2012

The Extent and Nature Of Chromosomal Rearrangements in B Lymphocytes

Isaac Andrew Klein

The Extent and Nature Of Chromosomal Rearrangements in B Lymphocytes

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Isaac Andrew Klein

June 2012

The Extent and Nature Of Chromosomal Rearrangements in B Lymphocytes

Isaac Andrew Klein, Ph.D.

The Rockefeller University 2012

Chromosomal rearrangements, including translocations, require formation and joining of DNA double strand breaks (DSBs). These events disrupt the integrity of the genome and are frequently involved in producing leukemias, lymphomas and sarcomas. Mature B cell lymphomas are unique among tumors in that they frequently carry clonal recurrent translocations. This may be the result of Activation Induced Cytidine Deaminase (AID) expression, which introduces a heretofore uncharacterized array of rearrangements. Despite the importance of these events, current understanding of their genesis is limited.

To examine the origins of chromosomal rearrangements we developed Translocation Capture Sequencing (TC-Seq), a method to document chromosomal rearrangements to a fixed DSB genome-wide, in primary cells. We examined over 180,000 rearrangements obtained from 400 million activated B lymphocytes to two loci, *IgH* and *c-myc.* Our data and analysis reveal that

proximity between DSBs, transcriptional activity and chromosome territories are key determinants of genome rearrangement. Specifically, rearrangements tend to occur in *cis* and to transcribed genes. We also find that AID induces rearrangement in specific hotspots. Hotspots are predominantly genic, transcribed, and are found as translocation partners in mature B cell lymphoma.

Amplius est amplius

(More is More)

ACKNOWLEDGEMENTS

First, my deepest thanks go to Michel. Michel, you provided a kind word, direction and support. You had confidence in my abilities, encouraging me to be creative, focused, determined and independent. You provided an ideal scientific environment: high expectations, brilliant colleagues and enormous resources. I am truly grateful for your friendship and mentorship.

Thanks to all the members of the Nussenzweig lab for stimulating discussions, scientific contributions and their friendship. Special thanks to Davide Robbiani for countless hours of one-on-one teaching and continuing guidance. Thanks to Mila Jankovic for valuable input, a kind ear and a bottomless chocolate drawer; and to Anne Bothmer, Anna Gazumyan and Michela Di Virgilio for advice, expertise and lunch table discussions. Zoran Jankovic provided outstanding support as laboratory manager, as did Virginia Menendez as the lab's administrative assistant. Thanks to Tom Eisenreich and David Bosque for their assistance in managing the mouse colonies.

Thanks to the Rockefeller Dean's office and the Tri-Institutional MD/PhD Program Office for their administrative support. Special thanks to Ruth Gotian who keeps the MD/PhD program running smoothly and to Olaf Andersen, the director of the

MD/PhD program, for his leadership and for his dedication to and faith in his students. Thanks to those who have served on my thesis committee: Sohail Tavazoie, Scott Keeney, Bob Darnell and Nina Papavasiliou; and to my outside examiners David Roth and Jayanta Chaudhuri. David Roth is also a former mentor of mine. Dave, thank you for your guidance and encouragement; working with you cemented my desire to do science. I'd also like to thank Sergei Borukhov, my first mentor. Sergei was a patient and devoted educator and I sincerely thank him for kindling my interest in science.

This project was made possible through the efforts of many collaborators. Thanks to Klara Velinzon and Svetlana Mazel for FACSorting. Also to Scott Dewell of the Rockefeller Genomics Resource Center and Gustavo Gutierrez of the National Institute of Arthritis and Musculoskeletal and Skin Diseases genome facility for high-throughput sequencing and guidance, as well as Christopher Mason of the Weill Cornell Medical College for assistance with data analysis. Thanks to Wolfgang Resch and Thiago Oliveira for extensive data analysis, and to Arito Yamane and Hirotaka Nakahashi for assistance with hypermutation analysis. Thanks as well to Rafael Casellas for his insights, resources and assistance with this project. I am very grateful for excellent fellowship support from the NIH (MSTP grant GM07739), the Cancer Research Institute (CRI Predoctoral Fellowship) and The Hearst Foundation (William Randolph Hearst Foundation Fellowship). This work was also supported by the Howard Hughes

Medical Institute, National Institute of Health grant #AI037526, NYSTEM grant #C023046, the Starr Foundation and the Intramural Research Program of the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health.

Finally, I'd like to thank my friends and family. To my wife Ellie, you are my best friend, my partner and my love; thank you for your companionship and understanding during these years. To my parents, you have given me everything, the most important of which is an example to live by. Thank you for your unconditional love, faith and support, and for instilling me with a love of learning. I dedicate this work to you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1:

Introduction

Chromosomal Translocations

Chromosomal rearrangements are a variation in the normal structure of the genome. They may be the result of translocation, deletion, inversion or insertion; thereby uniting disparate parts of the genome with or without the deletion of intervening sequence. Rearrangements can be physiological, as in the antibody diversification reactions of B lymphocytes. Or they may pathological, as in the Philadelphia chromosome observed in chronic myelogenous leukemia.

The chromosomal translocation is a genomic rearrangement caused by the exchange of arms between two non-heterologous chromosomes. They are commonly found in leukemias, lymphomas and sarcomas and are key pathogenic events in cancer. Translocations can confer a survival or proliferative advantage upon an incipient cancer cell by: juxtaposing proto-oncogenes to constitutively active promoters, deleting tumor suppressors, or producing chimeric oncogenes (Nussenzweig and Nussenzweig, 2010).

Translocations are the most common class of cytogenetic abnormality in hematologic malignancies (Kuppers, 2005). These events are often clonal, indicative of their role in triggering or maintaining transformation and progression, and are highly stereotyped within tumor classes. Within the hematologic malignancies, there are classical examples of the functional and diagnostic implications of translocations. The Philadelphia chromosome, a translocation between chromosome 9 and 22 (t(9;22)(q34.1;q11.2)), is observed in chronic myelogenous leukemia. This rearrangement juxtaposes the 5' region of the *BCR* gene to the 3' region of the *Abl* gene generating a constitutively active tyrosine kinase, BCR-Abl, which drives tumor growth (Sawyers, 1993). The discovery of this fusion oncogene led to the successful development of diagnostic tests to identify the tumor, and the development of a drug to halt tumor growth in patients (Druker et al., 2001). Another example is Burkitt's lymphoma, a mature B cell cancer associated with t(8;14)(q24;q32). This rearrangement places the immunoglobulin heavy chain regulatory elements next to the *c-myc* proto-oncogene. *c-myc* is thereby over expressed, which leads to malignant transformation (Gostissa et al., 2009).

Translocagenesis requires that three primary criteria be fulfilled. First, paired DNA breaks must occur on different chromosomes (Richardson and Jasin, 2000). DNA damage may occur by exogenous modes such as ionizing radiation and chemical exposure, or by endogenous modes such as replication errors, cellular

metabolism and genetically programmed damage. Second, translocation partners must be in proximity to one another in the nucleus, a function of the three-dimensional architecture of the genome (Roix et al., 2003). Recent studies have demonstrated that chromatin features, transcriptional activity, and hormone exposure may influence genomic architecture generally, and the proximity of specific translocation partners (Lieberman-Aiden et al., 2009; Mani et al., 2009). Finally, DSBs on heterologous chromosomes must be joined in *trans* (as opposed to faithful repair of the DSB). This is largely a function of the DNA repair pathway responsible for break resolution; both pathway choice and abnormalities in repair pathways may influence the genesis of chromosomal translocations.

Origins of DNA Double Strand Breaks

The genome is under constant threat of DNA damage from both exogenous and endogenous sources. Damage may be the result of environmental insults, cellular metabolism, genomic features and development. DNA damage, if repaired improperly, may result in genomic rearrangement. Nearly every DNA damaging agent or event has been shown to lead to chromosomal rearrangement, and in many cases, with implications for oncogensis or other human diseases.

Oxidative stress. In a cell, reactive oxygen species come from multiple sources including UV light, aerobic metabolism, and enzymatic activity. Oxidative stress occurs when there is a greater abundance of these reactive oxygen species (ROS) than the cell's capacity to neutralize them. If such an overabundance occurs, free oxidants may react with DNA to generate adducts and numerous downstream products including DNA breaks (Marnett, 2000). Oxide radical production and the resulting DNA damage has been shown to cause genomic rearrangements. For example, the absence of a peroxiredoxin (an antioxidant enzyme) in yeast causes increased frequency of chromosomal rearrangements (Huang and Kolodner, 2005). And, in diseases characterized by DNA repair defects such as xeroderma pigmentosum and ataxia telangiectasia, ROS are hypothesized to play a role in the observed genomic instability (Barzilai et al.,

2002).

Ionizing radiation and chemical compounds. Both Ionizing radiation (IR) and certain chemicals may react directly with DNA, altering its chemical structure and resulting in damage. Ionizing radiation, defined as electromagnetic waves with sufficient energy to remove electrons from target atoms, causes double stranded DNA breaks (DSBs) at the rate of ~35 per Gy (Rogakou et al., 1998). IR-induced breaks are known precursors to rearrangement. For example, children exposed to radiation from the Chernobyl disaster suffer from high rates of papillary thyroid carcinoma, which bear rearrangements of the *RET* proto-oncogene (Fugazzola et al., 1995). *RET* rearrangements are detected in human thyroid tissue upon exposure to ionizing radiation, and their incidence in tissue culture is directly proportional to the dosage of radiation (Caudill et al., 2005).

Clinically, ionizing radiation is employed in radiotherapy, where a focused beam is used to introduce DSBs in cancer cells, thereby inducing cell death. Ionizing radiation is also an essential tool for studying DNA damage in the laboratory setting. It has been used extensively to characterize the nature of the DNA damage response and the retinue of proteins that bind sites of damage. However, its usefulness is limited by our inability to apply irradiation in a spatially controlled manner.

Certain chemicals may also cause DNA damage. Clinically, many chemotherapeutic agents act in this way, thereby selectively targeting quickly growing cells that undergo rapid replication. For example, alkylating agents such as cisplatin form covalent bonds with DNA bases, generating crosslinks that elicit the DNA damage response, induce apoptosis, and interfere with replication (Lawley and Phillips, 1996). Topoisomerase inhibitors on the other hand, directly cause DSBs. The enzyme topoisomerase II generates then religates DSBs to relieve tension caused by supercoils and tangles. Rapidly dividing cancer cells are particularly dependent on this enzyme to facilitate replication and transcription. The drug etoposide binds to topoisomerase II preventing the religation of broken ends, thereby generating DSBs. These DSBs may serve as substrates for translocation (Libura et al., 2005). For example, TOP2 inhibition has been shown to generate breaks in the mixed lineage leukemia (*MLL*) cluster region. Patients treated for acute lymphoblastic leukemia (ALL) with TOP2 inhibitors may thereafter suffer rearrangement of the *MLL* gene, which can lead to therapy-related acute myeloid leukemias (Blanco et al., 2001).

Replicative stress. During DNA replication, stalling of replication forks may be the result of sequence features such as repeat rich sequences, loci with complex secondary structure, nucleotide deficiency, assorted forms of DNA damage or collisions between replication and transcription machinery (Mani and Chinnaiyan, 2010). Normally, fork stalling leads to ssDNA generation that affects downstream

fork repair and checkpoint activation (Allen et al., 2011). An unrepaired replication fork however, may give rise to unusual DNA structures and collapse into a double strand break, a phenomenon that may also occur upon a fork encounter with a single strand nick (Ammazzalorso et al., 2010).

DNA breaks originating from replication fork collapse may serve as intermediates in the formation of complex chromosomal rearrangements, structures that involve reorganization of sequence from breakpoints on three or more chromosomes. It has been proposed that broken 3' ends from collapsed replication forks anneal to extra-chromosomal single stranded DNA based on homology. Annealed ssDNAs then polymerize, and in a series of template changes followed by the continuation of replication, complex chromosomal rearrangements are formed (Hastings et al., 2009). These structures have been observed in female patients suffering from multiple recurrent abortions and in male patients with hypospermatogenesis or spermatogenic arrest (Kim et al., 2011).

Common fragile sites. In 1984 Glover and collegues notes that when aphidicolin (an inhibitor of DNA polymerase) is added to a cell culture, chromosomes develop non-random gaps on metaphase spreads. These gaps became more frequent with folic acid deprivation and appeared to match gaps found by previous studies of spontaneous chromosomal damage (albeit at a much lower frequency). The breaks were present in 12 of 12 cultures tested so they were

termed common fragile sites (Glover et al., 1984), and they were attributed to conditions of partial replicative stress. About 80 of these sites are currently known, however, they are coarsely mapped; generally they are only localized to chromosome bands. A sequence feature is very likely to underlie this phenomenon, but the exact cause of their emergence is unknown. Common fragile sites examined to date are mainly AT-rich sequences (Zlotorynski et al., 2003). And, it has been hypothesized that fragile sites appear in late replicating regions of the genome (Dillon et al., 2010). It is also possible that fragile sites represent long regions of ssDNA that tend to form secondary structures, leading to replication fork stalling and collapse (Dillon et al., 2010).

DSBs resulting from common fragile sites have been implicated in the formation of chromosomal rearrangements. In a study that examined all known cancer-associated recurrent translocations, 52% occurred within known fragile site positions (Burrow et al., 2009). In one specific case, the most frequently expressed fragile site in the human genome, *FRA3B*, is located within the tumor suppressor gene *FHIT* (Huebner and Croce, 2001). Deletions of *FHIT* are associated with cancer of the breast, lung, cervix, and esophagus.

Transcriptional stress. Transcription induces a multitude of changes on the DNA. It may change DNA conformation, strandedness, position and the local milieu of proteins. Additionally, it may interact with existing processes on the DNA such as

replication or DNA repair and with lesions such as altered bases. As a consequence of these changes and interactions, transcription can compromise genomic stability in the form of increased mutation and recombination rates.

The phenomenon of transcription-associated mutation was first described 40 years ago by studies in prokaryotes. In E. coli, alkylating agent induced mutation of the β-galactosidase gene is stimulated by transcription (Brock, 1971). And, in the presence of acridine (a mutagen) *lac* null mutations are reverted more frequently in the presence of transcription (Herman and Dworkin, 1971). The persistence of ssDNA at the transcription bubble may account for this instability; many mutagenic reactions are more efficient on ssDNA than dsDNA (Aguilera, 2002). For example, in bacteria, spontaneous deamination of cytosine is 140-fold more efficient on ssDNA than on dsDNA (Frederico et al., 1990). And, in E. Coli mutation of *LacI* and human *hprt* is more prominent on the non-transcribed strand (which is open ssDNA) as compared to the transcribed strand (which is in complex with RNA and multiple proteins) (Fix and Glickman, 1987). Finally, with an RNA polymerase mutant characterized as having a slow elongation rate transcription-associated mutation is higher than with a wild type polymerase. This may suggest that mutations rates depend on the length of time that DNA is held in an open, single stranded conformation (Beletskii et al., 2000).

In addition to increasing mutation rates, transcription appears to boost

recombination rates as well. Evidence for transcriptionally active recombinogenic DNA in eukaryotes was first described in the context of recombination hotspots, high levels of recombination were observed between nearby genes in the hotspot *HOT1* (in the ribosomal RNA genes) (Keil and Roeder, 1984). Later work showed that transcription by RNA polymerase I was responsible and required for the recombination (Voelkel-Meiman et al., 1987). Eventually, this phenomenon was shown in regions transcribed by RNA Polymerase II; transcription enhanced recombination 15 fold between directly repeated sequences of *GAL10* in S. cerevisiae (Thomas and Rothstein, 1989).

In 1990, Nickoloff and colleagues provided the first evidence for transcription associated recombination in mammalian cells. They found that transcription between heteroallelic neomycin genes on extrachromosomal plasmids was significantly stimulated in the presence of transcription (Nickoloff and Reynolds, 1990); the same effect was shown with intra-chromosomal recombination of integrated substrates (Nickoloff, 1992). Interestingly, transcription does not appear to increase the frequency of DSB induced recombination and does not change the spectrum of recombination events from a DSB. For example, when an I-SceI break is introduced in a direct repeat of a *neo* gene, transcription does not change the frequency of recombination between repeats, the repair pathway used or the size of deletions at the break (Allen et al., 2003). This suggests that transcription may be enhancing the generation of recombinogenic lesions, not

enhancing the recombinogenicity of existing lesions. Of note, transcription enhances the use of a donor sequence in DSB-induced recombination (Schildkraut et al., 2006).

There are several theories to explain the phenomenon of transcription-associated recombination. First, transcription may expose ssDNA that is more prone to damage and thereby, recombination. Given the propensity of transcription-associated mutations to occur on the exposed single strand of the transcription bubble, one might reason that this is a likely mechanism for recombination as well (Aguilera, 2002). In cells treated with a carcinogen 4-nitroquiniline-N-oxide transcription appears to synergize with DNA damage to drive recombination frequencies distantly above background levels (>4000 fold) (Garcia-Rubio et al., 2003). This suggests that transcription indeed increases the availability of DNA, and perhaps ssDNA, for damage and recombination. However, to our knowledge it has never been directly shown that recombination is more likely to occur to regions of ssDNA.

Another non-mutually exclusive possibility is that transcription associated recombination might be caused by replication fork/transcription bubble collisions. Replication and transcription can take place on the same strand at the same time; these processes might obstruct one another, resulting in stalled and collapsed replication forks. These events may result in rearrangement. For

example, when replication and transcription are initiated in opposite directions on an extrachromosomal substrate in E coli, up to 12 percent of the isolated plasmid sustains deletions (Vilette et al., 1996). In this situation, the replication fork has been shown to stall for several minutes, which may result in collapse and ensuing damage (Liu and Alberts, 1995). Supporting this model of transcription associated recombination Gottipati and colleagues have shown that the phenomenon tends to occur in S phase, suggesting that it is dependent on replication (Gottipati et al., 2008)

Genetically programmed DSBs. During T and B lymphocyte development, antigen receptor diversification occurs by virtue of specific DNA breaks and recombination events. The enzymes responsible for generating these breaks may contribute to genomic instability via promiscuous joining of physiological DSBs and generation of off-target DSBs. V(D)J recombination, catalyzed by a site-specific nuclease RAG1/2, generates antigen receptor variable region diversity by combining gene segments in myriad permutations. RAG1/2 introduces nicks at recombination signal sequences that convert to double strand breaks by transesterification (Roth, 2003). Unresolved ends from RAG breaks in the immunoglobulin loci may partner with other DSBs to generate oncogenic rearrangements, as in the *bcl-2/IgH* translocation observed in follicular lymphoma (Jager et al., 2000). Additionally, RAGs can function as a transposase, which may be a means for rearrangement production (Brandt and Roth, 2004). They

may also introduce breaks in various non-antigen receptor loci including those bearing non-B DNA structures or cryptic recombination signal sequences (Raghavan et al., 2001; Raghavan et al., 2004). However, the extent of these non-Ig breaks and their locations remain unknown.

Mature B cells suffer genomic insult from two additional diversification reactions catalyzed by Activation Induced Cytidine Deaminase (AID). Somatic hypermutation introduces non-templated mutations in the immunoglobulin genes during B cell activation in the germinal center. Class switch recombination changes the immunoglobulin constant region in a deletional recombination reaction catalyzed by double strand breaks in switch regions. In addition to acting on physiological targets, AID may generate DSBs throughout the genome (Nussenzweig and Nussenzweig, 2010). V(D)J recombination, somatic hypermutation and class switch recombination will be discussed in greater detail in later sections.

DNA Repair Pathways in Translocagenesis

Maintaining genomic integrity is key to cellular survival and proliferation. As such, multiple DNA repair pathways have evolved to recognize and repair DNA damage. Depending on the repair pathway, DNA breaks may be rejoined faithfully or may generate rearrangements. Additionally, defects in a pathway may result in unfaithful break repair. There are three major DNA repair pathways thought to repair DSBs; homologous recombination (HR), non-homologous end joining (NHEJ) and alternative non-homologous end joining (Alt-NHEJ).

Homologous recombination. HR is a primary pathway in the S and G2 phase of the cell cycle, during which time it repairs DSBs by using a sister chromatid template. In this pathway DNA ends must first be resected by exonucleases, leaving a 3' single stranded end. Exposed ssDNA is then bound by Replication Protein A (RPA) and Rad51 to form a nucleoprotein filament. This process also requires the MRN complex, BRCA1 and BRCA2. The filament carries out a homology search eventually forming a displacement loop on a homologous sequence. The invading strand then copies sequence from the template strand. Repair can result in crossover or non-crossover repair products (Moynahan and Jasin, 2010).

