding-end
TC-Seq	right_94d870af21498d917860453bded0bc2d3d97db4769f85cc21d5f96ed50e2c574_1876	chr6	68631963	68632041	78	0		0	Igκv10-96	coding-end
TC-Seq	right_9c0294cd2d7832a3d7849f638ba2d99cac45e5846787527ca4e545971433fc29_1885	chr6	68631963	68632041	78	0		0	Igκv10-96	coding-end
TC-Seq	right_568ab99d4508bbe82c768e69424701486a25c4ef189456877910810fa49861ee_1821	chr6	68631963	68631964	single junction	0		0	Igκv10-96	single junction
TC-Seq	right_c4a632cf5447973bd5e0caf27c6afde7e8f43b2dbe56d0bab07ace3eea847449_1770	chr6	68631966	68632052	86	0		0	Igκv10-96	coding-end
TC-Seq	right_1429fb834ccea81d00b742f23b0c483c6b46dd852614a16fcb2d1768057723ab_1728	chr6	68631966	68632052	86	0		0	Igκv10-96	coding-end
TC-Seq	left_9e5e11c7baee7a531c2e7ade405964c7cb7d23d6ab5a89272fa2548dc54e251f_2102	chr6	68631967	68632051	84	0		0	Igκv10-96	coding-end
TC-Seq	left_f3ab521e4114469af1a912ef54c2591a4f76c807ab2955bccc10e726857cb36d_2092	chr6	68631967	68632050	83	0		0	Igκv10-96	coding-end
TC-Seq	left_aa6222981392694df1030d166c8e8de87d624fce393ce31b6649a4d0a6b3f601_2109	chr6	68631967	68632050	83	0		0	Igκv10-96	coding-end
TC-Seq	left_d85ee0c7f195741e43aaca80e516a341fd7852f306218870398a14c8779b6c3f_2047	chr6	68631967	68632050	83	0		0	Igκv10-96	coding-end
TC-Seq	left_aca606622ff48809cfd15283bc3596701dabecb8c500ff3bb32ee17bf50b93c8_2128	chr6	68631967	68631968	single junction	0		0	Igκv10-96	single junction
TC-Seq	left_c2fb8fb8ef96d5fff313adfc7469457a4eebf7ac22d39e71f56bd788839b4b4f_2134	chr6	68631967	68632050	83	0		0	Igκv10-96	coding-end
TC-Seq	left_8e7af6320be4ad1e78667fad02b2da90e2721f1b8fba5abd3fa26bd9bea01d2d_2099	chr6	68631968	68632050	82	0		0	Igκv10-96	coding-end
TC-Seq	right_2e9c1733b37ce09e964add60dabe71c5654b2744d635f046ad1e15bd4594afaa_1870	chr6	68631971	68632054	83	0		0	Igκv10-96	coding-end
TC-Seq	right_ad2b65f60177cbc3dc9af8505cc1280992e933fbf4dae12ca62f6d51869c3a43_1845	chr6	68631971	68632054	83	0		0	Igκv10-96	coding-end
TC-Seq	right_b9fba1fabb74405f4738c5e245b4c4e4aa94cf3a3abfe257ace07c9a367ce5bf_1769	chr6	68632051	68632052	single junction	0		0	Igκv10-96	single junction
TC-Seq	left_87e114d2c088a8c7d793b3ad93162300496ee1d6f705b8db29553f80d3583584_2098	chr6	68632051	68632052	single junction	0		0	Igκv10-96	single junction
TC-Seq	left_b99e2fd7ed383a4f40a1f14a8317a4d9a72d46bbd3bf6d2561e7294cc8d9feef_2036	chr6	68680761	68680847	86	0		0	Igκv10-95	coding-end
TC-Seq	right_00e1f9bc83def641b933d033f8036550320dc6d853a1953881da42b352e6bb9b_1807	chr6	68680762	68680845	83	0		0	Igκv10-95	coding-end
TC-Seq	right_ed631a71bfa3dc1924853bdc97a49036dd5aa5296cace7dc7afcd56952efbc46_1809	chr6	68680762	68680845	83	0		0	Igκv10-95	coding-end
TC-Seq	left_0e470bd769986112c6d66c15f0f82cde314b094d31ef20e22264455857cf7123_2009	chr6	68704507	68704595	88	0		0	Igκv10-94	coding-end
TC-Seq	left_f112ca9d3ab79ce88699393bca360a85e2259781d2492d0edd10ca7ac6ecaab7_2024	chr6	68704507	68704594	87	0		0	Igκv10-94	coding-end
TC-Seq	left_797898ce0c97c4a6b550a6d878139f945bab43ea8057a8cd56b53e5104a25391_2029	chr6	68704507	68704594	87	0		0	Igκv10-94	coding-end
TC-Seq	left_e31a365c75b74f95727b121738ed7c236caed9d67491a5a217eacdce3f5361f8_2029	chr6	68704507	68704594	87	0		0	Igκv10-94	coding-end
TC-Seq	right_73c9af22d78d9d8bf0f3c6dc02b8565445ee375df7884a73cac09c5593f4a253_1769	chr6	68704508	68704509	single junction	0		0	Igκv10-94	single junction
TC-Seq	left_115fe70b4efb49d68b06036bf885e07d2a6fb5fe18910463325d75e3aebf7e5b_2082	chr6	68704509	68704593	84	0		0	Igκv10-94	coding-end
TC-Seq	left_1e292e11326d1860e853d9878612893c8bd9fb8668c33d71584cb5bf44a8e8aa_2012	chr6	68704509	68704593	84	0		0	Igκv10-94	coding-end