Homologous recombination plays an important role in maintaining genomic integrity; if the pathway is fractured cells suffer an increase in genomic instability (Sonoda et al., 1998). Specifically, in the absence of HR cells exhibit an increase of double strand breaks at fragile sites (Schwartz et al., 2005). Loss of certain HR components may also result in the accumulation of genomic abnormalities. For example, the loss of BRCA1 or BRCA2 leads to chromosomal radial structures, which are fusions between different chromosomes (Patel et al., 1998). Radial structures may break during mitosis resulting in breakage-bridge fusion cycles that give rise to translocations. Finally, loss of BRCA1 function is observed in inherited breast cancer, indicating the importance of this pathway in preventing oncogensis (Ford et al., 1994).

Non-homologous end joining. NHEJ is the so-called “error prone” repair pathway, which may occur throughout the cell cycle, but predominantly in G1 (Takata et al., 1998). Breaks are repaired by simply reattaching the ends, rather than copying from a homologous strand, so bases may be lost or mutated at the junctions. Briefly, in this pathway Ku70-Ku80 heterodimers bind the broken ends, which recruits DNA dependent protein kinase catalytic subunit (DNA-PKcs) to form a synaptic complex. Thereafter, the LigaseIV-Xrcc4 complex religates the broken ends (Lieber, 2008). Ataxia-telangiectasia mutated protein (ATM), a serine/threonine kinase, is responsible for phosphorylating numerous DNA repair

effectors, cell cycle checkpoint proteins and for amplifying the DNA damage signal (Derheimer and Kastan, 2010).

The classical NHEJ factors are required for faithful DSB repair, V(D)J recombination, and for efficient class switch recombination (Manis et al., 2002; Nussenzweig et al., 1996; Zhu et al., 1996). In the absence of NHEJ factors, there is a dramatic increase in persistent DSBs and chromosomal rearrangements (Difilippantonio et al., 2000). So, NHEJ suppresses aberrant rejoining events, including translocations, while encouraging appropriate intra-chromosomal repair. Consistent with this idea, translocations generated by RAGs, AID and homing endonucleases do not require the NHEJ factors Ku70, Ku80, DNA-PKcs or XRCC4, rather, translocation formation is increased in their absence (Boboila et al., 2010a; Callen et al., 2009; Wang et al., 2009; Weinstock et al., 2007). An intact NHEJ pathway is also required to suppress oncogensis. For example, mice lacking p53 and Ku80 develop early onset pro-B cell lymphomas characterized by recurrent chromosomal translocations (Difilippantonio et al., 2000).

Alternative non-homologous end joining. Alt-NHEJ is a poorly defined pathway, but is likely of great importance in generating chromosomal rearrangements. Alt-NHEJ is primarily characterized by presence of microhomology (~>4 nucleotides) and deletion at the repaired junction, beyond what is normally observed at

junctions generated by the NHEJ pathway (Lieber, 2008). The presence of microhomology at junctions was first described in 1986 (Roth and Wilson, 1986). Since then, the existence of a specific pathway has been demonstrated (Yan et al., 2007). While Alt-NHEJ is most robust in cells with a disabled classical NHEJ, it appears to perform repair even in the presence of an intact NHEJ pathway (Corneo et al., 2007). Several proteins have been implicated in Alt-NHEJ, including Mre11 and CtIP, which are notable for their involvement in end resection and ssDNA generation (Bennardo et al., 2008; Deriano et al., 2009; Lee-Theilen et al., 2011).

Alt-NHEJ appears to play a role in chromosomal translocation. In the absence of NHEJ factors Ku70 and Ligase4 or Xrcc4 and Ligase4, breaks are more likely to partner in translocations, and the junctions of those events bear microhomologies – indicating the role of alt-NHEJ in their formation (Boboila et al., 2010a; Simsek and Jasin, 2010). Additional evidence can be found in a mouse model of RAG-mediated lymphomagenesis. Deriano and colleagues generated mice that lack p53 and express a truncated form of RAG2, coreRAG2. coreRAG2 releases DNA ends after cleavage, allowing them to be accessed and repaired by the alt-NHEJ pathway (Corneo et al., 2007). These mice develop lymphomas bearing chromosomal rearrangements showing: the role of alt-NHEJ in generating genomic abnormalities, the importance of those abnormalities in

lymphomagenesis, and the role classical NHEJ plays in suppressing rearrangements and oncogensis (Deriano et al., 2011).

Spatial Proximity in Translocagenesis

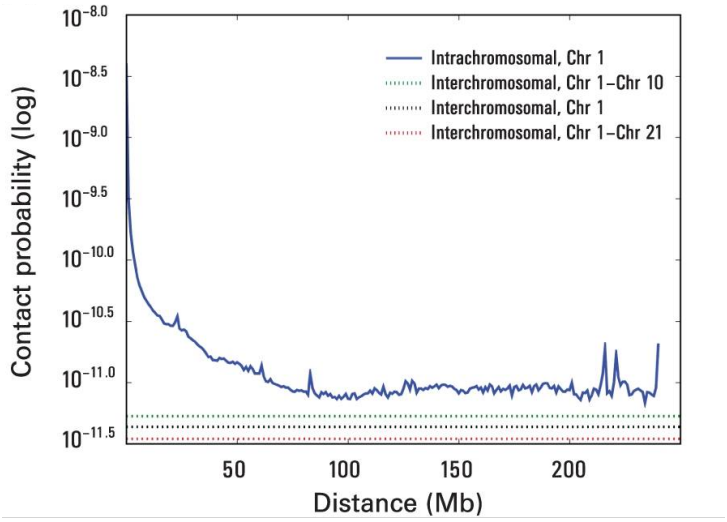
To form a rearrangement DSBs must be in proximity to one another. The proximity of two loci may be impacted by several factors including the constitutional architecture of the genome, transcription, chromatin accessibility and signaling pathways. While the 3D structure and dynamics of the genome are still emerging there appear to be certain principles dictating its form and, they may exert effects on the genomic rearrangement landscape.

1st order genomic architecture: chromosome territories. An organization of interphase chromosomes was first proposed in the 19th century, and Theodore Boveri coined the term chromosome territory in 1909. In Boveri's studies of the horse roundworm blastomere he noted that chromosome ends protruding from the nucleus maintained their position as the cell transitioned from prophase, during which individual chromosomes are visible, to interphase. He concluded that the chromosome territory is stably maintained during the cell cycle (Cremer and Cremer, 2010). Subsequent light microscopy studies supported these assertions (Stack et al., 1977), but in electron micrographs of the nucleus chromosome territorial integrity appeared to dissolve, so the idea was largely forsaken.

Recent studies have resurrected the concept of chromosome territories. Direct visual evidence was first obtained by in situ hybridization on total human DNA in cell hybrids containing only one human chromosome on a background of Chinese hamster ovary chromosomes (Manuelidis, 1985). Indeed, individual human chromosomes could be delineated during interphase. The advent of 3D fluorescent in situ hybridization and specific probes has allowed higher resolution studies of chromosome territories. These have confirmed their existence. For example, differential coloring of all 24 chromosomes in human fibroblasts followed by serial sections of laser confocal microscopy images showed excellent separation of distinct chromosome territories (Bolzer et al., 2005).

High throughput DNA sequencing has provided yet another tool for examining nuclear structure. In 2009 Lieberman-Aiden and colleagues developed Hi-C (High throughput chromosome conformation capture). In Hi-C, cells are cross-linked (to preserve long-rage chromatin interactions) then DNA is digested with restriction endonucleases. Ends are filled in with biotinylated nucleotides and the resulting blunt fragments are ligated under dilute conditions that promote ligation between cross-linked DNA fragments. The result is a pool in which fragments of DNA residing in close nuclear proximity are ligated and marked with a biotin. A high throughput paired-end library is assembled from biotinylated fragments, sequenced, and a catalog of nuclear interactions is developed. This study showed that intra-chromosomal contacts are more likely than trans-chromosomal

contacts, indicating the presence of chromosome territories (Figure 1.1) (Lieberman-Aiden et al., 2009).



Adapted from Lieberman-Aiden et al., 2009

Figure 1.1. Hi-C data indicates the presence of chromosome territories. The probability of intra-chromosomal contact between DNA fragments (solid line) exceeds the probability of trans-chromosomal contact (dotted lines) at any distance between fragments mapping to the same chromosome (Lieberman-Aiden et al., 2009).

The existence of chromosome territories may impact the rearrangement landscape across the genome. A 2009 report from Mahowald and colleagues demonstrated a system for cloning breakpoints from aberrantly resolved RAG

DSBs generated during V(D)J recombination in *Atm*$^{-/-}$ Abelson transformed pre-B cells. They showed that 60% of rearrangements occur to points on the same chromosome as the recombination substrate, rather than to trans-chromosomal targets. And, of these intra-chromosomal targets 87% were within 200kb of the recombination substrate. This study suggests that intra-chromosomal joining may be preferred to trans-chromosomal joining, however, they examined a small number of rearrangements limiting the certainty of the results (Mahowald et al., 2009). Firm proof that chromosome territories influence the landscape of rearrangements is still lacking, and whether this phenomenon occurs locally or over the entire chromosome is unknown. For this issue to be addressed, a much greater quantity of data is required in primary cells.

2nd *order genomic architecture: functional compartments.* In the interphase nucleus perhaps the most obvious forms of nuclear organization are the heterochromatic and euchromatic compartments. These compartments, which are easily visible by light microscopy, represent transcriptionally active open chromatin and transcriptionally inactive closed chromatin. In eukaryotes, heterochromatin tends toward the nuclear envelope while euchromatin lies centrally in the nucleus, and this organization relates to gene activation. For example, immunoglobulin loci only active in pro-B cells are positioned at the nuclear periphery in pro-T cells and hematopoietic precursors, but are centrally located in pro-B cell nuclei (Kosak et al., 2002). And, active and inactive alleles of

a mono-allelicly expressed astrocyte specific protein, *GFAP*, occupy different subnuclear compartments and radial positions in the nucleus depending on their functional status (Takizawa et al., 2008). These observations suggest that transcriptional activity is a mediator of genome organization.

Hi-C revealed transcription- and chromatin conformation-dependent levels of genome organization. First, gene rich chromosomes interact with one another. This suggests that in regard to chromosome relationships, genes influence the nuclear position. Hi-C data analysis also revealed two general spatial compartments within the nucleus. The presence of genes, hypersensitivity to DNAseI and high transcription levels (as measured by mRNA levels) accounted for the differences between these two compartments. Thus, the genome consists of two neighborhoods; active and inactive loci (Lieberman-Aiden et al., 2009). Whether this level of organization impacts the genomic rearrangement profile or the joining behavior of specific breaks remains unclear. Some bias for genic rearrangements has been demonstrated in deep sequencing experiments on human tumors (as will be discussed at the end of this section). However, rearrangements in tumors may be selected or a function of some cancer specific joining aberration, so it is difficult to draw general conclusions about the nature of rearrangements from these studies.

Proximity of specific partners. The importance of proximity in rearrangement propensity is demonstrated by studies of specific oncogenic rearrangement partners. Average proximity of loci correlates with their rearrangement frequency, indicating that proximity is a determinant of translocation. For example, Burkitt's lymphoma is characterized by translocations involving c-myc. Clinically, c-myc fusion to the immunoglobulin heavy chain (*IgH*) is common, while fusion to the immunoglobulin light chain kappa (*IgK*) is infrequent (Roix et al., 2003). In experiments using FISH in a cytogenetically normal B-lymphoblast cell line, Roix and colleagues demonstrated that *c-myc* is in closer proximity to *IgH* than *IgK* (Roix et al., 2003). In the same study, similar relationships were described for other translocation partners. This suggests that proximity might dictate rearrangement propensity.

Many rearrangements are cell-type specific. The proximity of specific translocation partners is also cell-type specific and related to the occurrence of rearrangement. For example, he *BCR-Abl* fusion of chronic myelogenous leukemia occurs in an early hematopoietic stem cell while the promyelocytic leukemia protein (*PML*) – retinoic acid receptor alpha (*RAR-alpha*) fusion of acute promyelocytic leukemia occurs in committed hematopoietic precursors (Mani and Chinnaiyan, 2010). In a population of cells containing early hematopoietic stem cells, *BCR* and *Abl* are in close proximity while *PML* and *RAR-alpha* are not (Neves et al., 1999). This is further evidence that proximity is

a determinant of rearrangement and suggests that the cell-type specificity of rearrangements may be a function of partner locus proximity.

Signaling pathways within an individual cell type may also influence the proximity of partner loci, and thereby effect rearrangement generation. This phenomenon can be observed in the *ETS* gene rearrangements that fuel prostate cancer. These rearrangements yield fusions of the 5' UTR of an androgen-regulated gene with *ETS* family genes (*ERG, ETV1, ETV4, ETV5*); of these, *TMPRSS2-ERG* is the most common (Helgeson et al., 2008; Lin et al., 2009). Treating human prostate cancer cells with dihydrotestosterone (DHT) induces proximity between these partners by a myosin and actin dependent mechanism (Mani et al., 2009). When subjected to irradiation, these cells then form *TMPRSS2-ERG* rearrangements. The authors of this study conclude that signaling pathways can effect the position of certain rearrangement partners. This phenomenon may play an important role in the formation of recurrent chromosomal aberrations.

Next Generation Analysis of Rearrangements in Tumors.

Next generation sequencing technologies have facilitated the in depth examination of cancer genomes. These technologies enable the generation of transcriptome, methylome, genome and epigenetic maps. Of particular relevance to this study, cancer genome sequences have provided a high-resolution view of some cytogenetic abnormalities in human cancer. While the events sequenced from tumors are highly selected, and may be formed by aberrant cancer-cell specific mechanisms, the landscape of rearrangements in cancer may reveal some principles of chromosomal rearrangement.

Rearrangements in tumors may be examined by transcriptome sequencing of cDNA. This limits the number of reads required to achieve complete coverage, and thereby limits the cost of analysis; however, only a small fraction of the genome can be viewed. In a study of the breast cancer cell line HCC1954 this technique was successfully employed to discover novel rearrangements. By mining the sequencing data for chimeric reads (those mapping to two disparate loci), Zhao et al discovered 7 cDNAs representing 6 transchromosomal events and 1 intrachromosomal event, all of which could be confirmed by FISH, genomic DNA PCR or cDNA PCR (Zhao et al., 2009). Some events appeared to have oncogenic significance. For example, they found truncating translocation of *MRE11A*, a protein involved in DNA repair. And, a translocation of *NSD1* (a gene

fusion partner in acute myeloid leukemia) was found to an intergenic sequence on chromosome 8 (Zhao et al., 2009). This approach can successfully discover some cytogenetic abnormalities. Unfortunately, the limited number of rearrangements obtained does not provide sufficient power to make any conclusions about the global rearrangement profile of this line or the general principles guiding rearrangement formation.

Whole genome paired-end sequencing provided a more comprehensive analysis of rearrangements. Paired-end genome sequencing involves reading both sides of all randomly generated 200 bp genomic DNA fragments, thereby providing a complete genome sequence as well as the location of any structural variations. In the sequences of 2 lung cancer cell lines, Campbell and colleagues discovered 103 somatically acquired rearrangements. Having such a number of rearrangements allowed the extraction of some general trends: rearrangements tended to originate from amplicons and rearrangements predominantly occur intra-chromosomally. Of the 103 events isolated, 81 were intra-chromosomal (Campbell et al., 2008). A subsequent study that captured 2,166 rearrangements by analyzing 24 breast cancer genomes added additional power to these analyses. Of more than 2,000 events, 1,708 were intra-chromosomal, and of these, 1,574 showed breakpoints within 2Mb of one another (Stephens et al., 2009). Paired-end sequencing of metastatic pancreatic cancer revealed a similar bias toward intra-chromosomal events (Campbell et al., 2010). This may suggest

that a) linear physical proximity encourages rearrangement, and b) chromosome territories might be a determinant of rearrangement. However, since these rearrangement profiles are from tumors, they may not be generalized to a normal cell.

The breast cancer genome sequence also revealed that 50% of rearrangements occur in genic regions (Stephens et al., 2009). This is significantly more than expected to occur in genic regions, which compose just 40% of the genome. However, rearrangements near genes are likely to alter gene expression or function and impact proliferative or metabolic capacity. So, whether this finding represented a bona fide predisposition to intragenic rearrangement or a consequence of selection could not be determined.

Analysis of the rearrangement junctions from breast cancer genomes showed that the majority had minor junctional microhomology, and 5-15% of the junctions used greater than 4 nucleotides (Stephens et al., 2009). This implicates both NHEJ and alt-NHEJ as the primary pathways generating rearrangements based on microhomology utilization. However, nucleotide deletion from a break is another important parameter in the joining process, and in this approach the extent of end deletion cannot be assessed because the precise breakpoints were unknown.

Mature B Cell Lymphomas

The yearly incidence of lymphoma is about 2 per 10,000; this represents 5% of all cancers in the US. About 95% of these cases are cancers of mature B cell origin (Kuppers, 2005). While nearly all cancers carry cytogenetic abnormalities, the mature B cell cancers distinguish themselves by their tendency to carry clonal, recurrent translocations that drive tumor growth. This predisposition to oncogenic rearrangement may be due to a relatively high DNA damage burden. Mature B cells divide rapidly and so are prone to an increased frequency of usual genomic insults. Perhaps more significantly, mature B cells express Activation Induced Cytidine Deaminase (AID) that normally introduces DNA damage in immunoglobulin genes to facilitate the antibody diversification reactions class switch recombination and somatic hypermutation. However, AID activity may also yield inappropriately joined physiological DSBs, off-target DSBs and off-target mutations.

Mature B Cell Lymphoma Origins and Pathogenesis

Developed B cells participate in the humeral immune response by producing antibodies to recognize and bind antigen. However, B cells must first complete a complex developmental program that includes de novo antigen receptor construction, activation, and affinity maturation. Development begins in the bone marrow with V(D)J recombination, which assembles the immunoglobulin heavy and light chain genes to form an antigen binding B cell receptor, the precursor to a secreted antibody. This process is performed at the DNA level and involves permanent alteration of the B cell genome. After presentation of a functional B cell receptor (BCR) on the surface, mature naïve B cells leave the bone marrow and home to the peripheral lymphoid organs where they form germinal centers. In germinal centers, mature naïve B cells may be activated by T cell dependent or independent mechanisms to expand, express AID, and undergo the immunoglobulin diversification reactions somatic hypermutation (SHM) and class switch recombination (CSR) (Durandy, 2003). These reactions alter the affinity of the BCR for its cognate antigen and change the effector function of the resulting antibody, respectively. B cell cancers can arise at any stage of development.

Origins of mature B cell lymphomas. B cell lymphomas can be traced to a cellular origin based on the B cell developmental stage they most closely resemble (Kuppers, 2005). Mature B cells lymphomas are so called because they originate

from B cells at this late stage of development, and therefore, show many of the characteristics of mature B cells. For example, Burkitt's lymphoma cells strongly express usual germinal cell B cell surface markers CD19, CD20 and CD22, while follicular lymphoma grows in follicles reminiscent of the germinal center morphology (Bellan et al., 2010; Carbone et al., 2009). In fact, the majority of mature B cell lymphomas are thought to derive from germinal center or post-germinal center cells. Consistent with this, the mature B cell lymphomas follicular lymphoma, Burkitt's lymphoma, diffuse large-cell lymphomas, MALT lymphoma and multiple myeloma all have hypermutated variable regions, indicating AID activity (and thereby germinal center residence) during or before malignant transformation (Kuppers et al., 1999). Some of the mature B cell lymphomas, such as follicular lymphoma, even exhibit ongoing somatic hypermutation, bolstering the case for their germinal center parentage (Bahler and Levy, 1992).

More recently, gene expression profiling has supplanted more traditional modes of tumor characterization such as selected cell surface markers and histological appearance. By profiling normal B cell subsets along with tumors then comparing their gene signatures, the origins of these lymphomas can be further elucidated. Using this method, follicular lymphoma, Burkitt's lymphoma and most diffuse large B cell lymphomas were shown to have a GC B cell gene expression signature (Alizadeh et al., 2000). Interestingly, a subset of diffuse large cell lymphoma (the activated B like DLBCL) most closely resembled *in vitro* activated

B cells. It should be noted that malignant transformation might change gene expression profiles and mask the true origin of tumors. However, together with the observation that many of these tumors carry mutated variable regions, these data strongly suggest that the mature B cell lymphomas indeed originate from GC or post-GC cells.

Genetic lesions in mature B cell lymphoma. B cell lymphomas are unique among cancers in that they almost all carry chromosomal translocations. Most of these rearrangements involve one of the immunoglobulin loci and an oncogene (Table 1.1). Consequently, the oncogene is brought under control of the immunoglobulin transcriptional regulatory elements thereby dysregulating its expression. For example, in diffuse large B cell lymphoma the *IgH* regulatory elements may partner with the *BCL2* locus to form the *IgH/BCL2* translocation. The BCL2 protein (an antiapoptosis factor) is constitutively expressed, which prevents cell death (Weiss et al., 1987) In Burkitt's lymphoma *c-myc* become constitutively expressed and drives uncontrolled cell proliferation (Dalla-Favera et al., 1983). In the mature B cell lymphomas unregulated expression of an oncogene is frequently coupled with deletion or inactivation of a tumor suppressor. In other words, a translocation provides cancer's accelerator, while deletion or mutation takes away the brakes. For example, in mantle cell lymphoma ATM is frequently mutated, this disables the DNA damage response and prevents the cell from blocking cell cycle progression in the case of genomic insult (Camacho et al.,

2002). And, in Burkitt's lymphoma p53 is mutated in nearly 50% of the cases (Gaidano et al., 1991).

Table 1.1. Genomic abnormalities in mature B cell lymphomas

Lymphoma	Chromosomal Translocation	Tumor suppressor gene mutations
Mantle cell lymphoma	*CCND1-IgH (95)*	*ATM (40)*
Follicular lymphoma	*BCL2-IgH (90)*	
Diffuse large B cell lymphoma	*BCL6-various (35)* *BCL2-IgH (30)* *c-myc-IgH/IgL (15)*	*CD95 (20)* *ATM (15)* *p53 (25)*
Burkitt's lymphoma	*c-myc-IgH/IgL (100)*	*p53 (40)* *RB2 (20-80)*
MALT lymphoma	*API2-MALT1 (30)* *BCL10-IgH (5)* *MALT1-IgH (20)* *FOXP1-IgH (10)*	*CD95 (80)*
Lymphoplasmacytoid lymphoma	*PAX5-Igh (50)*	
Multiple myeloma	*CCND1-IgH (20)* *FGFR3-IgH (10)* *MAF-IgH (10)*	*CD95 (10)*

Table 1.1. Genomic abnormalities in mature B cell lymphomas. Each mature B cell lymphoma subtype is shown with its associated chromosomal translocation(s) and tumor suppressor gene mutation(s). In parentheses are the percent of all cases in which the abnormality is estimated to exist. Adapted from (Kuppers, 2005).

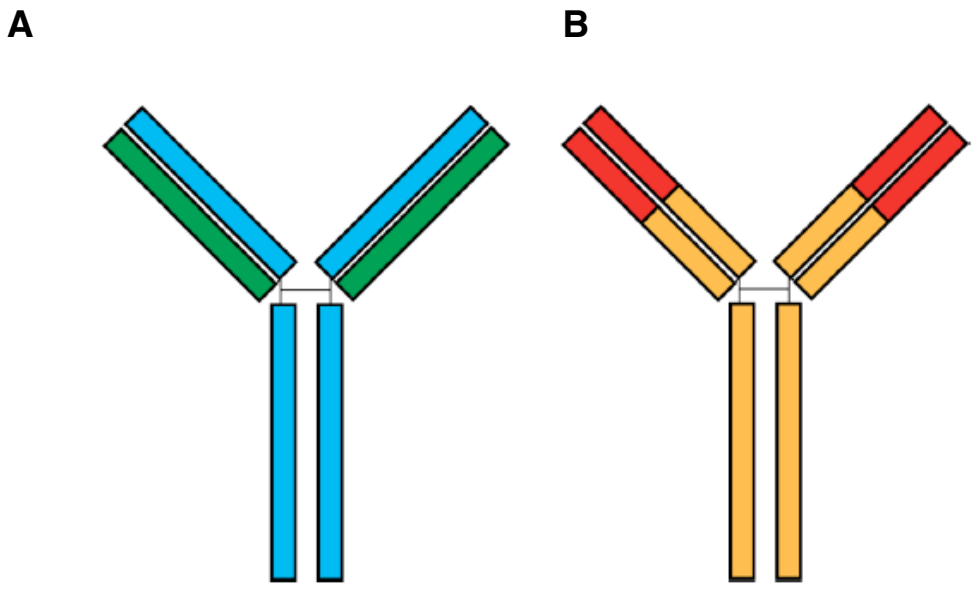
Of the three immunoglobulin loci, the heavy chain locus most frequently serves as a translocation partner. *IgH* is subject to all three of the antibody diversification reactions, and these reactions may generate the DSBs that serve as substrates for translocation (more on this in later sections). However, translocagenesis requires paired double strand breaks, and the origin of oncogene DSBs in mature B cell lymphomas has been much debated. The recurrent nature of translocations associated with mature B cell lymphomas might simply be a result of random events that are strongly selected. This must certainly be part of the picture whereas tumorigenesis requires a sustained selective advantage. However, oncogenes may also suffer specific recurrent targeted DSBs, as does the *IgH* locus.

Programmed DSBs in B Lymphocytes

B cells produce an antibody of two identical heavy and light chains. The heavy chains are connected to one another and to the light chains by disulfide bonds. The C-terminal portions of the heavy chains are called the constant region and are responsible for antibody effector functions. The N-terminal portions of the heavy and light chains compose the antigen-binding site or variable region (Figure 1.2). Antibodies recognize antigen in an effort to neutralize and mount a response to pathogens. However, a finite genome cannot code for a sufficient diversity of antigen receptors. B cells create this diversity *de novo* by somatic gene alteration in three reactions: V(D)J Recombination, Class Switch Recombination (CSR) and Somatic Hypermutation (SHM). Two of these reactions, CSR and V(D)J Recombination, involve DSB intermediates.

V(D)J Recombination. V(D)J recombination can be thought of as a "Pick 3 Lottery" where gene segments are combined in assorted permutations to generate the variable region of antibodies. It is a site-specific recombination reaction that, in B cells, occurs between immunoglobulin gene segments flanked by recombination signal sequences (RSSs). RSSs consist of a conserved heptamer and nonamer separated by either 12 or 23 bp of nonconserved sequence. Recombination reactions may only occur between 12- and 23-RSSs (van Gent et al., 1996). V(D)J is initiated by DSBs located between the gene

segment and its RSS. RSSs are joined to form signal sequences while gene segments are joined in a process that involves the alteration of sequences at the junction, thereby introducing additionally sequence, and therefore antibody, diversity.



Adapted from (Bothmer, 2011)

Figure 1.2. Schematic of an antibody. (A) Antibodies are composed of two identical heavy (blue) and light (green) chains connected by disulfide bonds. (B) The C-terminal portions of the heavy chains are called the constant region (orange), at the N-terminus of the heavy and light chains is the antigen-binding site (red).

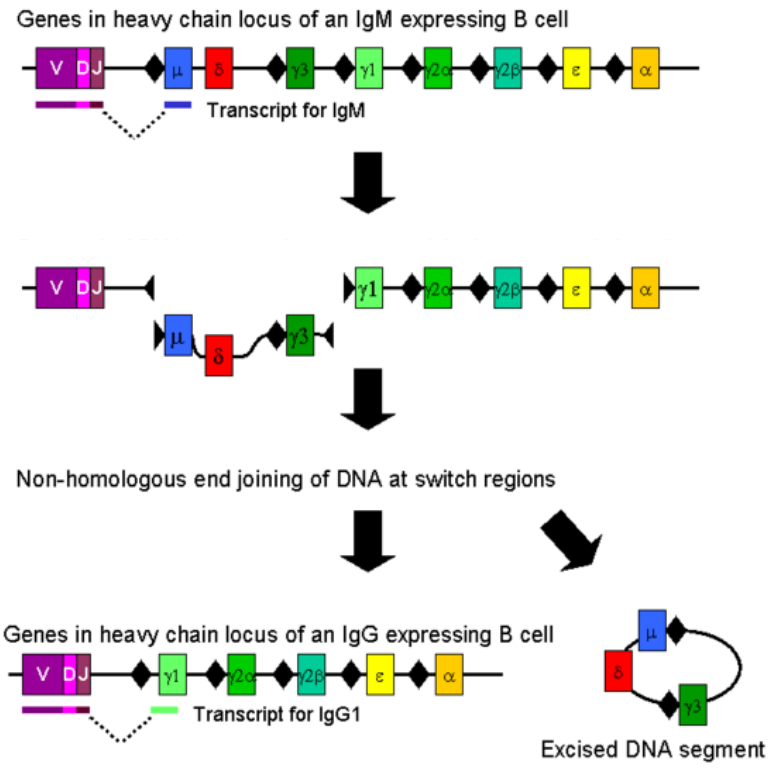
In B cells, V(D)J recombination occurs on the immunoglobulin heavy (*IgH*) and light chain loci. In mice, *IgH* consists of hundreds of V, 13 D and 4 J gene segments over a ~ 1 Mb region on chromosome 12. There are 2 light chain loci, *IgKappa* and *IgLambda*, which only contain V and J segments. B cell development begins in the bone marrow where pro-B cells undergo D-J rearrangement followed by V-DJ rearrangement on *IgH*. They are now pre-B cells, which undergo V-J rearrangement on the light chain and develop into immature B cells.

V(D)J is catalyzed by the RAG1 and RAG2 gene products. RAGs introduce a nick between two coding sequences and their RSSs, then catalyze a transesterification reaction in which the 3' hydroxyl group on the coding strand forms closed hairpin coding ends and blunt signal ends (McBlane et al., 1995). This process generates 4 ends that remain associated with RAGs in a post-cleavage complex (Qiu et al., 2001; Tsai et al., 2002). Both signal ends and coding ends are joined by the NHEJ machinery, requiring the activity of Ku70, Ku80, XRCC4 and Ligase4 (Roth, 2003). However, before the coding ends are joined, the endonuclease Artemis opens the hairpins to generate a substrate for nucleotide addition by terminal deoxynucleotidyl transferase (TdT) (Komori et al., 1993; Ma et al., 2002). The absence of any NHEJ factors leads to the complete lack of V(D)J recombination and the severe-combined immunodeficiency phenotype (Frank et al., 1998; Gu et al., 1997; Nussenzweig et al., 1996; Taccioli

et al., 1993).

The RAG post cleavage complex is thought to shepherd the ends to joining via the NHEJ pathway, excluding access to the ends by alt-NHEJ. This function is mediated by RAG2, which plays an important role in protecting RAG-generated ends from generating oncogenic rearrangements (Corneo et al., 2007; Deriano et al., 2011).

Overview of Class Switch Recombination. After antigen receptor assembly, B cells home to the peripheral lymphoid organs where they may undergo Class Switch Recombination (CSR). CSR is a deletional recombination event that exchanges an *IgH* constant region gene for one of several others. The constant region of an antibody (or isotype) determines its effector function. For example, IgG1 can activate complement while IgE causes leukocyte degranulation. In mice there are 8 constant regions: *Cμ, Cδ, Cγ3, Cγ1, Cγ2b, Cγ2a, Cε and Cα*. Naive B cells transcribe the *Cμ* constant region and express cell-surface IgM (IgD is expressed by virtue of alternative splicing to *Cδ*). During CSR IgM is replaced by one of the six other constant regions which code for the alternate antibody isotypes: IgG3, IgG1, IgG2b, IgG2a, IgE and IgA. The mouse constant region genes lie on a continuous stretch of chromosome 12. Just upstream of each constant region lies a switch region, a highly repetitive 1-5kb DNA sequence (Figure 1.3) (Chaudhuri and Alt, 2004).



Adapted from freely licensed media

Figure 1.3. Schematic of class switch recombination. At top, the structure of the *IgH* locus, colored rectangles represent constant region genes and black diamonds represent repetitive switch regions. AID introduces DSBs in switch regions and the intervening region is excised and circularized. NHEJ repairs DSBs, juxtaposing an alternative constant region to the VDJ gene segments.

Upon antigen encounter, B cells enter the germinal center and express AID (Muramatsu et al., 1999). AID was discovered by a subtractive hybridization screen in a B lymphoma cell line, and is required for CSR in mice and humans (Muramatsu et al., 2000; Muramatsu et al., 1999). AID deaminates cytidine residues on DNA, converting them to uracil (Dickerson et al., 2003; Petersen-Mahrt et al., 2002; Pham et al., 2003), primarily at WRCY motifs (W=A or T, R=A or G, Y=C or T) (Martin et al., 2002). During somatic hypermutation, AID introduces U:G mismatches into the immunoglobulin variable regions, which can then be processed into an array of mutations to diversify the antigen-binding motif (Peled et al., 2008). During CSR (which will be the focus of this section) mismatches are processed into switch region DSBs, which are joined with deletion of the intervening sequence (Figure 1.3) (Stavnezer et al., 2008).

AID targeting to the switch region. The mechanism by which AID targets switch region cytidines for deamination has been long-studied and recently, the molecular underpinning has been revealed. CSR was initially associated with transcription by the discovery of non-coding transcripts of class switched constant region genes in a lymphoma cell line (Stavnezer-Nordgren and Sirlin, 1986). Later work showed that certain cytokines stimulate germline transcripts and CSR. Specifically, TGF-beta induces IgA transcripts and switching to IgA in lipopolysaccharide (LPS) stimulated B cell cultures (Lebman et al., 1990). Additionally, deletion of the promoters responsible for sterile transcripts abolishes

CSR (Jung et al., 1993). Finally, AID activity on a transcribed gene has been shown to increase proportionally with transcription (Yoshikawa et al., 2002). Switch regions are repeat-rich and significantly enriched in guanine, so the formation of secondary structures has been hypothesized to play a role in AID targeting. Transcribed G-rich DNA is predisposed to forming stable RNA:DNA hybrids in R loops. Indeed, the presence of stable RNA-DNA hybrids (R loops) has been shown by bisulphate sequencing and has been detected *in vivo* (Yu et al., 2003).

The R loop as a potential substrate for AID is consistent with AID's activity on ssDNA, which is exposed in the transcription bubble. Both *in vivo* and *in vitro* assays have demonstrated that ssDNA is a high efficiency substrate for AID, while dsDNA is not (Bransteitter et al., 2003; Chaudhuri et al., 2003; Dickerson et al., 2003; Petersen-Mahrt et al., 2002; Pham et al., 2003; Ramiro et al., 2003; Sohail et al., 2003). For example, incubation of dsDNA with RNA polymerase in an *in vitro* transcription system with AID resulted in cytidine deamination on the non-transcribed strand (Pham et al., 2003). In B cells, AID has been shown to interact with PolII, demonstrating a physical interaction with the transcription machinery (Nambu et al., 2003; Pavri et al., 2010). Interestingly, AID appears to have activity on both the transcribed and non-transcribed DNA strand in SHM and CSR, although only the non-template strand is single stranded. AID may access ssDNA on the non-template strand by interaction with the ssDNA binding

protein Replication protein A (RPA) (Chaudhuri et al., 2004). The exosome complex may play a role in facilitating non-template strand access as well (Basu et al., 2011).

The distribution of PolII on the *IgH* locus provides additional insight into the mechanism of AID targeting. PolII ChIP experiments showed a high density of AID-independent PolII throughout the repetitive switch region, indicative of polymerase stalling (Rajagopal et al., 2009). Corroborating this finding, run-on analysis (a direct measure of transcription *in vivo*) also detected stalled transcripts in the switch region (Rajagopal et al., 2009). This stalling profile matched the hypermutation distribution well; together suggesting that stalling might precede and orchestrate AID targeting. Interestingly, G-rich switch region repeats appear to directly encourage stalling; in their absence PolII accumulation was not observed. Polymerase stalling results in a persistent transcription bubble and thereby, the prolonged exposure of ssDNA (Pavri et al., 2010). Given AID's proclivity for deamination of ssDNA, these data provide an attractive framework for AID targeting. Namely, switch regions bear high levels of stalled PolII, which exposes ssDNA, a high efficiency substrate for AID.

In 2010, Pavri and collegues performed an shRNA screen for factors that abrogate CSR in the CH12 B lymphoma cell line. They identified Suppressor of Ty5 Homolog (Spt5), as required for CSR (Pavri et al., 2010). Spt5 is a

transcription factor that associates with stalled PolII (Rahl et al., 2010), or PolII that accumulates in the promoter-proximal region. They showed that Spt5 directly associates with AID, and is required for AID recruitment to the *IgH* locus. ChIP-Seq in B cells of Spt5 revealed marked colocalization of Spt5 and PolII, with the *IgH* locus being the greatest area of enrichment for both, genome-wide. Additionally, AID ChIP-Seq correlated well with both Spt5 and PolII ChIP-Seqs; thousands of loci were found to load AID, and these loci also bore stalled PolII and Spt5. The promoter-proximal regions of these genes (an established area of stalling) were the primary sites of enrichment. Taken together, these data suggest that AID deaminates cytosines on ssDNA exposed by stalled PolII, and is targeted by interacting with Spt5. Indeed, the authors showed that Spt5 and AID interact in both primary B cell and fibroblasts. To confirm, the authors sequenced 10 genes that bore high levels of Spt5 for mutations. Generally, these genes suffered mutation, but 10 genes represent a very limited survey. A direct genome-wide measure of AID activity would be the ideal comparison, but the technology for such an experiment was not available.

U:G mismatches yield switch region DSBs. The U:G mismatch introduced by AID may be processed into DSBs. Several experimental approaches have been used to demonstrate switch region DSBs during CSR. First, the DNA damage response protein NBS1 and the DNA damage induced histone mark γH2AX form AID-dependent foci at the *IgH* locus in stimulated B cells (Petersen et al., 2001).

Second, linker-mediated PCR (LM-PCR), a technique to capture and amplify genomic DSBs at a particular locus, indicates AID-dependent DSBs at the switch regions (Schrader et al., 2005). But, how does a U:G mismatch result in a DSB? It has been shown that proteins in the base excision repair pathway (BER) effect this conversion. The mismatch is recognized and converted to an abasic site by Uracil DNA Glycosylase (UNG1) (Di Noia and Neuberger, 2002; Imai et al., 2003; Rada et al., 2002; Schrader et al., 2005). APE endonucleases can recognize this abasic lesion and cleave the DNA backbone to generate a single strand nick (Guikema et al., 2007). Thereafter, the precise mechanisms are unclear but it is thought that staggered nicks are processed into DSBs. The mismatch repair pathway, while not strictly required for class switching, may play a role in the generation of DSBs (Bardwell et al., 2004; Martin et al., 2003; Martomo et al., 2004; Schrader et al., 1999). Exo1 may be recruited the mismatch site by the MSH2-MSH6 complex, which then performs nucleotide excision followed by strand resection (Bardwell et al., 2004; Martomo et al., 2004). It has been hypothesized that paired nicks on opposite strands might be simultaneously resected thereby generating a DSB.

DSB repair at the switch region. DSBs in the switch region triggers the normal DNA damage response. DSBs are recognized by the MRN complex, which is followed by activation and accumulation of ATM, a serine/threonine kinase (Falck et al., 2005; You et al., 2005). ATM phosphorylates multiple downstream

effectors including the histone variant H2AX, converting it to γH2AX (Rogakou et al., 1998). γH2AX, in concert with other histone modifications and early DNA damage response factors, provides a scaffold for the focus formation of several DNA damage response factors such as MDC1 and 53BP1 (Schultz et al., 2000; Xie et al., 2007). The presence of these factors at the *IgH* locus in B cells undergoing CSR has been demonstrated by immuno-FISH (Petersen et al., 2001). The absence of these proteins results in reductions in CSR ranging from 3 fold, as in the case of ATM (Reina-San-Martin et al., 2004), to complete absence, as in the case of 53BP1 (Ward et al., 2004).

Sequencing of switch region junctions in class switched B cells shows that the majority do not contain significant microhomology (Yan et al., 2007), suggesting that the NHEJ pathway carries out the deletional recombination reaction. In mice with preassembled heavy and light chains that bear conditional deletion of the NHEJ factors XRCC4 or Lig4, CSR still occurs at 50% of wild type levels (Yan et al., 2007). Switch junctions in this system showed microhomology utilization, indicating that in the absence of NHEJ alt-NHEJ can contribute to efficient CSR. Similar results were obtained in a mouse lacking Ku70 and Lig4 (Boboila et al., 2010b). Thus, joining of switch region DSBs usually occurs by NHEJ, but in the absence of NHEJ pathway components, alt-NHEJ may perform the repair.

In contrast, the absence of 53BP1 results in complete abrogation of class

switching (Manis et al., 2004; Ward et al., 2004). Instead of joining DSBs in disparate switch regions, B cells lacking 53BP1 rejoin DSBs, producing intraswitch deletions. It has been hypothesized that this may be the result of DNA hypomobility in the absence of 53BP1. 53BP1 deficient cells show decreased movement of uncapped telomeres and B cells from *53BP1*$^{-/-}$ mice fail to perform efficient V(D)J recombination to distal gene segments (Difilippantonio et al., 2008; Dimitrova et al., 2008). Alternatively, 53BP1 has been shown to protect ends from resection; in its absence, ssDNA is exposed and available for microhomology utilization in joining (Bothmer et al., 2011; Bothmer et al., 2010). Thus, it is hypothesized that in the absence of 53BP1 switch region DSBs are resected, exposing switch region microhomology that is rejoined locally by the alt-NHEJ pathway.