TC-Seq	left_86a0a39851fce2ba8d3a9e4849711e5c6a58b572ab34457e70b5c4d48dc3ca35_2010	chr6	68704510	68704596	86	0		0	Igκv10-94	coding-end
TC-Seq	left_6931378a7321b08396a407561798b3b3c3a580365cf4b76abe93524f1ddbd9a1_1996	chr6	68704510	68704596	86	0		0	Igκv10-94	coding-end
TC-Seq	left_2a6930201e58c2d599ba8dd7d131ad81fe09eba7d5b89342fed10945c3687d67_2026	chr6	68704510	68704596	86	0		0	Igκv10-94	coding-end
TC-Seq	left_a39ba18c7473006f7c3509aa301a1583b489350cb5e3aa76564872aa9e7d86ca_2030	chr6	68704510	68704596	86	0		0	Igκv10-94	coding-end
TC-Seq	left_1347daba6aff5cf8fbf0835ccac2cc215e5405380995d7cc5ae708d2b837bb6a_2035	chr6	68704510	68704596	86	0		0	Igκv10-94	coding-end
TC-Seq	left_96e0397e9082eb9a25795cde9b91f7b477bea71be66a9b25a0878a9ca1a5ecfd_2041	chr6	68704510	68704596	86	0		0	Igκv10-94	coding-end
TC-Seq	left_b12d847ac36ddfd84af16db9d1806a361a8d910e98f4ba7ba7b4da28345163bc_2002	chr6	68704510	68704598	88	0		0	Igκv10-94	coding-end
TC-Seq	left_880d07624e0aebed0067e71d07c73059c3e30cd68c2d58adbb0162e259fd2702_2067	chr6	68704510	68704596	86	0		0	Igκv10-94	coding-end
TC-Seq	left_aab332fa3b7e09a1cde83414e557b2fd29e4cac1c116649e1ecd6aa97535b5b9_2008	chr6	68704510	68704598	88	0		0	Igκv10-94	coding-end
TC-Seq	left_4ae6381f6833e3faf1276a8c5e4467feae43523428082a948d68950a5032e69b_2036	chr6	68704511	68704593	82	0		0	Igκv10-94	coding-end
TC-Seq	right_6fa7ad21d2210e25f362cabd5f4a780d5ce3056ef187c4f7a666023b91a63b54_1819	chr6	68704511	68704598	87	0		0	Igκv10-94	coding-end
TC-Seq	left_0cab8bb9a1b269c6c1ba0e344bf83783ffdaf3caad3627bf2a3f292f42a7791f_1996	chr6	68704511	68704593	82	0		0	Igκv10-94	coding-end
TC-Seq	right_730b4ff73af3362b01cb619d316c1a4a2d69a3d61ef382a3d2e8827e820aaf6a_1793	chr6	68704592	68704593	single junction	0		0	Igκv10-94	single junction
TC-Seq	left_961e51a1cd5df5477cc887ec4de5c4ba169edea0aab223c2d422cab9a33bab5a_2052	chr6	68736302	68736382	80	0		0	Igκv19-93	coding-end
TC-Seq	left_5a5aa3cd3689a165da77aa943bdd5d7e8de5eb028e531670661826aaa7479432_1996	chr6	68736302	68736382	80	0		0	Igκv19-93	coding-end
TC-Seq	left_3c524009f85a5c5835d58f2250aee1448ab18584f0c43f2d213d7899245b0549_2076	chr6	69059689	69059779	90	0		0	Igκv4-78	coding-end
TC-Seq	left_54ab749c0106f0289eb185b5d5dcb24a54ea8d0b87ffe6a8c95ee40e2e5213c3_2123	chr6	69059689	69059690	single junction	0		0	Igκv4-78	single junction
TC-Seq	right_5c61424925459b418063a8ecacf3922eb760966a50d9fc3ad0e5e419c20cc5c9_1805	chr6	69059700	69059777	77	0		0	Igκv4-78	coding-end
TC-Seq	right_eb494d1e24ae5815a4c7b4dcc484377583d49ecb8d76ba41421306307c5d77d5_1811	chr6	69059780	69059781	single junction	0		0	Igκv4-78	single junction
TC-Seq	right_d1c9406d8c0b1ae749afb801a3759de739ade6b03a4a997e7161ae835639ee53_1720	chr6	69059780	69059781	single junction	0		0	Igκv4-78	single junction
TC-Seq	left_cc46f43ca5ea0a034acf88b2ed884d4cacce932ce08f242ac0c68153f3237d39_2097	chr6	69110908	69110909	single junction	0		0	Igκv4-77	single junction
TC-Seq	left_1881c6857aceeae9fa1e55abeb4e29ebde5af26654eadeb4f0b717e5cf6862fe_2076	chr6	69110908	69110909	single junction	0		0	Igκv4-77	single junction
TC-Seq	left_c1a47e9425e314697d4069d4067c2e0bdf2edcd5491f1459fec1da5bad3b30bd_2073	chr6	69110908	69110909	single junction	0		0	Igκv4-77	single junction
TC-Seq	right_cb2ebdd41832c72d8d876b1d6f5827c70e03adc01d7af9780dc4b8e62b5f86c2_1874	chr6	69184825	69184915	90	0		0	Igκv4-74	coding-end
TC-Seq	right_6438ccde1c8d4064baf2ee99043022168a06ebf31801a71d8ac9c1716dc7c501_1869	chr6	69184825	69184915	90	0		0	Igκv4-74	coding-end