AID Mediates Mutations and Translocations in Non-Ig Genes

AID generates physiological U:G mismatches in the switch region and variable regions to generate downstream DSBs and mutations, respectively. In the current model, AID is targeted to these regions by a physical interaction with Spt5, a PolII stalling factor. However, AID has been shown to mutate, translocate, and introduce DSBs in non-*Ig* genes. In fact, the study of AID activity at non-immunoglobulin loci has been instrumental in deciphering AID's targeting mechanism. Several non-Ig targets of AID have been implicated in lymphomagenesis, thereby associating AID with genomic instability in these tumors.

AID introduces mutations in non-Ig genes. Initial studies indicated that AID activity was associated with switch region transcription. Moreover, Ig genes could still be mutated when replaced with another promoter (Betz et al., 1994; Tumas-Brundage and Manser, 1997), and non-Ig sequences were mutated when in the context of the Ig locus. Therefore, it stood to reason that other transcribed genes might suffer AID mediated mutation. A first study isolated memory and naïve B cells from the peripheral blood of healthy donors and sequenced a select panel of genes. They found that *BCL-6*, a gene expressed in germinal center B cells and reported to carry mutations in non-Hodgkin's lymphoma, was significantly mutated (Shen et al., 1998). Importantly, the mutation spectrum was similar to

that found in Ig genes. A subsequent study had similar findings, but also reported that *BCL-6* sequencing from B cell tumors showed that only those cancers with a germinal center or post-germinal center phenotype carried mutations (Pasqualucci et al., 1998). Additionally, *CD95* was shown to bear a similar mutation spectrum in the GC B cells of healthy donors (Muschen et al., 2000). Later work expanded the roster of expressed, mutated non-Ig genes to include *Pim1, c-myc, Pax5, CD79a and CD79b* (Gordon et al., 2003; Pasqualucci et al., 2001). Importantly, mutation at these loci displayed a similar profile to variable region mutations but at a much lower frequency; a preference for transitions over transversions and specific targeting to WRCY motifs. These papers suggested that AID might target non-Ig genes for mutation.

A broader study of AID dependent mutation in germinal center B cells suggested that AID might very widely mutate transcribed genes. Liu and colleagues isolated germinal center B cells from AID deficient and wild-type mouse peyer's patches, then sequenced 118 genes, in a region 1.5kb downstream from the transcription start site based on their expression in the germinal center. While mutation frequencies were about 100 times lower than immunoglobulin mutation rates, 25% of the genes examined bore significant AID dependent mutations (Liu et al., 2008). Consistent with previous studies, *BCL6* was the most highly mutated gene in the experimental set. Interestingly, when the same genes were sequenced in germinal center B cells from mice lacking UNG and MSH2, several genes

including the *c-myc* proto-oncogene, appeared to have increased mutations rates. This suggested that some genes targeted by AID might undergo error-free repair by the base excision repair and mismatch repair pathways.

This study was the most comprehensive analysis of non-*Ig* AID targeting at the time, and surely demonstrated that AID triggers mutation at several locations throughout the genome. However, low-throughput sequencing is a relatively cumbersome technique, prohibiting its application in a truly comprehensive study. Indeed, these experiments delivered a limited picture of AID's role in genomic instability. First, only genes were examined for mutation; this excluded the possibility of discovering AID's role in targeting non-genic regions, should it have one. Second, from thousands of potentially mutated genes only a small percent were examined (118 of ~25,000 or 0.5%), while these result were generalized to conclude that 25% of all genes are mutated, it was not a true measure of AID promiscuity. Third, the selection criteria for genes required that they be in the top decile of microarray expression values; as such, the effects of transcription on non-Ig AID-dependent mutations could not be effectively assessed. Fourth, they only sequenced a 1 kb region of each gene in the same location – 1.5 kb downstream from the transcription start site. Whether this is the primary site of AID targeting in the gene body was unexamined. Finally, while mutations may be a proxy for DSBs, especially in the absence of UNG, DSBs and subsequent translocation were not directly measured. In fact, no system to perform an

unbiased, large-scale analysis of AID dependent translocations has yet been be devised.

Expanding on this work, a recent genome-wide study revealed the full spectrum of potential AID targets. Yamane and colleagues performed AID-ChIP followed by deep sequencing and showed that AID is associated with almost 6,000 genes in activated B cells. This confirmed the assertion from Liu and colleagues that AID might associate with 25% of all genes. As expected, the *IgH* locus loaded the greatest quantity of AID, while genes previously shown to be mutated by AID or translocated in mature B cell lymphomas appeared throughout the list. For example, *mir-142* (a translocation partner of *c-myc* and mutation target of AID (Gauwerky et al., 1989; Robbiani et al., 2009)) was among the 15 genes with the greatest AID levels. Importantly, the presence of AID as determined by ChIP-Seq was predictive of AID-mediated mutation, suggesting that AID both binds and mutates promiscuously (Yamane et al., 2011).

When compared to histone modification and PolII ChIP-Seq landscapes, several trends emerged. AID was shown to primarily associate with transcribed genes as determined by mRNA-Seq, and to associate with open chromatin as determined by the histone modification H3K4 trimethylation. Within the gene body, AID bound predominantly at the promoter proximal sequence, and colocalized with putative sites of RNA PolII accumulation and stalling. This meshed well with

previous data suggesting that AID targeting depends on PolII stalling and ssDNA availability (Pavri et al., 2010; Rajagopal et al., 2009). However, this study was marked by some limitations. First, AID association is a substitute evaluation for AID activity, so while AID ChIP-Seq may define sites of AID association it is not a direct measure of activity. Second, after surveying 13 AID associated genes for mutations, it was concluded that AID ChIP-Seq is a good proxy for AID activity, but this is a relatively small-scale analsis. Third, the authors conclude that AID association correlates well with AID-mediated mutation, but the question of AID mediated translocations and DSBs was left unaddressed.

AID-mediated translocations. AID is also responsible for initiating translocations, events commonly observed in mature B cell lymphomas. For instance, formation of the *c-myc/IgH* translocation (a hallmark of Burkitt's lymphoma) in the IL-6 induced plasmacytoma lymphoma model, requires the presence of AID (Ramiro et al., 2004). Moreover, the same translocation occurs in primary B cell in culture within hours of AID expression, and only in the presence of UNG (Ramiro et al., 2006). The requirement for UNG indicates that the conversion of U:G mismatches to DSBs is required for translocation, and strongly implicates AID in directly initiating the DSB.

Translocations require paired DSBs (Richardson and Jasin, 2000), and while AID had been conclusively shown to initiate DSBs in *IgH*, the etiology of the DSB in *c-*

myc was unclear. Fragile DNA elements prone to form non-B-DNA configurations, including Z-DNA, triple-helical H-DNA, and G-quadruplex DNA have been proposed to account for breaks in *c-myc* (Liu and Levens, 2006; Marcu et al., 1992; Wierstra and Alves, 2008). Recently, direct evidence for AID generating the DSBs in *c-myc* came from genetic experiments using the site-specific yeast meganuclease I-SceI. In mice bearing I-SceI sites in both *IgH* and *c-myc* I-SceI is sufficient to produce *c-myc/IgH* translocations, and when I-SceI is only present at *IgH*, I-SceI and AID together produce *c-myc/IgH* translocations. However, in the absence of AID, I-SceI-mediated DSBs in *IgH* did not generate *c-myc/IgH* translocations (Robbiani et al., 2008). Thus, AID is responsible for the breaks in *c-myc* that partner with *IgH* to form the *c-myc/IgH* translocation.

AID's ability to introduce DSBs at *IgH* and non-*Ig* loci such as *c-myc* may play a role in mature B cell lymphomagenesis *in vivo*. In humans, AID is expressed in mature B cells lymphomas that carry translocations to *IgH*. For example, AID is expressed in activated B cell-like diffuse large B cell lymphoma (ABC-DLBCL) (Lenz et al., 2007) and, ABC-DLBCL has a high frequency of illegitimate switch region recombinations and deletions, as well as translocations involving *IgH* (Lenz et al., 2007) to genes targeted for mutation by AID, such as *c-myc* and *BCL6*. These findings associate AID with lymphomagenesis, but a mouse model finally proved that AID causes mature B cell lymphomas,

Several transgenic mice constitutively over-expressing AID have been generated and many of these mice develop cancer. However, the tumors are not of B cell origin, instead they are epithelial tumors and T cell lymphomas (Muto et al., 2006; Okazaki et al., 2003; Shen et al., 2008). These tumors carried mutations in oncogenes, but did not bear any translocations (Okazaki et al., 2003). Additionally, in conflict with the IL-6 induced plasmacytoma mouse (which bore tumors only in the presence of AID), mice with a B cell specific AID transgene did not succumb to any tumors (Muto et al., 2006). In 2009, Robbiani and colleagues published the first direct evidence that AID contributes to lymphomagenesis *in vivo*. *p53*$^{-/-}$ mice overexpressing AID from a transgene in B cells by virtue of V kappa promoter succumbed quickly to mature B cell lymphomas (Robbiani et al., 2009). This was in contrast to the control group of *p53*$^{-/-}$ mice, which perished from T cell lymphomas and sarcomas. Important, B cell lymphomas in these mice carried clonal reciprocal translocations, including the *c-myc/IgH* translocation observed in Burkitt's lymphoma. This mouse model also yielded tumors bearing a clonal translocation between *c-myc* and *mir-142*. The *c-myc/mir-142* translocation juxtaposes the *c-myc* coding region to the *mir-142* regulatory elements, thereby dysregulating *c-myc* expression. Sequencing of the *mir-142* locus from germinal center B cells in WT and AID$^{-/-}$ mice revealed that *mir-142* is a bona fide target of AID. This discovery provided further evidence for AID's role in lymphomagenesis in humans; the *c-myc/mir142* translocation has been reported in a human aggressive B cell leukemia (Gauwerky et al., 1989).

These studies nicely show that AID is capable of initiating DSBs and translocations. However, to date *c-myc* and *mir142* are the only genes conclusively shown to suffer AID dependent translocation. B cells overexpressing AID harbor high levels of translocations and chromosome breaks (Robbiani et al., 2009). But, the entire array of genes susceptible to DSBs by AID is unknown because a method for high-resolution unbiased mapping of DSBs or their resulting translocations does not exist.

CHAPTER 2:

Translocation Capture Sequencing

Recent deep-sequencing studies have revealed hundreds of genomic rearrangements in human cancers (Campbell et al., 2008; Pleasance et al., 2010a; Pleasance et al., 2010b; Stephens et al., 2009). These studies have shed some light on the nature and extent of genomic abnormalities. However, rearrangements from tumors may be highly selected, and cancer cells might have altered DNA repair pathways that give rise to a different rearrangement profile. Thus, the physiologic constraints that govern the genesis of these events remain unclear because methods for mapping chromosomal translocations in primary cells do not yet exist. Here, we sought to develop a genome-wide strategy to document chromosomal rearrangements in short-term cultures of primary cells.

In our system, DSBs are induced at the *c-myc* (chromosome 15) or *IgH* (chromosome 12) loci, which were engineered to harbor the I-SceI meganuclease target site (Robbiani et al., 2008) (*c-myc*$^{I\text{-}SceI/I\text{-}SceI}$ or *IgH*$^{I\text{-}SceI/I\text{-}SceI}$ B cells respectively, hereafter referred to as Myc^{I} and IgH^{I}) (Figure 2.1). I-SceI, an intron encoded endonuclease present in the mitochondria of *Saccharomyces cerevisiae,* recognizes an 18 base pair sequence absent from the mouse genome. The I-SceI break system was first developed in mammalian cells to

examine DSB resolution (Rouet et al., 1994). It has since been widely used to study many aspects of DSB repair including DSB end processing, chromosomal translocation, pathway selection, and to perform gene targeting (Bothmer et al., 2010; Brunet et al., 2009; Windbichler et al., 2007). Here, the unique break at I-SceI sites serves as bait for other endogenous breaks in the cell, thereby generating rearrangements with one known partner.

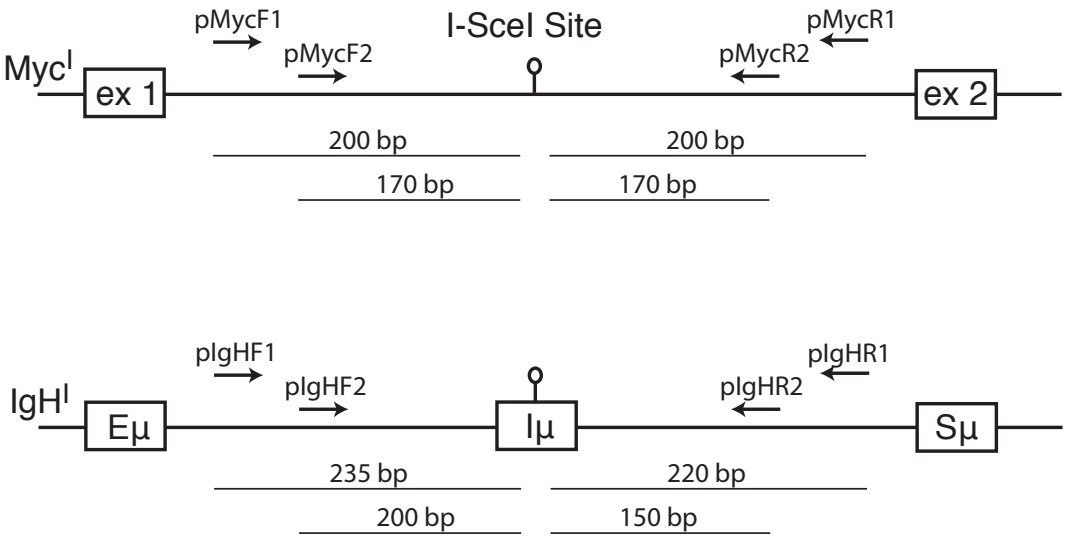


Figure 2.1. Schematic of loci containing I-SceI sites. The *IgH* locus of IgH^{I} B cells is shown below, while the *c-myc* locus from Myc^{I} B cells is shown above. Each is displayed with the site-specific primers used for TC-seq and their approximate distance from the I-SceI site. Nested primers were placed at least 150 bp from the I-SceI site to allow for capture of events generated from moderately resected ends.

To generate genomic DNA bearing rearrangements to I-SceI sites, mouse splenic B cells were isolated by immunomagnetic depletion with CD43 beads and stimulated to undergo isotype switching in cultures with LPS and IL-4. The CD43 antigen is expressed on most T cells, NK cells, granulocytes, monocytes, and macrophages, on hematopoietic stem cells, and platelets. Among B cells, CD43 is expressed on activated B cells and plasma cells but not on resting B cells (naive B cells) (from Miltenyi Biotech). On culture days one and two activated B cells were infected with retroviruses to express I-SceI (with mCherry by virtue of an embedded IRES element), and in some cases AID (with GFP by virtue of an embedded IRES element). On day 4 singly infected cells were harvested for genomic DNA while doubly infected cultures were sorted for double positive cells then harvested (Figure 2.2).

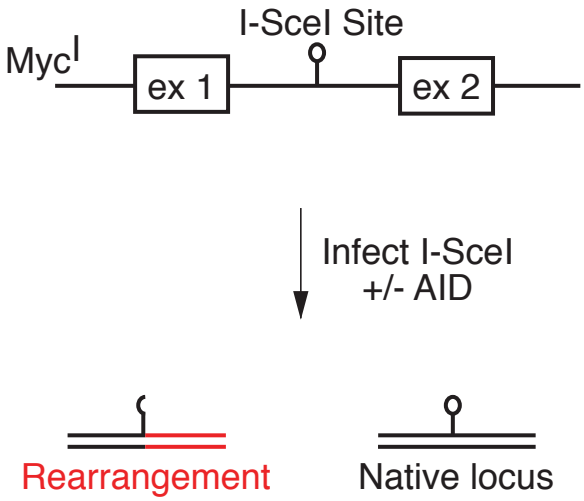


Figure 2.2. Schematic of concept for rearrangement generation. B cells bearing I-SceI sites are isolated and infected with retroviruses to express I-SceI endonuclease with or without AID. I-SceI cleaves the target sequence, which may then join to endogenous DSBs to generate chromosomal rearrangements with one known partner. Alternatively, the I-SceI induced break might religate and assume the conformation of the native locus.

We sought a method to isolate and sequence rearranged I-SceI sites. Of course, classical PCR could not be utilized; rearrangements to the I-SceI site only contain one known partner sequence. Essentially, this task requires a mechanism for amplifying a genomic region using only one site-specific primer. A solution was developed in 1989 to perform *in vivo* footprinting of enhancers and methylation analysis (Mueller and Wold, 1989; Pfeifer et al., 1989). By ligating DNA to double stranded linkers then using primers that anneal to the sequence of interest and the linker, one can perform single-sided PCR (dubbed ligation-mediated PCR or LM-PCR). Since then, this technique has been used in various applications including the investigation of genetically programmed DSBs in lymphocytes and HIV integration site mapping (Papavasiliou and Schatz, 2000; Schroder et al., 2002; Zhu and Roth, 1995). We designed linkers to allow for high-fidelity PCR with low background by including the following features.

1) Strand asymmetry; to allow for novel sequence generation in the case of site-specific primer extension (Figure 2.3).
2) Thymine tail; to prevent linker-linker dimer formation, and to accommodate genomic DNA fragment adenine tailing so that intermolecular ligation events do not occur (Figure 2.3).
3) Dideoxycytosine block; to prevent double stranded linker generation in the absence of site-specific primer extension (Figure 2.3)

4) AscI cleavage site; to remove linkers prior to deep-sequencing. Additionally, AscI cleavage will leave an 8 nucleotide barcode on fragments, indicating that it was a product of linker-mediated amplification (Figure 2.3).

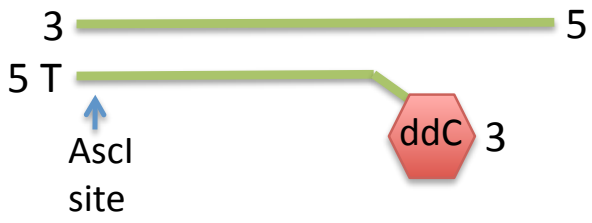


Figure 2.3. Structure of LM-PCR linker. The double stranded asymmetric linker (dsaLinker) for LM-PCR contains a T-tail, 5' overhang, AscI site and a dideoxycytosine blocked 3' end.

First, genomic DNA was sonicated to a mean size of 800 bp, blunted, A-tailed and ligated to dsaLinkers. Then, purified I-SceI enzyme was applied to eliminate unrearranged target loci. (Figure 2.4A). Each sample was then divided into 2 equal parts, for TC-Seq with either forward or reverse primers. Forward and reverse preparations were processed separately for the entire protocol. DNA was subjected to single-primer PCR with biotinylated site-specific primers. This step generates a complete double stranded linker on genomic DNA fragments containing the I-SceI adjacent sequence (Figure 2.4B).

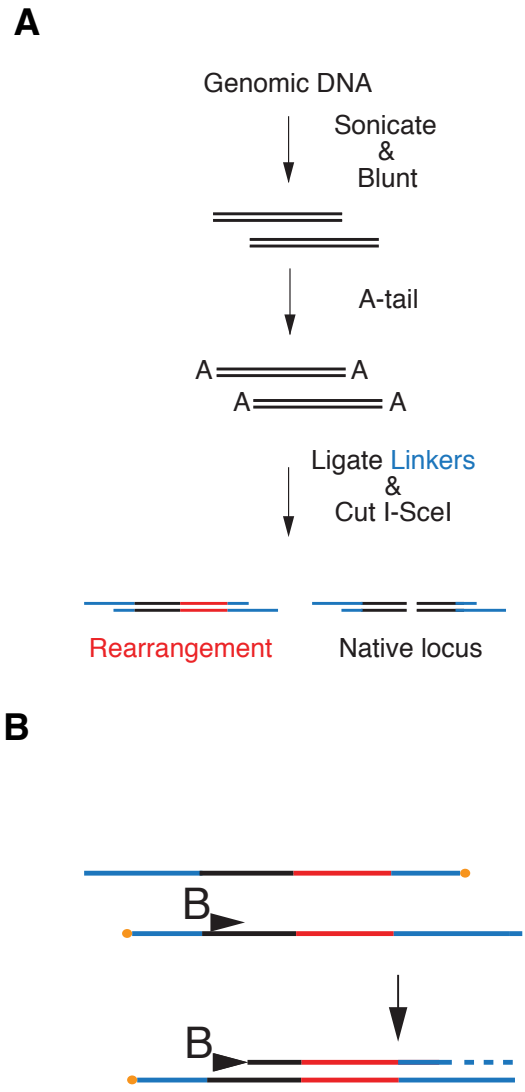


Figure 2.4. Translocation sequencing schematic I. (A) To prepare genomic DNA for TC-Seq it is sonicated, blunted, A-tailed, ligated to dsaLinkers and cut with I-SceI. (B) Single-primer PCR. Prepared genomic DNA fragments were subjected to 15 cycles of PCR with a single biotinylated primer annealing either upstream or downstream of the I-SceI site (pIgHF1, pIgHR1, pMycF1 or pMycR1) (Figure 2.1).