TC-Seq	right_dffe106ee1ef3c36c4837b5b9ffd73f3411c2f32c26c1da0ccb349c1d58ca532_1845	chr6	69184825	69184915	90	0		0	lgkv4-74	coding-end
TC-Seq	left_744a39bd9de69ba2381375 1b6f0104ed95b4806f0197dd24b4d62b78691364a_2011	chr6	69184830	69184914	84	0		0	lgkv4-74	coding-end
TC-Seq	left_4edfcd19a748ee4563f673765a2f56bd62c7d36f38c5e85a65554172af15e0c9_2052	chr6	69184830	69184914	84	0		0	lgkv4-74	coding-end
TC-Seq	right_51612c6b621db181d75c9375e6a064135add32e4e32acd26974aae8cd798c1bd_1793	chr6	69184914	69184915	single junction	0		0	lgkv4-74	single junction
TC-Seq	left_4ac1aec8ab034943b8cfadca914feaea571292e6fa8d5abc01c1ff8549ae5984_2000	chr6	69197586	69197587	single junction	0		0	lgkv4-73	single junction
TC-Seq	left_7f4b1cb655b4a182655451 1ac87ca660e4fbd2cb5d5ae54c88ac f14b955df783_2042	chr6	69197671	69197672	single junction	0		0	lgkv4-73	single junction
TC-Seq	left_86da7440dd69dc1b16e91d4e3366bab3362dba960c30f7818f6c32c820be3d36_1998	chr6	69243161	69243162	single junction	0		0	lgkv4-71	single junction
TC-Seq	left_6a5d724b27cd27bd6a14356e9464f2bf0a7c156f351727a6fbaa013a01452815_2036	chr6	69243162	69243163	single junction	0		0	lgkv4-71	single junction
TC-Seq	right_1a2f7ad2a457758b5a051f3d7f31fd09961b32b58c338aed010761141ff7afb6d_1810	chr6	69243162	69243163	single junction	0		0	lgkv4-71	single junction
TC-Seq	left_fefcae5bac9ca8f43229d5e365fb1923372488b17acfd45d17ac80e6dca629fa9_2115	chr6	69243246	69243247	single junction	0		0	lgkv4-71	single junction
TC-Seq	right_ec9fbe46dcad31a1f9ff81d3e3f0d58c2c0989f93ae751b01eb2f00ce280ca5f_1744	chr6	69243246	69243247	single junction	0		0	lgkv4-71	single junction
TC-Seq	right_817e5d0bb557e267fbd6055cd9211795b3437ce5d4ca7a4edcfed99ad958a5b3_1811	chr6	69243251	69243252	single junction	0		0	lgkv4-71	single junction
TC-Seq	left_a11413255a3b57b2b1973a831c25a551faa4850d80440b6eb7268bae18d7ce7c_2006	chr6	69500256	69500341	85	0		0	lgkv4-58	coding-end
TC-Seq	left_de5f025bfed969c29fa2bdf3a2bad9c3bb2e70d1f78127bacdc5e3d044135831_2075	chr6	69500256	69500257	single junction	0		0	lgkv4-58	single junction
TC-Seq	left_7dc12cfdd9b0669a0576a1d6fbbb4c5aa48b48117006e23d8cd335bbc925a1d92_2006	chr6	69500256	69500257	single junction	0		0	lgkv4-58	single junction
TC-Seq	right_61b87737e84ca1a0fd5d1b38fa6ce69988ba7ad96dbeae8368d7e4346443299_1834	chr6	69500258	69500345	87	0		0	lgkv4-58	coding-end
TC-Seq	right_349ad2f9acfb159d6425851b46d40128b11c3e9774f269ce72d9a5e40e4d7296_1807	chr6	69500258	69500345	87	0		0	lgkv4-58	coding-end
TC-Seq	right_5ec497abd418a7419c256756 9ee0d6e9b43 44b2a93d16a8 1c48cc6c1861b729_1805	chr6	69500258	69500341	83	0		0	lgkv4-58	coding-end
TC-Seq	right_7a1cd1868ed0558597130f8fd04f228 44de1f4f15bb5bd6a0d6272015e2d835_1795	chr6	69500258	69500345	87	0		0	lgkv4-58	coding-end
TC-Seq	right_95ab98f4b6b07162194e2ad2a6f823373add58384fabca4ec71236dabeaae2cb_1848	chr6	69500258	69500345	87	0		0	lgkv4-58	coding-end
TC-Seq	right_0c07e6494616f56e95aeaf00a8baa23b8d53476548216 3a749de258c075136f_1859	chr6	69500258	69500345	87	0		0	lgkv4-58	coding-end
TC-Seq	right_73081c35f4b8f8a1dd732b0cf56be01cf069a3fe1a3775afca1a6461e1fc733b_1863	chr6	69500258	69500345	87	0		0	lgkv4-58	coding-end
TC-Seq	right_22ef155c553925 80c660fd3e3940dafc2874cadcc36c8f16d68e2cf8b3a55255_1874	chr6	69500258	69500341	83	0		0	lgkv4-58	coding-end
TC-Seq	right_d75b5aed6682e875 88d4ca703bc892a592ab3c29d6d6dca205d06f50e7c72be_1793	chr6	69500258	69500259	single junction	0		0	lgkv4-58	single junction
TC-Seq	right_a565bba51c3d1f30724f980f5926bc0b0e27e9e923191ef45c1c407dec3233a1_1848	chr6	69544359	69544360	single junction	0		0	lgkv4-57-1	single junction
TC-Seq	right_2a33072e7630d2b824d0c0d7500f7c2e54462b34036f1ba31983c20dddc20d02c7_1869	chr6	69544359	69544360	single junction	0		0	lgkv4-57-1	single junction
TC-Seq	right_52e055c99eea12f932f8dccd7b9ede6caca25c82b4a2c667 9d35d375db6476a6d_1884	chr6	69544359	69544360	single junction	0		0	lgkv4-57-1	single junction
TC-Seq	left_b31fb0e596ec0051df78d68c39657c893a5e6ef7dd8c292cbf2263c189e65dc48_2167	chr6	69544361	69544448	89	0		0	lgkv4-57-1	coding-end
TC-Seq	left_02409458bc161c02297a242a0047e259e4799349c3cd17187c05be18cd2281c9_1991	chr6	69544361	69544448	87	0		0	lgkv4-57-1	coding-end
TC-Seq	left_79b2fc82182 6b706fbf345640b6da6eef827051256ee7228be5b31a72f4c4620_1997	chr6	69544369	69544370	single junction	0		0	lgkv4-57-1	single junction
TC-Seq	right_3489c3c8995b4bd1060db482e5fa1ea937cdbcb2c85a73ae50030a7c70f64597_1847	chr6	69607275	69607362	87	0		0	lgkv4-55	coding-end
TC-Seq	right_f25321ea8ff49df86280e76309e4858dfb4f97be709143e682aa8166e8fc230b_1844	chr6	69607275	69607362	87	0		0	lgkv4-55	coding-end
TC-Seq	right_44f22759aeca9e955c374c2963ad5d983fc74659eb4ce57ada1601d70a177f47_1903	chr6	69607275	69607276	single junction	0		0	lgkv4-55	single junction
TC-Seq	right_2e745c64f4e7afe0fd17ab2e87d12a06ae929597c7bd3fcd62f605e0cc3fcd30_1897	chr6	69607275	69607276	single junction	0		0	lgkv4-55	single junction
TC-Seq	left_b3e79a79cdb104e2bf2ef5b4b90c817cce11159d22e5a9416864c5f909561 9d0_1986	chr6	69700767	69700856	89	0		0	lgkv4-50	coding-end
TC-Seq	left_ab9e1dbc7759d1a0c6f9d4c1abfce2f9fbe749db15ebb8519a34a43057008010_1979	chr6	69700767	69700856	89	0		0	lgkv4-50	coding-end
TC-Seq	left_57ea500a08152c7aca37ff4556255 8d4a14cb8292a2672e5787cd7fd7f5061b7_2029	chr6	69700771	69700859	88	0		0	lgkv4-50	coding-end
TC-Seq	left_3c81136 7e894cb318b86e4c6446 5d19d349b32d8cb63cd1abf7519e38e5467c7_2223	chr6	69700855	69700856	single junction	0		0	lgkv4-50	single junction
TC-Seq	left_e84ac02e5d66409b17e89a540a8ae9f85f2b046370735 0c0cc09ef2b349d148_2065	chr6	70074015	70074078	63	0		0	lgkv6-32	hybrid-end
TC-Seq	left_678d89e03809138f3c520bfdc0925a2bb38eb77b97ad728ee6ae1715e2c899cc_1989	chr6	70074015	70074078	63	0		0	lgkv6-32	hybrid-end
TC-Seq	left_ff869bb2d7401d52130cae1f30bbd40e2c8d3e340feb1ebf25906ddaf67bb7f2_2027	chr6	70117060	70117061	single junction	0		0	lgkv6-30	single junction
TC-Seq	left_0089bc66a7640c849c2e5cfbe667193d6bd1bef022ab7701f11233a376661c5b_2010	chr6	70117138	70117247	109	0		0	lgkv8-30	coding-end
TC-Seq	left_ee637d80babed8f031ed1fc1074cea9603ef3f1e97e2289af115f3fe7a415bba_2152	chr6	70314854	70314893	39	0		2	lgkv8-21	hybrid-end
TC-Seq	left_0008cdf621182c5b7ec685e77c1a4760db8564ff918f90a8adde1924a8424a28_2152	chr6	70314854	70314893	39	0		2	lgkv8-21	hybrid-end
TC-Seq	left_e07a27229b1ecc45568838cf5aac630cf5657fe8a2850b1d90e4205491b25cd2_2041	chr6	70703994	70704175	181	0		0	lgkv3-1	coding-end
TC-Seq	left_b96f358ccb6cdec4dabcefd0a8e7f7193579ed7d57c26660ebd9a74a4b1b647d_2057	chr6	70703994	70704175	181	0		0	lgkv3-1	coding-end
TC-Seq	right_d3b3fea94b837a08f3f1e0b299ae1816edff19256eee2be1e0a7c7cf285625759_1790	chr6	70703996	70703997	single junction	0		0	lgkv3-1	single junction
TC-Seq	right_73ce95d2ae4722d8c4f0c4f87267b756e09f821031d2fc8f6d819c4170c2c3a7_1981	chr6	70703997	70704177	180	0		0	lgkv3-1	coding-end
TC-Seq	left_3c53bc9c6667b46a63e3ac80c43dd2ca932183d4005a5ce8c5d6daad546f2c0a_1976	chr6	70703997	70704177	180	0		0	lgkv3-1	coding-end
TC-Seq	left_c763b2f44be78325b178ad1e1315a2639b063668ffb1d6ef4b7615460505 46ee61_1990	chr6	70703998	70704174	176	0		0	lgkv3-1	coding-end
TC-Seq	left_00c9b5e753 1315ffe50bb2e8333251fc7caa5de52b376ca62f0b5eee2f3eb40_1995	chr6	70703998	70704174	176	0		0	lgkv3-1	coding-end
TC-Seq	left_75d636fb03a0677f246fb72de13e41cab1093e5ee2c6d9595018ddbe3f6479360_1997	chr6	70703999	70704000	single junction	0		0	lgkv3-1	single junction
TC-Seq	left_2a56f163572d476cdee635eb209b4bd4b51a536e0adf98f5729ef841a3bd00673_2019	chr6	70703999	70704000	single junction	0		0	lgkv3-1	single junction
TC-Seq	left_e45fb906602240fa05d5ea09fb5054d620785b780ab450e58227c1918e93ddc00_2013	chr6	70704058	70704174	116	0		0	lgkv3-1	coding-end
TC-Seq	left_e07a77adeefadd813254e91f6b110949ceaad6236571b3e1a41c8986cadaeb47_2007	chr6	70704058	70704174	116	0		0	lgkv3-1	coding-end
TC-Seq	left_3600a4e03f9020cb981a6bb629b942e2253081 5a2fbd5794a698cc8dd11f8356f_2076	chr6	70704058	70704174	116	0		0	lgkv3-1	coding-end