Next, PCR reactions were spiked with primer that anneals to the newly generated linker sequence. Higher molecular weight products were isolated from the reaction by agarose gel electrophoresis followed by streptavidin bead purification. A semi-nested PCR was performed (with primers pLinker and pIgHF2, pIgHR2, pMycF2 or pMycR2) (Figure 2.1), and higher molecular weight products were isolated by agarose gel electrophoresis. Finally, linkers were removed by AscI digestion and fragments were assembled into an Illumina library for Paired End deep-sequencing. 36x36 or 54x54 nucleotide sequencing was performed on an Illumina GAII (Figure 2.5) (See methods for an in-depth TC-Seq protocol including primer sequences).

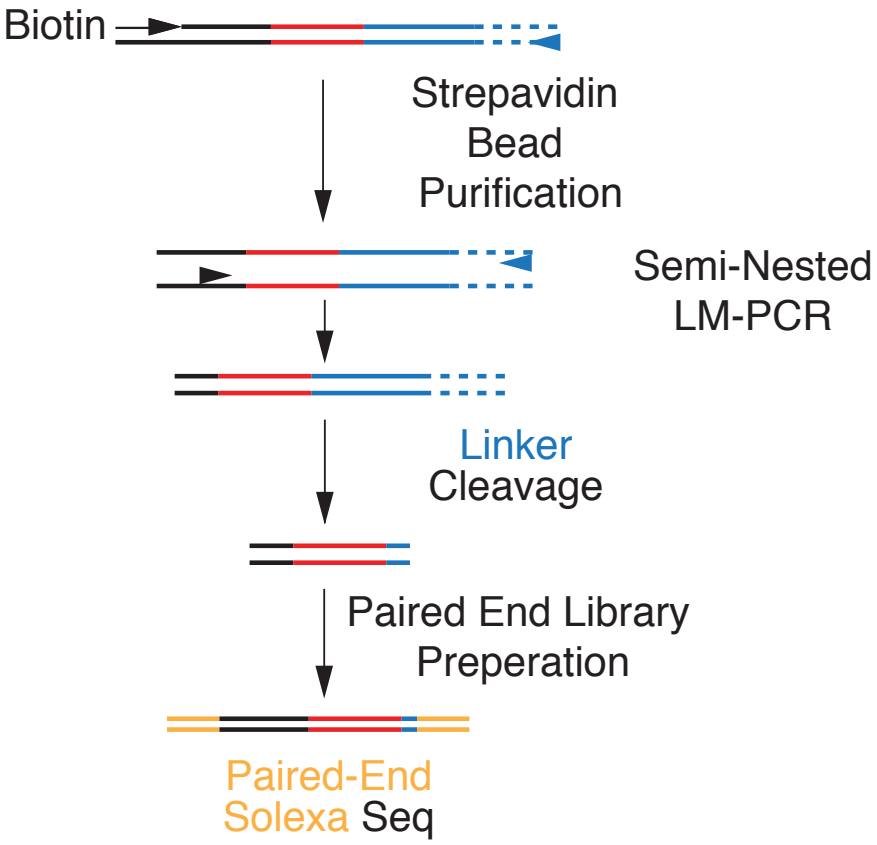


Figure 2.5. Translocation sequencing schematic II. Following LM-PCR, products were isolated by strepavidin coated magnetic bead purification and subjected to semi-nested LM-PCR. Linkers were cleaved off fragments by digestion with the restriction endonuclease AscI then were processed into paired-end libraries for deep sequencing.

TC-Seq libraries were generated from 50 million B cells each. Experiments were performed in duplicate (biological replicates) for a total of 100 million B cells assayed per condition. To determine AID's contribution to the genomic rearrangement landscape TC-Seq libraries to *IgH* and *c-myc* were assembled from AID deficient, AID sufficient and AID over-expressing (by retroviral infection) B cells (Table 2.1). Illumina sequencing runs were repeated on some samples (technical replicates) to increase depth and ensure saturation.

Next, sequencing data were mined for useful reads. Each end of the paired-end sequences was matched against the relevant bait primer (IgHF2, IgHR2, MycF2 or MycR2) plus genomic sequences allowing up to two mismatches with the bowtie short read aligner. For read pairs longer than 2 x 36 nts, 10 nts were trimmed off the 3' end of each read. Each read pair with a single match to one of the site-specific primers was then checked for a perfect match to the linker barcode on the second arm. After AscI cleavage, eight nucleotides remain ligated to the genomic DNA, this was termed the "barcode" and it indicates linker-specific amplification. If the barcode was present, this arm was designated a target arm, linker sequence was trimmed, and the remainder was aligned against the mouse genome (NCBI 37/mm9) with bowtie allowing up to 2 mismatches and requiring unique alignments in the best alignment stratum (command line options: -v2 --all --best --strata -m1). Exactly identical alignments (same position, same strand) were combined into a single putative rearrangement event, events supported by

a single alignment were not considered in any analyses. Since TC-Seq involves multiple rounds of PCR, it is unlikely that individual reads represent actual biological events. We also removed reads closer than 1 kb to their respective bait; these alignments represent rejoining of the I-SceI break, uncut, or unrearranged loci. Rearrangement positions are given as the position of the 5' end of the read in the alignment. It should be noted that these reads are displaced from the actual rearrangement breakpoints. Based on the distance of the primers from the I-SceI site and agarose gel size selection, the average displacement is expected to be ~250 bp. For all analyses, data from technical and biological repeats were pooled to increase saturation (Table 2.1).

Table 2.1. Data quantity per experiment. Bait – the locus containing I-SceI for each experiment. AID – the status of activation induced cytidine deaminase (wt=wild type, ko=knockout, rv=retroviral over-expression). Replicate – indicates which biological replicate the data set represents, identical letters indicate technical replicates of the same experiment, biological replicates are indicated by different letters. Sequenced pairs – total number of paired reads obtained. Valid pairs – total number of paired reads containing a matched linker and site-specific primer. Left – number of valid aligned reads paired with a forward primer. Right – number of valid aligned reads paired with a reverse primer. Total – total number of valid, aligned reads paired with a site-specific primer, for tabulated rows (gray shading) this value is the sum of the replicate experiments. Rearrangements [>1 kb from bait; >= 2 alignments] – total number of valid, aligned, unique reads represented by more than 1 read instance and lying greater than 1kb from the associated site-specific primer, for tabulated rows (gray shading) this value is the sum of the replicate experiments.

Table 2.1. Data quantity per experiment.

bait	AID	Repli-cate	sequenced pairs	valid pairs	left	right	total	Rearrange-ments [>1kb from bait; >= 2 alignments]
cMyc	ko	A	37,496,617	16,759,723	4,071,732	6,600,751	10,672,483	16,172
cMyc	ko	B	31,561,878	16,112,165	5,763,288	5,113,161	10,876,449	14,816
		A+B						28,548
cMyc	rv	A	42,942,899	13,132,287	2,610,943	458,513	3,069,456	25,653
cMyc	rv	A	39,981,085	12,332,625	2,979,621	513,095	3,492,716	26,465
		A					6,562,172	39,789
cMyc	rv	B	39,119,951	15,277,084	4,997,453	1,563,572	6,561,025	27,001
		A+B					13,123,197	63,772
IgH	rv	A	35,443,959	11,779,691	1,448,672	2,315,695	3,764,367	35,167
IgH	rv	A	10,860,307	2,797,362	212,458	970,485	1,182,943	21,677
IgH	rv	A	16,897,163	4,582,262	386,947	1,531,206	1,918,153	26,455
		A					6,865,463	43,315
IgH	rv	B	5,586,756	1,820,820	247,532	226,915	474,447	19,720
		A+B					7,339,910	51,312
IgH	ko	A	42,015,461	9,945,416	890,202	4,233,574	5,123,776	67,156
IgH	ko	B	12,289,033	2,723,023	701,452	334,221	1,035,673	1,247
		A+B						
IgH	wt	A	40,136,125	20,570,387	5,169,695	2,536,311	7,706,006	31,601
IgH	wt	B	25,885,221	8,005,979	1,507,154	1,665,017	3,172,171	14,104
		A+B					10,878,177	37,945

CHAPTER 3:

AID-Independent Rearrangements In Activated B Cells

In the absence of AID, DSBs arise as by-products of normal cellular metabolism including transcription, oxidative stress and DNA replication (Branzei and Foiani, 2010). To examine the distribution of AID-independent DSBs that lead to rearrangements we examined rearrangements to I-SceI sites in *c-myc*, in the absence of AID. Consistent with a global distribution of DSBs, we mapped 28,548 unique rearrangements between the I-SceI site and every chromosome in $Myc^{I}AID^{-/-}$ B cells (100 million cells assayed, Figure 3.1).

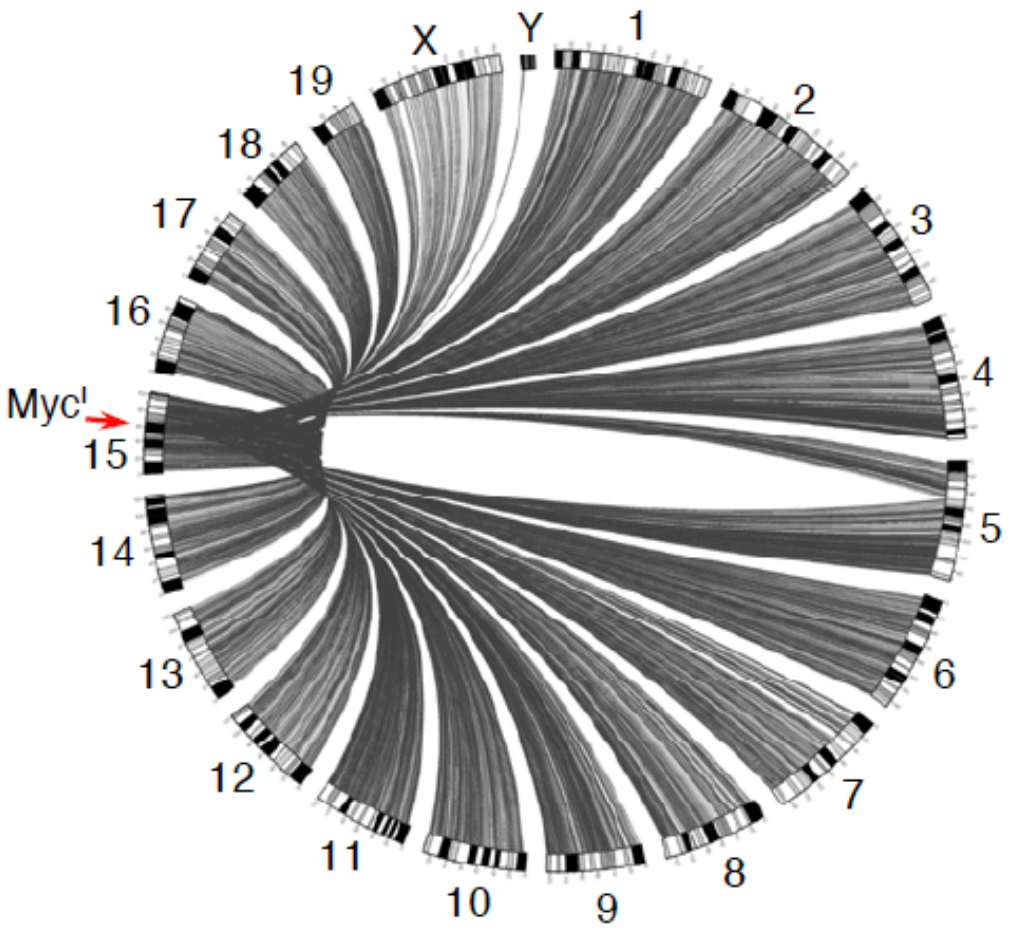


Figure 3.1. Genomewide view of rearrangements to Myc^{I} in AID$^{-/-}$ B cells. Rearrangements were mapped according to genomic position and displayed by the Circos circular visualization software package (http://circos.ca). Each line originating from *c-myc* on chromosome 15 represents a single rearrangement event. The red arrow indicates the I-SceI site embedded in *c-myc.*

To determine whether there is a genome-wide bias for rearrangement, these events were characterized based on location, transcriptional status and histone modification of the locus. First, we examined chromosomal location and found a marked enrichment of intra-chromosomal rearrangements on chromosome 15, with approximately 125 events per mappable megabase (11,066 rearrangements), or ~40% of all events genomewide (Figure 3.2). In contrast, translocations between Myc^{I} and other chromosomes were evenly distributed throughout the genome (Figure 3.2).

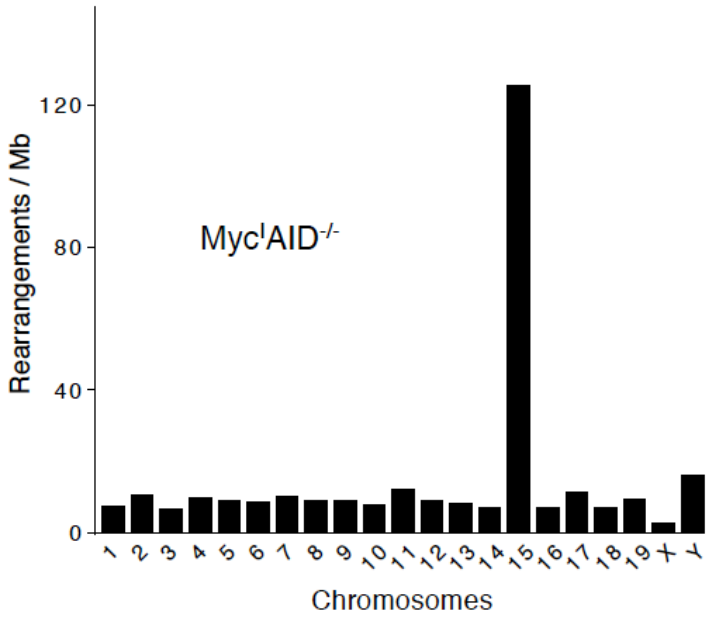


Figure 3.2. Rearrangements to each chromosome in $Myc^{I}AID^{-/-}$ B cells. Rearrangement partners to I-SceI sites in *c-myc* were binned based on chromosome, each bin was divided by the number of mappable megabases of sequence for that chromosome in the mm9 genome build.

Notably, 86.7% (9,591 of 11,066) of all intra-chromosomal rearrangements were localized within a 350 kb domain surrounding the I-SceI site (from -50 kb to +300 kb; Figure 3.3). This is consistent with the observation that 92% of intra-chromosomal rearrangements in the breast cancer genome involve aberrant joining of DSBs within 2 Mb of each other (Stephens et al., 2009), and that 87% of RAG-mediated intra-chromosomal rearrangements in Abl-transformed pre-B cells lie within 200 kb of a recombination substrate (Mahowald et al., 2009). The asymmetrical distribution of events in the direction of *c-myc* transcription and the adjacent *Pvt1* gene is also consistent with the idea that gene expression facilitates rearrangement (Thomas and Rothstein, 1989). I-SceI-proximal events may be the result of either resection and rejoining of I-SceI breaks, bona fide rearrangements between I-SceI and random DSBs, or a combination of DNA end resection and balanced translocations. Regardless of the precise molecular mechanism, the abundance of these events reveals a strong preference for DSBs to be resolved by ligation to a proximal sequence, a DNA repair strategy that may minimize gross genomic alterations.

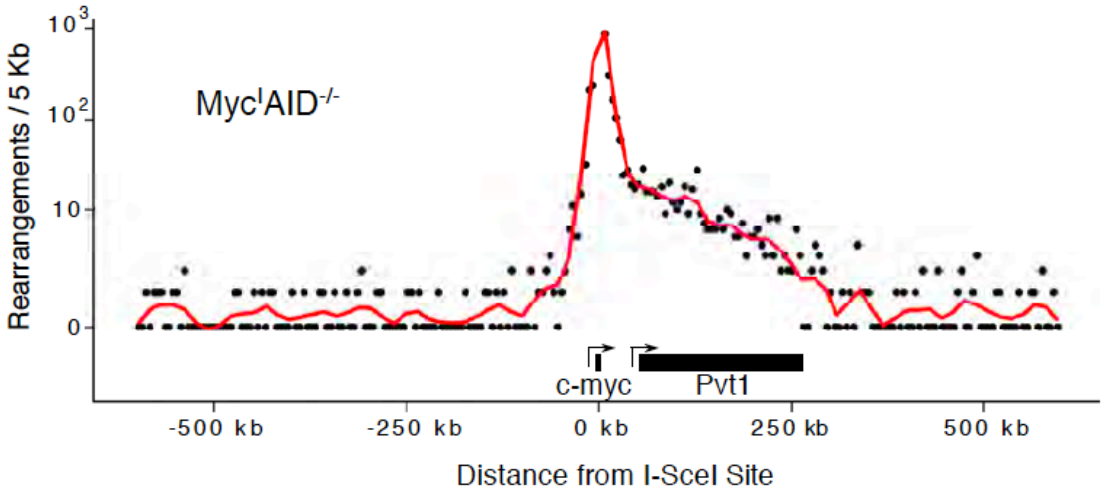


Figure 3.3. Profile of rearrangements around the *c-myc* I-SceI site. Local rearrangements in Myc^{I} were mapped in 5 kb bins around the I-SceI site. Black dots represent values at individual bins, the red line represents a smoothed moving average of bin contents. Data were plotted on a logarithmic scale. Below, schematic of the *c-myc* locus containing the position of the adjacent gene, *pvt1.* Arrows indicate the direction of transcription.

Recent cancer genome sequencing experiments uncovered a modest but highly significant preference for cancer-associated rearrangements to occur in genes, which compose only 41% of the human genome. For example, in 24 sequenced breast cancer genomes, 50% of all rearrangements involved genes (Stephens et al., 2009). Whether this bias resulted from selection or some inherent feature of DSB formation and repair specific to cancer cells could not be determined. To ascertain whether a similar bias is seen in primary cells in short term cultures, AID-independent rearrangements in $Myc^{I}AID^{-/-}$ B lymphocytes (excluding 1 Mb of DNA around the I-SceI site) were classified as genic or intergenic. Local events were omitted from the analysis because of the observed bias for events around the break. Consistent with the human tumor studies, 51% (9,677 of 19,246) of the events were associated with genes (Figure 3.4A). Because only 40% of the mouse genome is genic, this represents a small (1.25-fold) but significant difference (permutation test $P < 0.001$) relative to intergenic regions.

Next, we performed a composite density analysis of rearrangements to genes and intergenic regions. In this approach events are mapped along the gene body or regions between genes but, instead of mapping by nucleotide position they are mapped according to distance from the transcription start site as a percent of the entire locus body. For example, a translocation that maps to base 250 of a 1000 bp gene would be given the value of 25%. Finally, all genes and intergenic regions are combined to generate a merged graph of reads along all gene bodies

and intergenic spans. Using this approach we find that genic rearrangements were particularly enriched at transcription start sites (TSS) (Figure 3.4B). Importantly, there appears to be no enrichment at any length along the composite intergenic region, indication that the TSS is indeed unique in its enrichment.

A

		Total Captured Rearrangements		
		Genic	Intergenic	% genic
Myc^{I}	$AID^{-/-}$	9677	9569	50.3
	AID^{RV}	26221	25101	51.1
IgH^{I}	AID^{WT}	5651	5723	49.7
	AID^{RV}	9475	8026	53.9

P (permutation test) < 0.001 for all

B

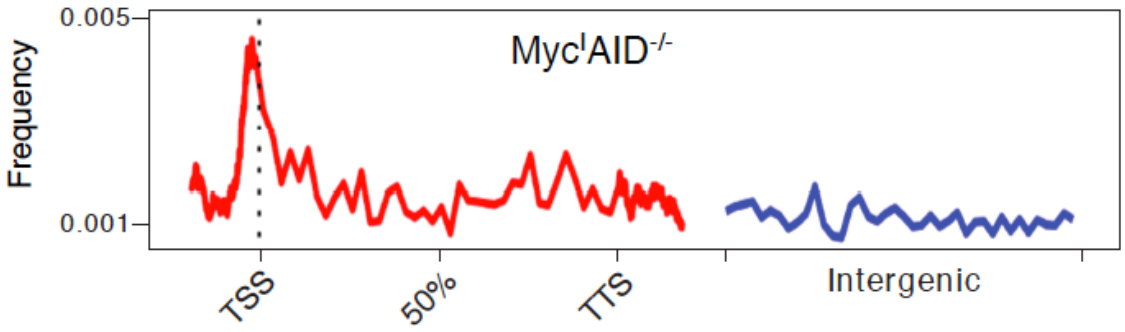


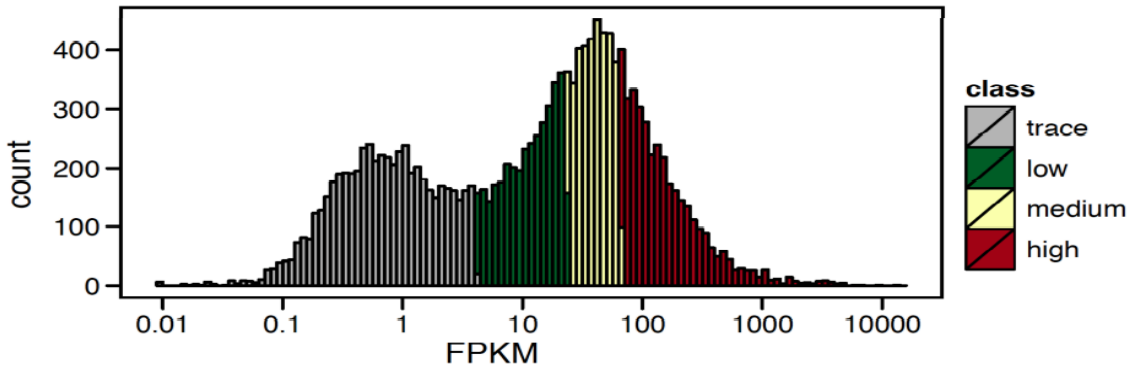
Figure 3.4. Rearrangements to genic and intergenic regions.

(A) Rearrangements, excluding the 1 Mb around I-SceI, were categorized as genic or intergenic for each sample. (B) Composite density profile of genomic rearrangements from Myc^{I} to genes (red line) and intergenic (blue line) regions. TSS = transcription start site, TTS = transcription termination site.

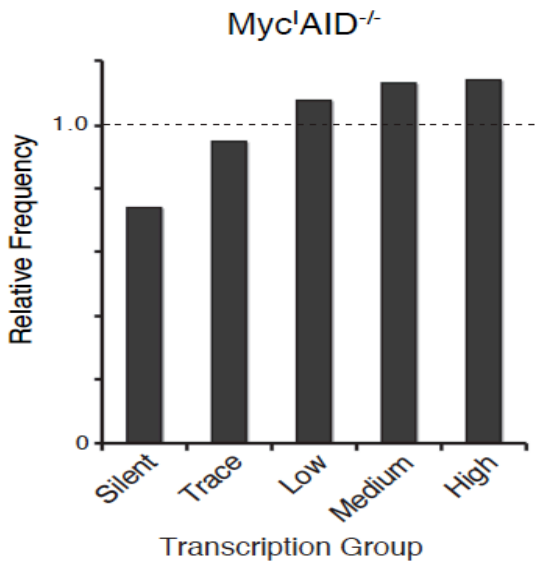
Next, we asked whether transcriptional activity is a determinant of rearrangement. mRNA-Seq data from LPS and IL-4 activated B cells was divided into quintiles of expression and the number of rearrangements to each quintile was tabulated (Yamane et al., 2011). We then performed 1000 computational Monte Carlo experiments to determine the expected value in each transcription quintile assuming a random distribution along the genome, using their average as an expected frequency. Finally, we calculated relative frequency in each quintile (fe) – this number indicates the percent divergence of the observed value from the expected. Consistent with the preference for genic rearrangements, there was a bias to transcribed genes. Fewer rearrangements than expected occurred at silent (fe = 0.74, $P < 0.001$) and trace (fe = 0.95, $P < 0.001$) transcribed genes, while more than expected occurred at low (fe = 1.08 $P < 0.001$), medium (fe = 1.13, $P < 0.001$), and highly (fe = 1.14, $P < 0.001$) transcribed genes (Figures 3.5A and 3.5B).

Figure 3.5. Rearrangements in transcription level gene groups. (A) Genes were divided into silent (0 FPKM (fragments per kilobase of exon per million fragments mapped)) and expressed genes based on mRNA-Seq (Yamane et al., 2011). The expressed group showed a bimodal distribution and was subdivided into trace expressed genes (<= 4.04 FPKM) and more highly expressed genes using a normal mixture model. The higher group was divided again into tertiles (low > 4.04 FPKM; medium > 23.5 FPKM; high > 65 FPKM). (B) Relative frequency (fe) of rearrangements in genes that are either silent or display trace, low, medium or high levels of transcription in activated B cells as determined by RNA-Seq. Dashed line indicates the expected rearrangement frequency based on a random model. $P < 0.001$ for all (permutation test).

A


count
400
300
200
100
0
0.01
0.1
1
10
100
1000
10000
FPKM
class
trace
low
medium
high

B


Myc^IAID^-/-
Relative Frequency
1.0
0
Silent
Trace
Low
Medium
High
Transcription Group

To confirm a bias for transcription and open chromatin, we categorized genes as either bearing or lacking rearrangements. Next, we asked whether genes in these two categories were overrepresented or underrepresented in the population of genes that bear activating histone marks and PolII as determined from ChIP-Seq on these entities in LPS and IL-4 activated B cells (Yamane et al., 2011). We find that rearrangements occur more frequently than expected (in a random distribution model) to genes bearing PolII and the activating histone marks H3K4 trimethylation, H3 acetylation, and H3K36 trimethylation ($P < 0.001$ for all, Figure 3.6). Similarly, rearrangements occur less frequently than expected to genes that do not load PolII or carry activating histone marks. Thus, there is a propensity for a DSB to recombine with gene rich regions of the genome and more specifically to transcription start sites of actively transcribed genes.

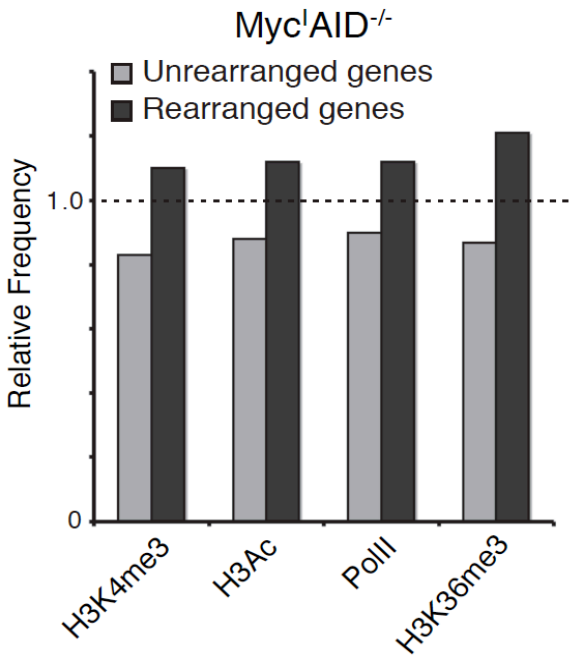


Figure 3.6. Rearrangements in activating histone mark- and PolII-associated gene groups. Relative frequency of rearrangements in PolII, H3K4 trimethylation, H3 acetylation and H3K36 trimethylation-associated gene groups (Yamane et al., 2011). Dashed line indicates expected frequency based on a random model. $P < 0.001$ for all samples (permutation test).

CHAPTER 4:

Rearrangements to *c-myc* and *IgH* in the Presence of AID

Processing of AID induced U:G mismatches can result in DSBs in *Ig* and non-*Ig* genes such as *c-myc* (Robbiani et al., 2008). To determine whether AID-mediated DSBs can be captured by TC-Seq we examined the *IgH* and *c-myc* loci in B cells expressing retrovirally encoded AID ($IgH^{I}AID^{RV}$ or $Myc^{I}AID^{RV}$). IgH^{I} B cells expressing both I-SceI and AID showed extensive AID-dependent rearrangement between the I-SceI site and downstream switch (S) regions (Figure 4.1). The frequency of rearrangements resembled the pattern of AID-mediated CSR in LPS+IL-4 cultures (e.g. IgG1>>IgG3>IgE), with 18,686 mapping to *Sγ1*, 3,192 to *Sγ3*, and 1,433 to *Sε* (Table 4.1). Notably, rearrangements mapped preferentially to the switch regions of the immunoglobulin constant region genes, which is the physiological site of AID activity.

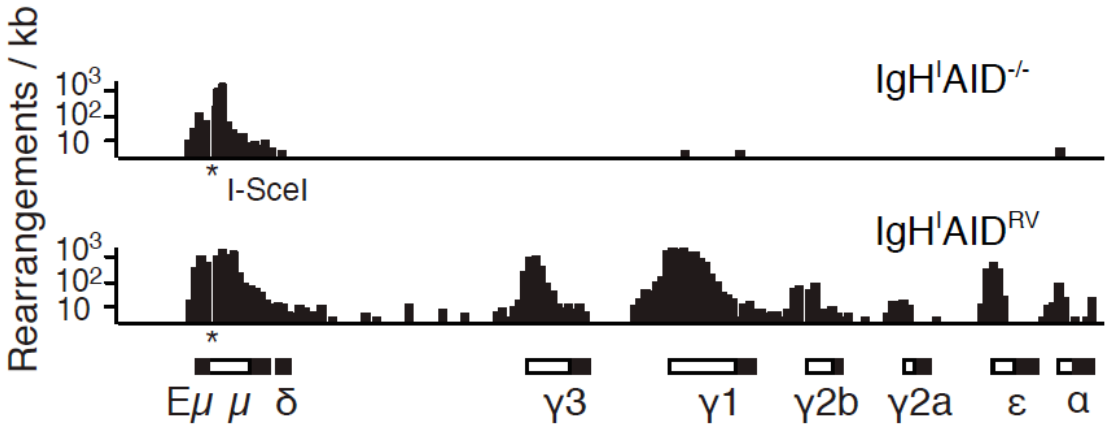


Figure 4.1. Rearrangements from IgH^{I} to downstream switch regions. Rearrangements per kb to I-SceI sites (indicated with an asterisk) in *IgH* in AID$^{-/-}$ (top panel) or AID^{RV} cells (bottom panel). White boxes in the schematics below each graph represent *Ig* switch domains while black boxes depict constant regions.

Table 4.1. Rearrangement abundance in specific loci.

		Rearrangements			
Gene	**Location**	**IgH^{I} AID^{RV}**	**IgH^{I} AID^{WT}**	**c-myc^{I} AID$^{-/-}$**	**c-myc^{I} AID^{RV}**
Ig mu	chr12:114644476-114668744	8,475	7,849	5	3,463
Ig gamma3	chr12:114594089-114606686	3,192	4,682	2	253
Ig gamma1	chr12:114563338-114582604	18,686	11,723	0	2,209
Ig gamma2b	chr12:114542220-114557966	284	513	0	19
Ig gamma2a	chr12:114525733-114535551	61	240	0	4
Ig epsilon	chr12:114507208-114518138	1,433	372	1	14
Ig alpha	chr12:114496093-114499984	10	5	0	1
c-myc	chr15:61814486-61825559	45	3	7,386	5,404

Table 4.1. Rearrangement abundance in specific loci. Total rearrangements to immunoglobulin constant domains (switch + constant) and *c-myc*. 100 million B cells were used per experiment.

Furthermore, TC-Seq detected translocations between *c-myc* and *IgH*, and these were entirely dependent on AID (Figures 4.2 and Table 4.1). In two biological replicate samples totaling 100 million B cells, we observed 45 translocations from IgH^I to *c-myc* (the I-SceI DSB was in *IgH*), and 5,925 from Myc^I to *IgH* (the I-SceI DSB was in *c-myc*) (Table 4.1). Additionally, TC-Seq tags mapping to *c-myc* from IgH^I correlate well with *c-myc/IgH* translocation breakpoints sequenced from primary B cells (Figure 4.2) (Robbiani et al., 2008). This suggests that TC-Seq reads are an accurate proxy for breakpoints. Furthermore, the data corroborate previous findings showing that AID induced breaks at *c-myc* are rate limiting for *c-myc*/*IgH* translocations (Robbiani et al., 2008) and suggest that AID-dependent *IgH* breaks are two orders of magnitude more frequent than those at *c-myc*. We conclude that TC-Seq captures rearrangements and translocations between DSBs in IgH^I or Myc^I and known AID targets.

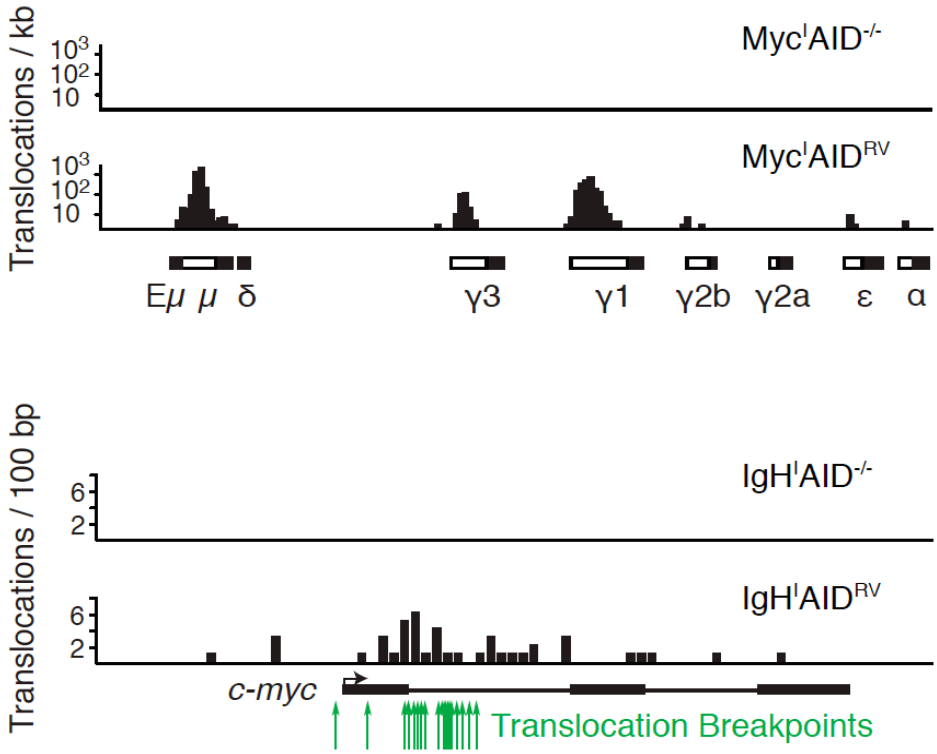


Figure 4.2. Screenshots of *c-myc/IgH* translocations. Rearrangements per kb to I-SceI sites in c-*myc* in AID$^{-/-}$ or AID^{RV} cells (top panel). White boxes in the schematics below each graph represent *Ig* switch domains while black boxes depict constant regions. Translocations per 100 bp from IgH^{I} to *c-myc* in AID$^{-/-}$ or AID^{RV} cells (bottom panel). Green arrows indicate *c-myc/IgH* translocation breakpoints sequenced from primary B cells (Robbiani et al., 2008).

As was the case for AID deficient samples, Myc^{I}AID^{RV} and IgH^{I}AID^{RV} libraries were enriched in intra-chromosomal rearrangements: 17% (10,633 of 63,772 total events) for Myc^{I} and 70% (36,019 of 51,312) for IgH^{I} (Figure 4.3A and 4.3B). The latter likely had more rearrangements on chromosome 12 due to the activity of AID at the *IgH* locus (Figure 4.3).

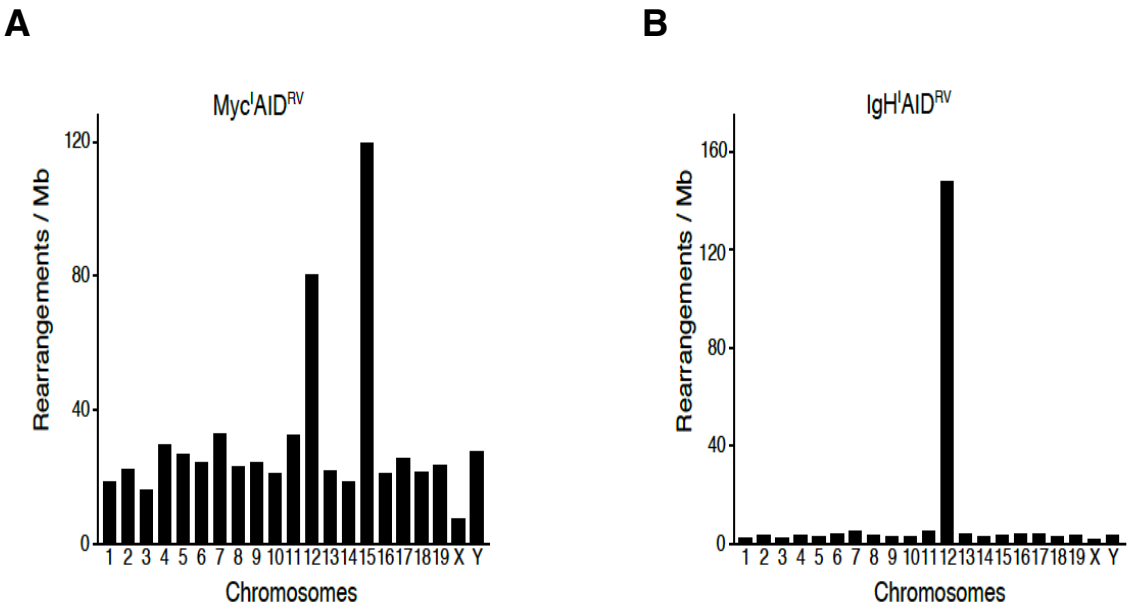


Figure 4.3 Chromosomal distribution of rearrangements in AID-sufficient Myc^{I} and IgH^{I} B cell genomes. (A) Bar graph displaying total rearrangements per mappable megabase per chromosome in Myc^{I}AID^{RV} and (B) IgH^{I}AID^{RV} B lymphocytes. In both cases cells were transduced with retroviruses (RV) expressing AID.

Expression of AID did not alter the distribution of events around Myc^{I}, with 72.5% (7,707 of 10,633) mapping within -50 kb to 300 kb of the break (Figure 4.4). A notable exception was an additional cluster of rearrangements associated with *Pvt1* exon 5 (Figure 4.4). These events coincided precisely with documented chromosomal translocations isolated from AID sufficient mouse plasmacytomas (Cory et al., 1985; Huppi et al., 1990) and likely represent an AID hotspot.

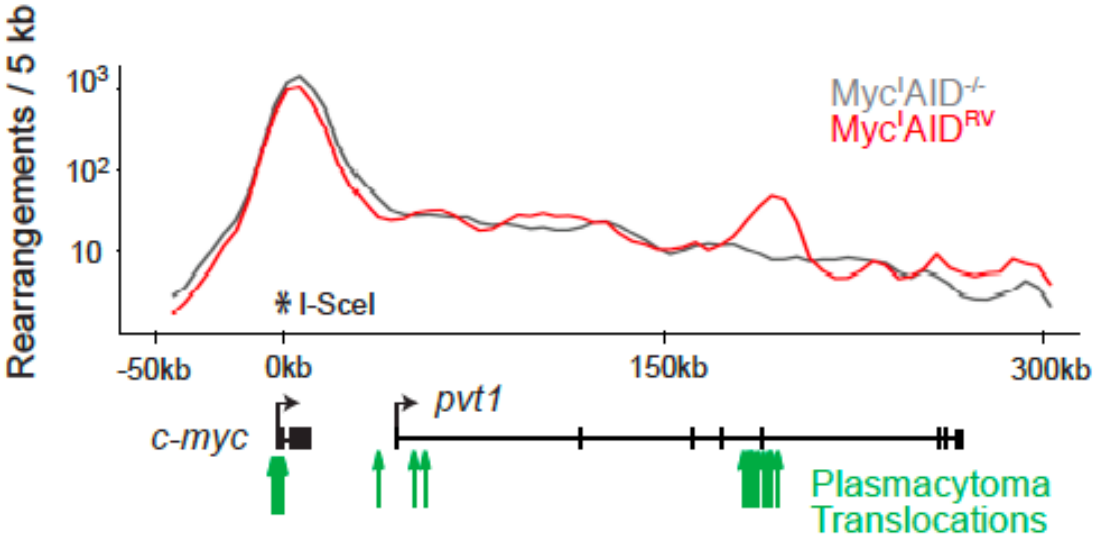


Figure 4.4. Profile of local rearrangements in Myc^{I} B lymphocytes, in the presence and absence of AID. Rearrangements per 5 kb around Myc^{I}, I-SceI site indicated with an asterisk. Distribution in the absence (gray line) or presence (red line) of AID. Green arrows indicate plasmacytoma translocation breakpoints.