TC-Seq	right_e79a1756ba0d10b85ddb600ac33514715bbb0170712ffb29eeb472836369b7_979	chr6	70704122	70704123	single junction	0		0	Igκv3-1	single junction
TC-Seq	left_8d29fe1784e5f6eb60a6774a4fc4b318f8e0dfa46e024159cd0dfef422dfed75_1986	chr6	70704171	70704172	single junction	0		0	Igκv3-1	single junction
TC-Seq	right_e20804963acbcb62762163816927c0d11e062ac76da254497f9a12a64ce9a8c0_1843	chr6	70704173	70704174	single junction	0		0	Igκv3-1	single junction
TC-Seq	left_fc786193b46412f53a869983ba9b57421a4dd0ebfafd68711c14533d7911c8a7_2027	chr6	70704176	70704177	single junction	0		0	Igκv3-1	single junction
TC-Seq	right_662657ede797f9de86c3c62bff2eadd44a1e50d53aa394be299751d0fe24a7da_1815	chr6	70704177	70704178	single junction	0		-1	Igκv3-1	single junction
TC-Seq	left_1144bfa2f2e21bc6e6d2739b4ab0860a3574fb952835e7f8a879b26b45348691_1979	chr6	70722561	70722915	354	0		0	Igκj1	hybrid-end
TC-Seq	left_dc6bd1577d634a9adec1aa939ff0ae9c3444013862d8e2456bdcd249085c4c11_1990	chr6	70723549	70723550	single junction	0		0	Igκj4	single junction
TC-Seq	left_f217d98c200d4cb9aa37f95cafc91e943486cd633c851b536fe18df786d56f0d_2034	chr6	70723550	70723631	81	0		0	Igκj4	hybrid-end