In agreement with the Myc^{I}AID$^{-/-}$ samples, translocations between IgH^{I} or Myc^{I} and other chromosomes were evenly distributed throughout the genome, except for the Myc^{I} capture sample, which displayed a marked bias for chromosome 12 due to creation of DSBs at the *IgH* locus by AID (Figure 4.1). Similar to Myc^{I}AID$^{-/-}$ samples, rearrangements in both cases were more likely to occur in regions that are genic, transcriptionally active, recruiting PolII, and associated with activating histone marks (Figures 4.5 and 4.6A). Furthermore, intragenic rearrangements were enriched at transcription start sites of genes (Figure 4.6B).

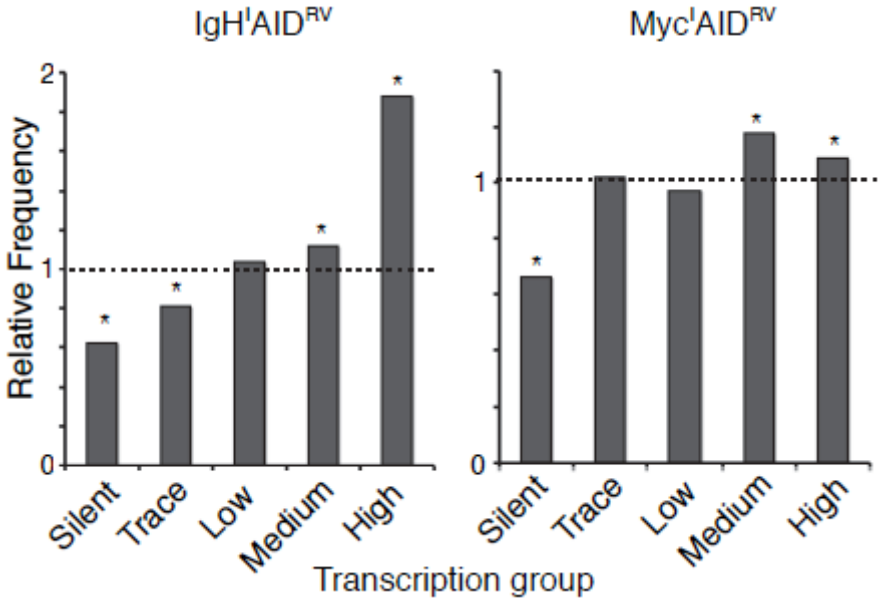


Figure 4.5 Rearrangements to transcription-level gene groups in AID^{RV} B cells. Relative frequency of rearrangements in transcription-level gene groups to I-SceI sites in AID sufficient B cells. IgH^{I}AID^{RV} is on the left while Myc^{I}AID^{RV} is on the right. Dashed line indicates expected frequency based on a random model. Asterisks highlight values with a $P < 0.001$ (permutation test).

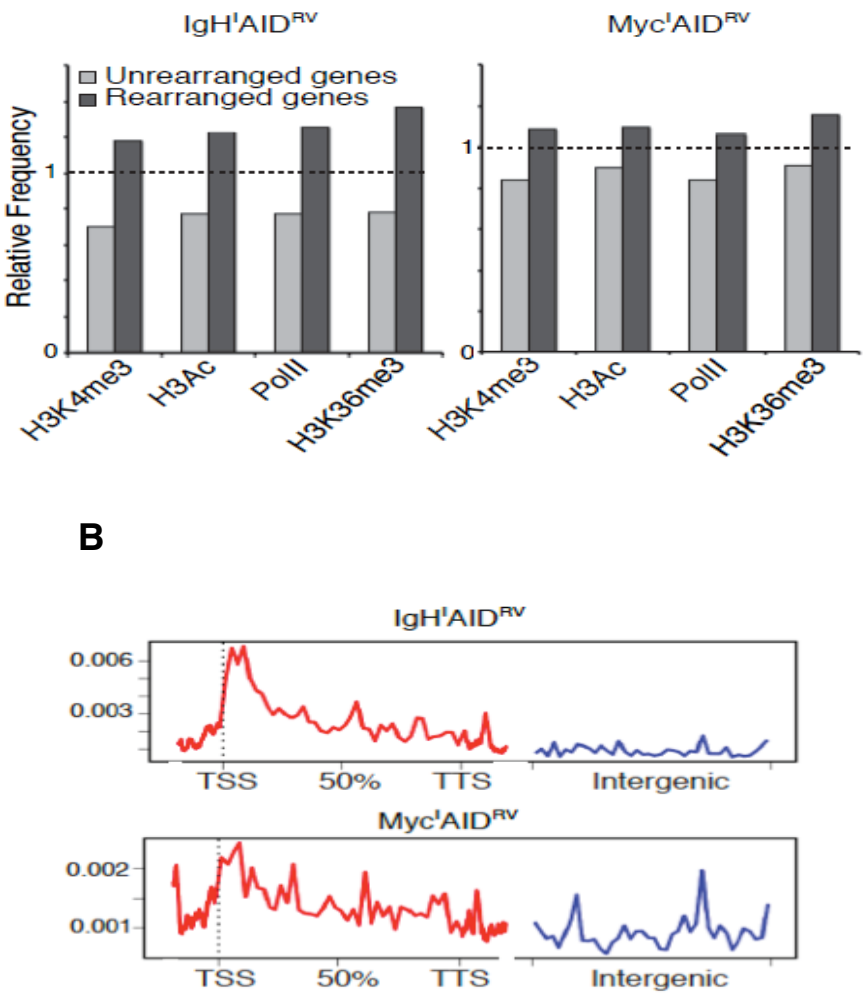


Figure 4.6. Rearrangements to histone and PolII-associated gene groups and gene body distribution in AID sufficient cells. (A) Relative frequency of rearrangements in PolII-associated or activating histone mark-associated gene groups (Yamane et al., 2011). IgH^{I} (left) or Myc^{I} (right) AID^{RV} captures. Dashed line indicates expected frequency based on a random model. $P < 0.001$ for all (permutation test). (B) Composite density profile of rearrangements, genic (red) and intergenic (blue) regions in IgH^{I} (upper) or Myc^{I} (lower) AID^{RV} captures.

In contrast to recent studies that used Nbs1 as an indirect marker of AID mediated damage (Staszewski et al., 2011), we found little or no difference in rearrangements to genomic repeats in the presence of AID (Table 4.2). Thus, AID does not dramatically alter the general profile of rearrangements.

Table 4.2. Rearrangements in intergenic repeats in $Myc^{I}AID^{RV}$ and $Myc^{I}AID^{-/-}$ samples.

Repeat Type	$Myc^{I}AID^{-/-}$	Frequency $Myc^{I}AID^{-/-}$	$Myc^{I}AID^{RV}$	Frequency $Myc^{I}AID^{RV}$	p-value
CA	507	0.0271	1528	0.028	0.5361
LINE	1084	0.0579	3137	0.0574	0.197
SINE	862	0.0461	2645	0.0484	0.7991
Others	2610	0.1395	7048	0.1291	0.0003
Total	18707		54605		

Table 4.2. Rearrangements in intergenic repeats in $Myc^{I}AID^{RV}$ and $Myc^{I}AID^{-/-}$ samples. For each category, the number of events in each repeat type and the fraction of all events falling in that type are shown. A Fisher‘s exact test was performed to compare AID sufficient and AID deficient frequencies. Total = number of rearrangements analyzed in that sample.

Next, we examined whether IgH^{I} and Myc^{I} capture DSB targets at similar rates. Indeed, in AID sufficient samples, total translocations from a given chromosome to IgH^{I} or Myc^{I} occurred at roughly similar frequencies (Figure 4.7).

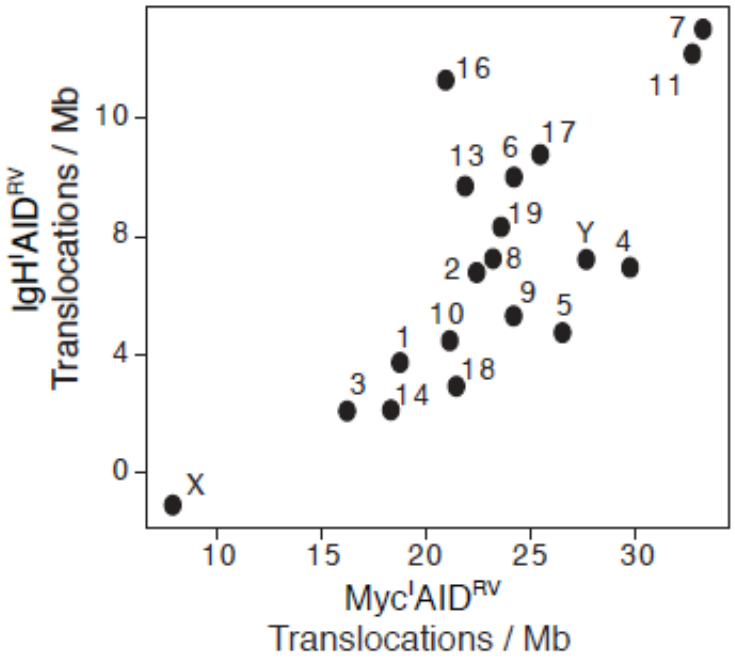


Figure 4.7. Translocations from Myc^{I} and IgH^{I} to transchromosomal targets. Graph comparing the number of translocations per mappable megabase to each chromosome from IgH^{I} (y axis) or Myc^{I} (x axis).

This similarity could be explained by the close physical proximity of *IgH* and *c-myc*, as suggested by studies with EBV-transformed B lymphoblastoid cells (Roix

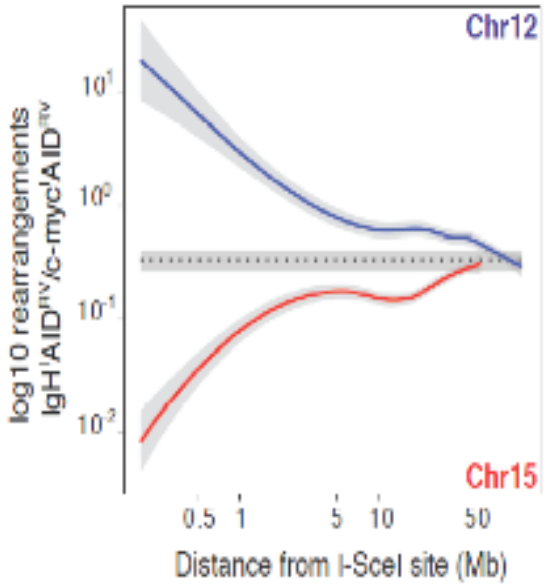
et al., 2003). Alternatively, the correlation in trans-chromosomal joining might represent random ligation between I-SceI DSBs in *IgH* or *c-myc* and DSBs on other chromosomes. This might also be a function of gene density; chromosome 7 and 11 are relatively gene rich while chromosome 3 is relatively gene poor. We conclude that extra-chromosomal DSBs ligate DSBs in IgH^{I} and Myc^{I} at similar rates.

To examine the nature of the intra-chromosomal rearrangement bias we calculated the ratio of IgH^{I} to Myc^{I} captured events for each 500 kb segment of the genome and compared the values for chromosome 12 and 15 to the trans-chromosomal joining average of 500kb windows genome-wide (Figure 4.8A). This analysis revealed that DSBs are preferentially captured intra-chromosomally and this effect diminishes at a rate inversely proportional to the distance from the I-SceI site ($d^{-1.29}$) (Figure 4.8B). This effect was most prominent locally but was evident at up to ~50 Mb away from the I-SceI break. We conclude that paired DSBs are preferentially joined intra-chromosomally and that the magnitude of this effect decreases with increasing distance between the two lesions.

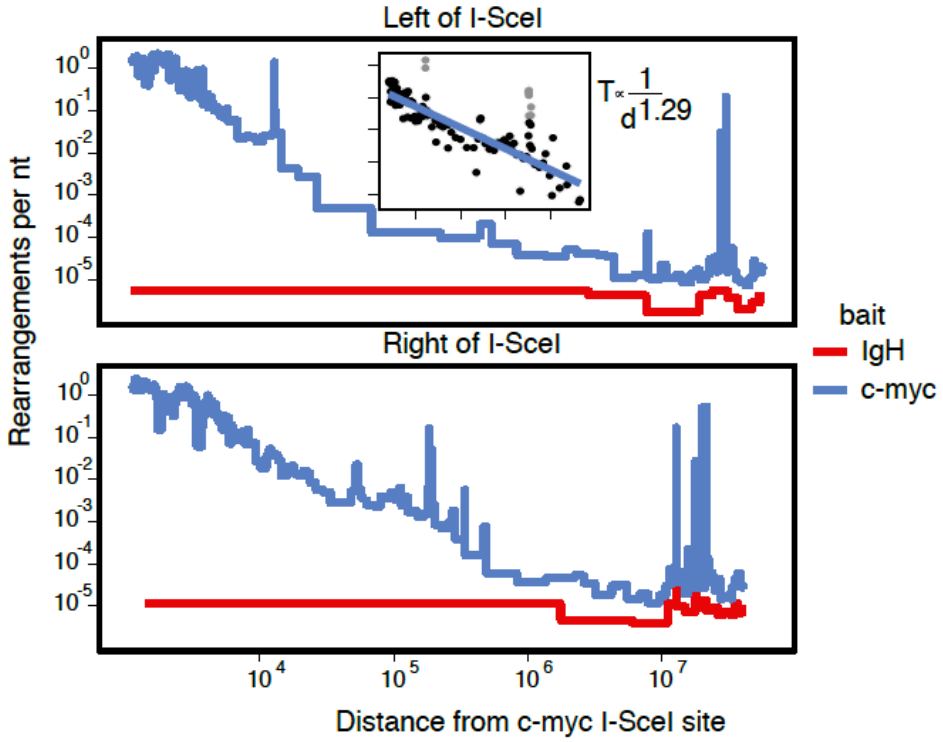
Figure 4.8. Distance dependent intra-chromosomal rearrangements.

(A) Ratio of IgH^{I} captured to Myc^{I} captured events in 500 kb bins moving away from the I-SceI capture site (both directions combined). Dotted line represents the average trans-chromosomal joining frequency computed on all chromosomes other than 12 or 15. Gray areas show 2 standard deviations around the mean. (B) Modeling intra-chromosomal rearrangement bias. Rearrangements to IgH^{I} (red) and Myc^{I} (blue) in the combined AID^{RV} samples were sorted into variable-sized bins of ≥20 events. Rearrangement density was calculated per bin to the left and right of *c-myc* separately as a step plot. When both directions were combined and plotted as a scatter plot on a log-log scale (inset graph, hotspots in grey) from 5 kb to 1 Mb, a roughly linear relationship suggested that the rearrangement density decreased as a function of a power of the distance. Fitting a linear model to the log transformed data excluding hotspots allowed the exponent of this power law to be estimated as approximately -1.3 (R^2=0.81).

A


Chr12
Chr15
log10 rearrangements IgH'AID-/-/c-myc'AID-/-
10^1
10^0
10^-1
10^-2
0.5
1
5
10
50
Distance from I-SceI site (Mb)

B


Left of I-SceI
Right of I-SceI
Rearrangements per nt
T∝1/d^1.29
bait
IgH
c-myc
10^4
10^5
10^6
10^7
Distance from c-myc I-SceI site

CHAPTER 5:

Rearrangement Hotspots in Activated B Cells

To determine whether there are hotspots for rearrangement, we searched the B cell genome for local accumulations of reads in AID deficient and sufficient samples. TC-Seq hotspots were defined as a localized enrichment of rearrangements above what is expected from a uniform genomic distribution. We removed likely artifacts; namely hotspots containing >80% of reads within DNA repeats, and those with footprints of <100nt (because translocations are amplified from randomly sonicated DNA (Figure 2.4A), deep-sequence tags associated with bona fide rearrangements are unlikely to map within a small region). We chose to omit these repeat alignments because while repeats may be analyzed in bulk (Table 4.2), the alignments from the reads may not be reliable. We identified 34 hotspots captured by Myc^{I} in the absence of AID (Table 5.1). There were 31 hotspots in 17 genes and 3 in non-genic regions. 17 of the hotspots were in *Pvt1*, within 500 kb of the I-SceI site.

Table 5.1. TC-Seq hotspot list for Myc^{I}AID^{KO} B cells.

Hotspot position	Total Rearran-gements	Closest gene idIname	Closest gene (nts)	Repeat type and cryptic I-SceI
chr15:61868606-61882447	61	19296IPvt1	in gene	all
chr15:61937605-61948372	48	19296IPvt1	in gene	all
chr15:61884209-61895375	37	19296IPvt1	in gene	
chr17:39980042-39983459	36	751530IMir715	in gene	
chr15:61905101-61913127	27	19296IPvt1	in gene	all
chr15:61896784-61903467	26	19296IPvt1	in gene	line, sine, LTR, DNA, simple, DNA
chr2:155630784-155631968	26	17391IMmp24	in gene	
chr7:112726654-112731060	19	11785IApbb1	in gene	sine, line, near cryptic IsceI site
chr15:61919292-61924833	18	19296IPvt1	in gene	sine, ltr
chr15:61928582-61933207	17	19296IPvt1	in gene	line, sine, LTR, DNA, simple (GT, GA)
chr15:81735124-81735608	14	11429IAco2	in gene	
chr15:61950188-61954544	12	19296IPvt1	in gene	
chr15:61914451-61917481	11	19296IPvt1	in gene	
chr1:137751600-137752931	11	21956ITnnt2	2762	near cryptic IsceI site
chr15:61996073-61999322	8	19296IPvt1	in gene	
chr9:107230861-107231335	7	69536IHemk1	in gene	sine
chr15:16219195-16219312	7	12565ICdh9	488543	near cryptic IsceI site
chr1:37704468-37704668	7	72097I2010300C02Rik	in gene	
chr15:62050249-62052811	7	19296IPvt1	in gene	
chr14:46096170-46097095	6	218952IFermt2	in gene	
chr15:61989531-61991172	6	19296IPvt1	in gene	
chr2:155633401-155634287	5	17391IMmp24	in gene	sine, line
chrX:11376878-11378216	5	71458IBcor	235647	LTR, near cryptic IsceI site
chr15:62025428-62027589	5	19296IPvt1	in gene	line, ltr, simple (CCAA)

chr15:61987000-61987927	5	19296lPvt1	in gene	
chr15:62075361-62075946	4	19296lPvt1	in gene	line
chr8:96450914-96451204	4	14681lGnao1	in gene	simple (CA)
chr7:6872248-6872359	4	57775lUsp29	in gene	simple (ca)
chr19:44375405-44375796	4	20250lScd2	in gene	near cryptic IsceI site
chr2:132665496-132665672	3	66634lMcm8	in gene	Sine, Simple (GT)
chr15:61975151-61975421	3	19296lPvt1	in gene	
chr6:85346261-85346531	3	52055lRab11fip5	21633	
chr15:62042419-62042698	3	19296lPvt1	in gene	
chr7:80884795-80885210	3	100038347lFam174b	in gene	

Table 5.1. TC-Seq hotspot list for Myc^{I}AID^{KO} B cells. For each hotspot the number of rearrangements constituting the hotspot, the closest gene, the distance to that gene, and the type of repetitive DNA elements contained in the hotspot are shown. Hotspots attributed to cryptic I-SceI sites are indicated in the last column.

In addition, 2 hotspots occurred within 5 kb of cryptic I-SceI sites (each bearing one mismatch to the 18- base pair recognition sequence). For example, one such hotspot at chr15:16219195-16219312 containing a 1-off I-SceI recognition sequence bore 8 rearrangements (Figure 5.1). A genome-wide search for rearrangements within 5 kb of cryptic I-SceI sites (83 within the mouse genome with 1 or 2 mismatches to the canonical I-SceI recognition sequence) yielded 57 events, 7 times more than expected in a random distribution model. When allowing up to 6 mismatches we find a total of 5 out of 17 AID-independent hotspots near putative cryptic I-SceI sites. Although I-SceI has been used to generate a unique DSB in gene targeting and DNA repair experiments, our data suggest that DNA recognition by I-SceI can be promiscuous in the mouse genome, as demonstrated for other yeast endonucleases (Argast et al., 1998).

Next, we searched the AID sufficient data sets for AID dependent hotspots. These were defined as genomic locations where there was at least a 10-fold enrichment over background rearrangement levels. In contrast to $AID^{-/-}$, we found 157 hotspots in 83 genes captured by $IgH^{I}AID^{RV}$ and 60 hotspots in 37 genes by $Myc^{I}AID^{RV}$ in 100 million B cells. (Tables 5.2 and 5.3). This suggests that AID is generating targeted genomic instability above background levels.