Igκ coordinates: chromosome 6, 67,555,636 - 70,726,966
insention sizes and species are not avaible for single junctions (only 1 defined breakpoint)

Table S4: Overview of *IG/TCR* insertions detected in human cancer

Study	Study ID	Type of cancer	Number of patients analyzed	Number of patients with insertions from *IG/TCR* loci	Origin of insertions
Holmfeldt et al., 2013	phs000341.v2.p1	hypodiploid acute lymphoblastic leukemia[BM]	17	1	Jα35
Zhang et al., 2012	phs000340.v3.p1	early T-cell precursor acute lymphoblastic leukaemia[BM]	11[C]	2	Jh6; Jα53[RE]
Okosun et al., 2014	EGAS00001000399	follicular lymphoma[LN, BM]	6[C]	2	Vα13-2; Vα38-2DV8[RE]

all analyzed samples are whole genome sequences of bone marrow ([BM]) or lymph node ([LN]) cells
[C] control samples were "in remission" and therefore included in the cancer analysis
[RE] insertion was detected in the remission sample of a patient that later relapsed

Table S5: Primer list

Name	Notes	Sequence
p2		CTCGAGTTATTAATCAAACAGTCTTCTAAGG
p3		CTCGAGTTATTATTCCTCTGAGTCTTCAAAGGGA
p4		GGATCCGCCACCATGGGATCAAGATCGCCAAAAA
p5		GTCTCCTGCTTGCTTTAACAGAGAGAAGTTCGTGGCTCCGCTTCCTTTCAGGAAAGTTTCGGAGGAG
p6		GCCACGAACTTCTCTCTGTTAAAGCAAGCAGGAGACGTGGAAGAAAACCCCGGTCCTATGTCCCTGCAGATGGTAACAG
p58	original: DHL, PMID 1900081	GGAATTCGMTTTTTGTSAAGGGATCTACTACTGTG
p96	original: JH4, PMID 1900081	TCCCTCAAATGAGCCTCCAAAGTCC
p98	original: VHQ52, PMID 1900081	CGGTACCAGACTGARCATCASCAAGGACAAYTCC
p113		TTGGGGGAAACCAGAGGGAATCC
p114		GGGAGGGGGTGTCAAATAATAAGAG
p195	original: Jk1, PMID 19467709	AATCAGCAGTTCTCTGTCAGAGAAGCC
p196		GCTACCCACTGCTCTGTTCCTC
p199		ACCTCATGTCAGATTTGTGGGAGAAATG
p200		ACTTAGCCTATCTAACTGGATCAGCCTC
p205	original: 3'Jk5.3, PMID 14581608	GCTTATTCTCCGATCCAATCTCTTGGATGG
p206	original: 3'Jk5.2, PMID 14581608	CACTGTATGCCACGTCAACTGATAATGAGC
p207		CATTGTGCTSACCCAATCTCCAGC
p210		CAAAGGAGGACGCTGAGAGTGG
p243		CAGGAGCCCAAGAAGCATCC
p244	poison primer	CCTCATCTATGCTGCATCCAACG
p245		TCAGTTGAGAATTCTTTGTTTGGCTCTAC
p247		TGAATGTAGCGGCCGGTTAGG
p251		ACTCCCTACTATCAGTGACGCTCG
p255	poison primer	AACTGGTCTCAGAAGCCTAAGACG
p256		AACCCTCCCTAGGTAGACAATTATCC
p257		CCTCCTTAACACCTGATCTGAGAATGG
p258	poison primer	AGGCTACCCTGCTTCTTTGAGC
p274a		CACTCTTTCCCTACACGACGCTCTTCCGATCTAGGAAGACTGCGGTGAGTCG
p274b		CACTCTTTCCCTACACGACGCTCTTCCGATCTCAGGAAGACTGCGGTGAGTCG
p274c		CACTCTTTCCCTACACGACGCTCTTCCGATCTACAGGAAGACTGCGGTGAGTCG
p274d		CACTCTTTCCCTACACGACGCTCTTCCGATCTTACAGGAAGACTGCGGTGAGTCG
p275a		GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATCCGATTCGAGCTCGC
p275b		GACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGATCCGATTCGAGCTCGC
p275c		GACTGGAGTTCAGACGTGTGCTCTTCCGATCTATGATCCGATTCGAGCTCGC
p275d		GACTGGAGTTCAGACGTGTGCTCTTCCGATCTCATGATCCGATTCGAGCTCGC
p305	original: 3'Jk5.1, PMID 14581608	GAACTGACTTTAACTCCTAACATGAAAACC
p306	original: Vk DEG, PMID 14581608	GGCTGCAGSTTCAGTGGCAGTGGRTCWGGRAC
pNextflex common		AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC
pNextflex index5		CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGC
pNextflex index6		CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACGTGTGC

all custom primers were designed using Geneious (Kearse et al., 2012)

Acknowledgements

The work presented herein was carried out in the Laboratory of Molecular Immunology at The Rockefeller University in New York from March 2012 - December 2016. In the following, I would like to express my sincerest gratitude to the people who supported me during this time:

My supervisor in New York, **Prof. Michel C. Nussenzweig**, for giving me the opportunity to join his laboratory for my doctoral (and also my master) thesis. It has been an incredible experience, both personally and professionally. I am very thankful for his continuous guidance and support throughout my entire project and beyond.

My supervisor in Stuttgart, **Prof. Monilola Olayioye**, for supporting me from abroad and for reviewing my final thesis in Germany.

Prof. Albert Jeltsch and Prof. Ralf Takors for kindly accepting the obligation to act as third and fourth reviewers.

Prof. Roland Kontermann for kindly accepting the obligation to act as chair of my doctoral defense and to examine my thesis.

Prof. Christina Wege, **Prof. Wolfgang Hauber** und **Prof. Günter Tovar** for kindly accepting the obligation to examine my thesis.

Associate Prof. Davide F. Robbiani for his guidance and support over the years. He co-supervised my project and also helped me with the final IC-Seq tissue culture experiments.

Thiago Y. Oliveira for helping me with the complex data analysis of my TC-Seq and IC-Seq libraries. Several aspects of this project would not have been possible without his help.

Klara Velinzon, **Yelena Shatalina** and **Neena Thomas** for helping me with FACS sorting.

David Bosque, **Thomas Eisenreich** and **Susan Hinklein** for helping me with the maintenance of my mouse colonies.

Connie Zhao of the Rockefeller Genomics Resource Center for helping me with the difficult sequencing of my TC-Seq and IC-Seq libraries.

Prof. Patricia Q. Cortes for providing me protocols and reagents for V(D)J PCRs.

All members of the Nussenzweig lab for helpful discussions and support during late-night hours.

My family for their continuous support despite the long distance and my frequent nightshifts during visits at home.

My family-in-law for always welcoming me in their home and making me a part of their family.

My beloved wife for her endless patience and loving support. I would have never made it without her help. #Team Cuesta-Rommel.

The **German Academic Exchange Service (DAAD)** for the financial support of my thesis.

Curriculum vitae

Philipp Rommel

Contact Information

Email	Philipp.Rommel@outlook.com or prommel@rockefeller.edu

Education

Expected Apr. 2017	**Ph.D. (Doktor rer. nat.) in Technical Biology**, University of Stuttgart, Stuttgart, Germany Grade: pending.
Feb. 2012	**Bachelor and Master (Diplom) in Technical Biology**, University of Stuttgart, Stuttgart, Germany Major subjects: Immunology and Cell Biology Minor subjects: Biochemical Engineering and Industrial Genetics (Nucleic Acid Technology) Grade: excellent.

Research and Industry Experience

Mar. 2012 - today	**Ph.D. thesis (Doktorarbeit) at The Rockefeller University**, New York, USA Laboratory of Molecular Immunology, **Prof. Michel C. Nussenzweig** *"RAG1/2 induces genomic insertions by mobilizing DNA into RAG1/2-independent breaks"* Grade: pending.
Feb. 2011 - Feb. 2012	**Master thesis (Diplomarbeit) at The Rockefeller University**, New York, USA Laboratory of Molecular Immunology, **Prof. Michel C. Nussenzweig** *"Influence of DNA End Resection on Antibody Class Switch Recombination"* Grade: excellent.
Nov. 2010 - Jan. 2011	**Internship at Roche Diagnostics**, Mannheim, Germany **Department of Marketing Applied Science and Molecular Diagnostics** Conducting market research and developing marketing strategies for the virology team, supporting sales agents with training programs and customer data.
Oct. 2009 - May 2010	**Bachelor thesis (Studienarbeit) at the Albert Einstein College of Medicine**, New York, USA Department of Biochemistry, Laboratory of **Prof. Marion Schmidt** *"Monitoring the Activity and Inhibition of the 20S Proteasome from S. cerevisiae"* Grade: excellent.
2006 - 2010	**Technician at the University of Stuttgart**, Stuttgart, Germany Working part-time as research assistant (Hilfswissenschaftler) in the Department of Animal Physiology, the Department of Bioinformatics and the Department of Biocatalysis.

Additional Skills

Languages	German (native language), English (fluent), Spanish (conversational) and French (basic).
Computer	Hardware support (expert), programming in Python (intermediate) and JavaScript (beginner).

Scholarships, Fellowships and Grants

Jan. 2014 - Dec. 2014	Doctoral fellowship from the German Academic Exchange Service (DAAD).
Oct. 2011	Travel grant from the German Academic Exchange Service (DAAD).
Feb. 2011	Travel grant from the Erwin Riesch Foundation.
Oct. 2009 - May 2010	Research scholarship for bachelor thesis (Studienarbeit) from the Heinrich J. Klein Foundation.

Publications

In press	**RAG1/2 induces genomic insertions by mobilizing DNA into RAG1/2-independent breaks.** **Rommel PC**, Oliveira TY, Nussenzweig MC, Robbiani DF *The Journal of Experimental Medicine*, in press.
Aug. 2015	**Plasmodium Infection Promotes Genomic Instability and AID-Dependent B Cell Lymphoma.** Robbiani DF, Deroubaix S, Feldhahn N, Oliveira TY, Callen E, Wang Q, Jankovic M, Silva IT, **Rommel PC**, Bosque D, Eisenreich T, Nussenzweig A, Nussenzweig MC *Cell*, 2015 Aug 13;162(4):727-37. PMID: 26276629.
Jul. 2013	**Fate mapping for activation-induced cytidine deaminase (AID) marks non-lymphoid cells during mouse development.** **Rommel PC**, Bosque D, Gitlin AD, Croft GF, Heintz N, Casellas R, Nussenzweig MC, Kriaucionis S, Robbiani DF *PLOS ONE,* 2013 Jul 8;8(7):e69208. PMID: 23861962.
Jan. 2013	**RPA accumulation during class switch recombination represents 5'-3' DNA-end resection during the S-G2/M phase of the cell cycle.** Yamane A, Robbiani DF, Resch W, Bothmer A, Nakahashi H, Oliveira T, **Rommel PC**, Brown EJ, Nussenzweig A, Nussenzweig MC, Casellas R *Cell Reports,* 2013 Jan 31;3(1):138-47. PMID: 23291097.
Jan. 2013	**Mechanism of DNA resection during intrachromosomal recombination and immunoglobulin class switching.** Bothmer A, **Rommel PC (shared first authorship)**, Gazumyan A, Polato F, Reczek CR, Muellenbeck MF, Schaetzlein S, Edelmann W, Chen PL, Brosh RM Jr, Casellas R, Ludwig T, Baer R, Nussenzweig A, Nussenzweig MC, Robbiani DF *The Journal of Experimental Medicine*, 2013 Jan 14;210(1):115-23. PMID: 23254285.
Dec. 2011	**Blm10 protein promotes proteasomal substrate turnover by an active gating mechanism.** Dange T, Smith D, Noy T, **Rommel PC**, Jurzitza L, Cordero RJ, Legendre A, Finley D, Goldberg AL, Schmidt M *The Journal of Biological Chemistry*, 2011 Dec 16;286(50):42830-9. PMID: 22025621.
Feb. 2010	**Simultaneous fluorescent monitoring of proteasomal subunit catalysis.** Wakata A, Lee HM, **Rommel P**, Toutchkine A, Schmidt M, Lawrence DS *Journal of the American Chemical Society*, 2010 Feb 10;132(5):1578-82. PMID: 20078037.

References

Ph.D. and master thesis advisor	**Michel C. Nussenzweig, M.D., Ph.D.** Investigator, Howard Hughes Medical Institute Senior Physician Zanvil A. Cohn and Ralph M. Steinman Professor The Rockefeller University, Laboratory of Molecular Immunology, New York, USA Michel.Nussenzweig@rockefeller.edu
Co-advisor	**Davide F. Robbiani, M.D., Ph.D.** Research Associate Professor The Rockefeller University, Laboratory of Molecular Immunology, New York, USA Davide.Robbiani@rockefeller.edu
Former colleague	**Michela Di Virgilio, Ph.D.** Group Leader, DNA Repair and Maintenance of Genome Stability Max-Delbruck Center for Molecular Medicine (MDC), Berlin, Germany Michela.DiVirgilio@mdc-berlin.de

Declaration of academic integrity (Eidesstattliche Erklärung)

I hereby assure that I performed this work independently without further help or other materials than stated. Passages and ideas from other sources have been clearly indicated.

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen, als die angegebenen, Quellen und Hilfsmittel benutzt habe. Wörtlich oder sinngemäß übernommenes Gedankengut habe ich als solches kenntlich gemacht.

Name (Name): Philipp Rommel

Signature (Unterschrift): _____

Date (Datum): _____