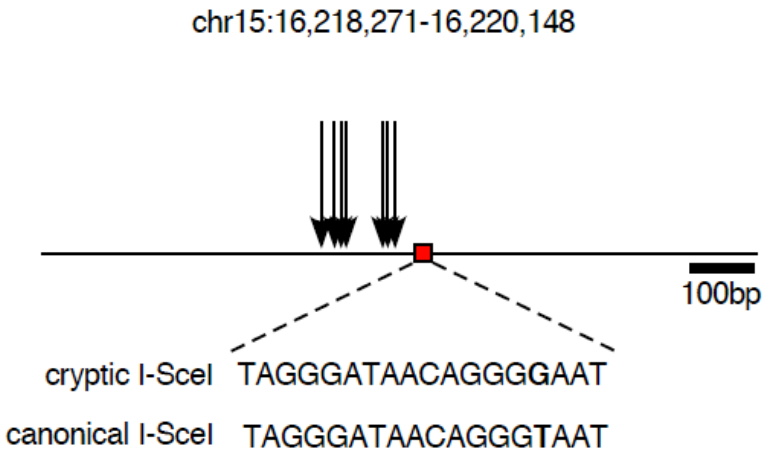


Figure 5.1. I-SceI cryptic site screenshot. Screenshot showing a TC-Seq hotspot associated with a cryptic I-SceI site located on chromosome 15 in $Myc^{I}AID^{-/-}$ B cells. Rearrangements accumulate within 200 bp of a site that differs by 1 base (T->G) from the canonical I-SceI recognition sequence.

Table 5.2. TC-Seq hotspots in Myc^{I}AID^{RV} B cells.

Hotspot position	Total Rearran-gements	Closest gene idlname	Closest gene (nts)
chr12:114569045-114578835	2063	IgHIIgH	in gene
chr12:114600202-114603869	259	IgHIIgH	in gene
chr7:132695991-132699453	141	16190Il4ra	in gene
chr13:43880182-43885076	132	12522ICd83	in gene
chr15:62001300-62004099	128	19296IPvt1	in gene
chr15:74761629-74764081	85	17069ILy6e	21399
chr11:61466811-61469564	74	71520IGrap	in gene
chr15:62009859-62012742	69	19296IPvt1	in gene
chr15:74785846-74789190	57	17069ILy6e	in gene
chr17:29627912-29629491	49	18712IPim1	in gene
chr12:114665845-114667844	48	IgHIIgH	in gene
chr15:82040871-82042918	28	20788ISrebf2	5481
chr10:53793003-53795422	27	17155IMan1a	in gene
chr7:52091757-52092909	25	14204Il4i1	in gene
chr15:34012782-34013570	24	67154IMtdh	in gene
chr15:79724869-79726419	23	80287IApobec3	in gene
chr16:84706683-84708187	20	387173IMir155	6197
chr4:44714489-44715772	20	18507IPax5	in gene
chr16:84703741-84705601	19	387173IMir155	8783
chr8:129115681-129117132	17	270110Irf2bp2	in gene
chr15:62157722-62158587	15	19296IPvt1	75192
chr16:19061770-19063197	14	IgLlambdaIgLlambda	in gene
chr7:132747538-132748234	13	60504Il21r	in gene
chr4:44716165-44717120	13	18507IPax5	in gene
chr11:87569331-87570257	12	387160IMir142	108
chr15:62000011-62000612	11	19296IPvt1	in gene
chr4:44722019-44722683	11	18507IPax5	in gene
chr6:70676375-70677272	11	IgLkappaIgLkappa	in gene
chr9:44333604-44334132	9	12145ICxcr5	in gene
chr4:44981291-44982237	9	76238IGrhpr	12045
chr11:86398867-86399824	9	75909ITmem49	in gene
chr15:74827015-74828130	8	110454ILy6a	in gene
chr11:61470062-61470857	7	71520IGrap	in gene
chr8:35871344-35872045	7	319520IDusp4	in gene
chr6:129134035-129135134	7	93694IClec2d	in gene
chr15:62160533-62161287	6	19296IPvt1	78003
chr15:61903483-61904141	6	19296IPvt1	in gene

chr15:74784468-74785172	6	17069|Ly6e	308
chr7:132750068-132750375	6	60504|Il21r	in gene
chr8:73223046-73223403	6	16478|Jund	in gene
chr9:107199973-107201059	6	12700|Cish	in gene
chr8:87501360-87502454	6	16477|Junb	in gene
chr15:82980236-82980359	5	104307|Rnu12	12
chr12:114550576-114550791	5	IgH|IgH	in gene
chr3:95463197-95463519	5	17210|Mcl1	in gene
chr10:17443755-17444131	5	17684|Cited2	in gene
chr10:59681817-59682230	5	53374|Chst3	in gene
chr10:59680759-59681209	5	53374|Chst3	in gene
chr12:114551620-114552139	5	IgH|IgH	in gene
chr15:79913571-79914224	5	27367|Rpl3	in gene
chr6:127100691-127101260	4	12444|Ccnd2	in gene
chr15:61897189-61897630	4	19296|Pvt1	in gene
chr15:74756575-74757098	4	17069|Ly6e	28382
chr15:62156632-62157160	4	19296|Pvt1	74102
chr13:46059069-46059856	4	20238|Atxn1	in gene
chr12:114566773-114567144	4	IgH|IgH	in gene
chr15:62020486-62020920	4	19296|Pvt1	in gene
chr9:44143288-44143844	4	15270|H2afx	in gene
chr17:35138064-35138737	4	22321|Vars	in gene
chr15:62105382-62105511	3	19296|Pvt1	22852

Table 5.2. TC-Seq hotspots in Myc^{I}AID^{RV} B cells. For each hotspot the genomic location (chromosome:base#-base#), the number of events in that hotspot and its most proximal gene is shown. If a hotspot is in a gene the last column indicates this, if the hotspot is in an intergenic region the last column gives the number of bases away from the nearest gene. Only AID-dependent hotspots are shown.

Table 5.3. TC-Seq hotspots in IgH^{I}AID^{RV} B cells. (Pages 106-109)

Hotspot position	Total Rearrang-ements	Closest gene id|name	Closest gene (nts)
chr12:114563677-114583617	18010	IgH|IgH	in gene
chr12:114592880-114605961	3415	IgH|IgH	in gene
chr12:114464291-114473409	397	IgH|IgH	24991
chr7:132695160-132701616	355	16190|Il4ra	in gene
chr12:113826127-113831888	306	70435|Inf2	in gene
chr12:114477632-114487105	198	IgH|IgH	11295
chr13:43880627-43885981	158	12522|Cd83	in gene
chr12:114545002-114549279	149	IgH|IgH	in gene
chr16:84703733-84708651	142	387173|Mir155	5733
chr11:61466722-61471367	129	71520|Grap	in gene
chr12:114550562-114553023	120	IgH|IgH	in gene
chr12:114487888-114495960	103	IgH|IgH	2440
chr6:129131713-129135370	95	93694|Clec2d	in gene
chr6:70672635-70676968	90	IgLkappa|IgLkappa	in gene
chr12:81210545-81214030	89	12192|Zfp36l1	in gene
chr17:29627857-29630975	74	18712|Pim1	in gene
chr12:114529738-114534115	61	IgH|IgH	in gene
chr12:114460798-114463471	60	69195|Tmem121	33065
chr7:132746979-132749248	58	60504|Il21r	in gene
chr16:58715676-58718789	52	71223|Gpr15	in gene
chr7:52091646-52094659	49	100328588|Nup62-il4i1	in gene
chr8:129115389-129117233	43	270110|Irf2bp2	in gene
chr15:61817197-61819111	42	17869|Myc	in gene
chr12:116745900-116747837	38	IgH|IgH	in gene
chr10:53793549-53795607	38	17155|Man1a	in gene
chr12:115526476-115529230	35	IgH|IgH	in gene
chr11:87569636-87571044	31	387160|Mir142	in gene
chr12:115984499-115986323	29	IgH|IgH	in gene
chr12:114560586-114562879	29	IgH|IgH	in gene
chr12:116861441-116863848	27	IgH|IgH	in gene
chr12:117150666-117151414	27	IgH|IgH	in gene
chr12:88539032-88541058	27	382620|Tmed8	in gene
chr12:115821793-115822684	26	IgH|IgH	in gene
chr12:114096947-114100451	26	56696|Gpr132	in gene
chr16:19061101-19062590	26	IgLlambda|IgLlambda	in gene
chr12:116152849-116154287	25	IgH|IgH	in gene
chr13:63896445-63900869	25	76251|0610007P08Rik	15758
chr3:135352008-135353935	24	18033|Nfkb1	in gene
chr9:107199367-107200994	20	12700|Cish	in gene
chr17:51970428-51972180	20	20230|Satb1	in gene
chr12:114606862-114609279	20	IgH|IgH	in gene
chr12:114557649-114560033	19	IgH|IgH	in gene
chr3:95462675-95465208	19	17210|Mcl1	in gene
chr11:106173700-106175830	19	15985|Cd79b	in gene
chr12:114835673-114836176	18	IgH|IgH	in gene
chr11:115487516-115488927	18	67283|Slc25a19	in gene
chr12:116886227-116888201	17	IgH|IgH	in gene
chr12:115097413-115098823	17	IgH|IgH	in gene

chr11:86398657-86400144	17	75909|Tmem49	in gene
chr16:19065068-19066725	17	IgLlambda|IgLlambda	in gene
chr8:35870456-35872041	15	319520|Dusp4	in gene
chr12:70298030-70299253	15	109065|1110034A24Rik	in gene
chr8:73223130-73224638	15	16478|Jund	in gene
chr12:117158582-117159779	14	IgH|IgH	in gene
chr12:115630002-115631365	14	IgH|IgH	in gene
chr9:107201768-107203198	13	12700|Cish	in gene
chr14:122361101-122363562	13	321019|Gpr183	in gene
chr12:116532327-116532749	12	IgH|IgH	in gene
chr12:116383843-116384804	12	IgH|IgH	in gene
chr12:56591584-56593525	12	18035|Nfkbia	in gene
chr12:117086102-117086768	11	IgH|IgH	in gene
chr12:70397988-70398524	11	69554|Klhdc2	in gene
chr4:44725889-44726961	11	18507|Pax5	2577
chr7:133556483-133558100	10	12478|Cd19	in gene
chr13:53014444-53014988	10	11992|Auh	in gene
chr12:115560605-115561181	10	IgH|IgH	in gene
chr4:44980880-44981502	10	76238|Grhpr	12780
chr13:23713187-23714124	10	50709|Hist1h1e	in gene
chr3:96073011-96074020	10	319189|Hist2h2bb	in gene
chr12:87027653-87028719	10	53314|Batf	in gene
chr4:44715829-44717064	10	18507|Pax5	in gene
chr2:158492484-158494071	10	228852|Ppp1r16b	in gene
chr13:53072809-53074649	10	18030|Nfil3	in gene
chr9:44333703-44334267	10	12145|Cxcr5	in gene
chr12:115076892-115077564	9	IgH|IgH	in gene
chr12:115391414-115392751	9	IgH|IgH	in gene
chr2:166888980-166890860	9	68949|1500012F01Rik	in gene
chr12:115776774-115777349	9	IgH|IgH	in gene
chr8:87501548-87502456	9	16477|Junb	in gene
chr12:114610038-114611149	8	IgH|IgH	in gene
chr9:44412166-44412743	8	13209|Ddx6	231
chr16:58672312-58673522	8	12892|Cpox	in gene
chr1:108068896-108070111	8	98432|Phlpp1	in gene
chr12:115232761-115233011	8	IgH|IgH	in gene
chr12:115707425-115707742	8	IgH|IgH	in gene
chr12:114955011-114955614	8	IgH|IgH	in gene
chr6:136752518-136753244	8	320332|Hist4h4	in gene
chr12:114555430-114556333	8	IgH|IgH	in gene
chr10:59681922-59682995	8	53374|Chst3	in gene
chr12:116175390-116175840	7	IgH|IgH	in gene
chr12:115008112-115009040	7	IgH|IgH	in gene
chr8:97269761-97270072	7	20299|Ccl22	in gene
chr12:116733901-116734666	7	IgH|IgH	in gene
chr12:117072263-117072562	7	IgH|IgH	in gene
chr6:129130592-129131098	7	93694|Clec2d	in gene
chr12:115717148-115717715	7	IgH|IgH	in gene
chr12:116229318-116230137	7	IgH|IgH	in gene
chr4:44714015-44714884	7	18507|Pax5	in gene
chr12:114553758-114554896	7	IgH|IgH	in gene
chr15:74785673-74786897	7	17069|Ly6e	in gene
chr16:10784789-10785123	7	12703|Socs1	in gene

chr15:74763628-74764024	7	17069|Ly6e	21456
chr12:115134720-115134990	6	IgH|IgH	in gene
chr7:132750171-132750297	6	60504|Il21r	in gene
chr17:35136839-35137019	6	22321|Vars	832
chr12:116831379-116831823	6	IgH|IgH	in gene
chr4:44721635-44722251	6	18507|Pax5	in gene
chr12:115414770-115415436	6	IgH|IgH	in gene
chr15:79911701-79912444	6	27367|Rpl3	in gene
chr13:51942199-51943030	6	23882|Gadd45g	in gene
chr6:48655721-48656684	6	231931|Gimap6	in gene
chr15:82041126-82042151	6	20788|Srebf2	5736
chr1:36365470-36366278	5	214855|Arid5a	in gene
chr12:116396756-116396922	5	IgH|IgH	in gene
chr11:5421596-5421866	5	22433|Xbp1	in gene
chr12:116930165-116930446	5	IgH|IgH	in gene
chr12:115251470-115251824	5	IgH|IgH	in gene
chr3:96024349-96024764	5	319176|Hist2h2ac	in gene
chr15:74787652-74788420	5	17069|Ly6e	in gene
chr12:114729327-114730211	5	IgH|IgH	in gene
chr12:114497586-114498517	5	IgH|IgH	in gene
chr6:48696841-48697785	5	317757|Gimap5	in gene
chr11:61472162-61473172	5	71520|Grap	in gene
chr12:55797689-55797969	5	66266|Eapp	837
chr12:117107309-117107806	5	IgH|IgH	in gene
chr12:92827477-92828227	5	83602|Gtf2a1	in gene
chr2:34812256-34812641	4	22029|Traf1	in gene
chr12:115609552-115610249	4	IgH|IgH	in gene
chr12:115855867-115856592	4	IgH|IgH	in gene
chr6:48634895-48635772	4	107526|Gimap4	in gene
chr12:115117746-115118111	4	IgH|IgH	in gene
chr9:32347803-32348302	4	14247|Fli1	in gene
chr2:127952144-127952647	4	12125|Bcl2l11	in gene
chr12:115298061-115298583	4	IgH|IgH	in gene
chr12:113858703-113859237	4	11565|Adssl1	in gene
chr17:88463676-88464261	4	225055|Fbxo11	in gene
chr1:162966442-162967042	4	14455|Gas5	in gene
chr13:21871724-21872334	4	56702|Hist1h1b	in gene
chr15:34012549-34013176	4	67154|Mtdh	in gene
chr6:48689187-48689841	4	16205|Gimap1	in gene
chr7:108257838-108258501	4	207278|Fchsd2	in gene
chr16:10783339-10784041	4	12703|Socs1	in gene
chr10:59413651-59414365	4	74747|Ddit4	in gene
chr14:70688210-70689003	4	19057|Ppp3cc	in gene
chr12:115252782-115253686	4	IgH|IgH	in gene
chr4:134479505-134480566	4	27981|D4Wsu53e	in gene
chr12:115187722-115188045	4	IgH|IgH	in gene
chr13:63895201-63895639	4	76251|0610007P08Rik	20988
chr16:58720437-58720884	4	71223|Gpr15	1600
chr7:132634208-132634668	4	67711|Nsmce1	in gene
chr12:116011816-116012535	4	IgH|IgH	in gene
chr10:59634442-59634618	3	53374|Chst3	9657
chr12:117238411-117238513	3	IgH|IgH	in gene
chr1:38720806-38720934	3	16764|Aff3	in gene

chr4:44724505-44724640	3	18507|Pax5	1193
chr12:116550486-116550640	3	IgH|IgH	in gene
chr5:114281912-114282032	3	20345|Selplg	1402

Table 5.3. TC-Seq hotspots in IgH^{I}AID^{RV} B cells. For each hotspot the genomic location (chromosome:base#-base#), the number of events in that hotspot and its most proximal gene is shown. If a hotspot is in a gene the last column indicates this, if the hotspot is in an intergenic region the last column gives the number of bases away from the nearest gene. Only AID-dependent hotspots are shown.

Next, we sought to characterize the nature of AID targets. Consistent with a translocation coupled targeting mechanism, 80% of the AID-dependent hotspots captured by *c-myc* and 90% of those captured by *IgH* were within genes. For example, we found robust AID-dependent hotspots on *Il4i1* and *Pax5* (a recurring *IgH* translocation partner in lymphoplasmacytoid lymphoma (Kuppers, 2005)) (Figure 5.2 and Tables 5.2 and 5.3).

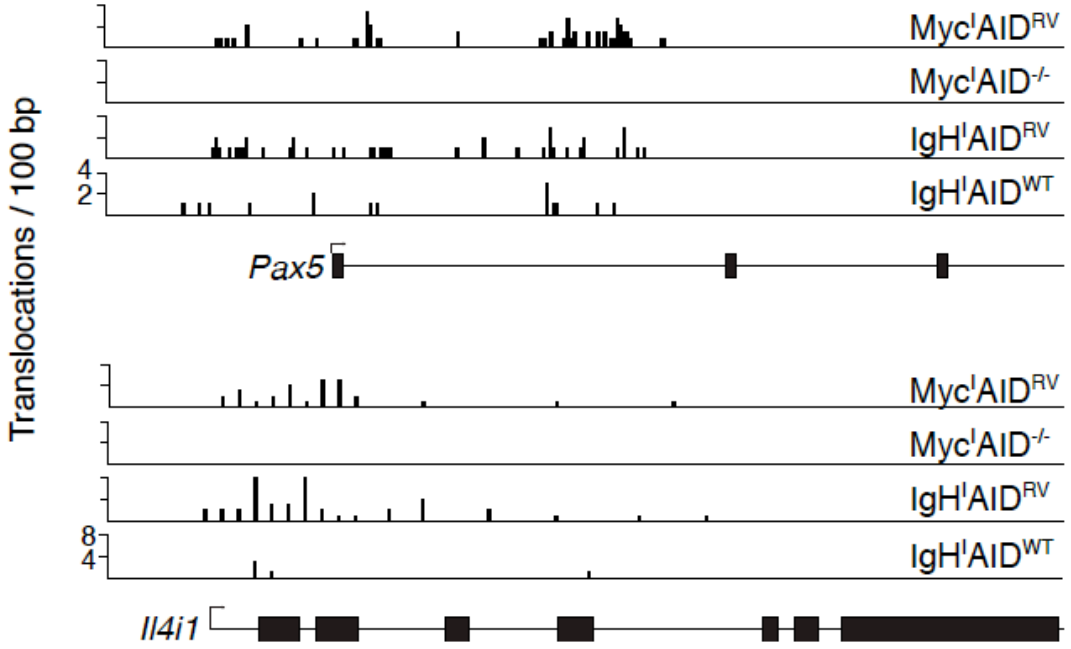


Figure 5.2. Screenshots of AID-dependent rearrangement hotspots. Translocations per 100 bp present at *Il4i1* and *Pax5* genes in all samples. Arrows indicate the direction of transcription. Black boxes represent exons while lines represent introns.

AID-dependent hotspots were similar for IgH^{I} B cells expressing wild type levels (WT) or retrovirally over-expressed (RV) AID, however the number of TC-Seq captured events per hotspot was decreased in the former (Figures 5.2 and 5.3). Therefore, translocations to AID targets occur in cells expressing physiological levels of AID and hotspots are not dependent on AID over-expression. 3 hotspots observed in AID^{WT} samples but not AID^{RV} samples are likely due to experimental variability. We conclude that AID produces substrates for translocations in a number of discreet sites throughout the genome, and these sites are mainly in genes.

Figure 5.3. Overlap of hotspots in AID^{WT} and AID^{RV} IgH^{I} samples. Venn diagram showing the number of shared and exclusive AID-dependent hotspot-bearing genes in IgH^{I}AID^{RV} and IgH^{I}AID^{WT} experiments.

Consistant with equal capture rates of extrachromosomal targets by *c-myc* and *IgH* (Figure 4.7) Genes containing AID-dependent hotspots overlapped between IgH^{I}AID^{RV} and Myc^{I}AID^{RV} samples (Figure 5.4).

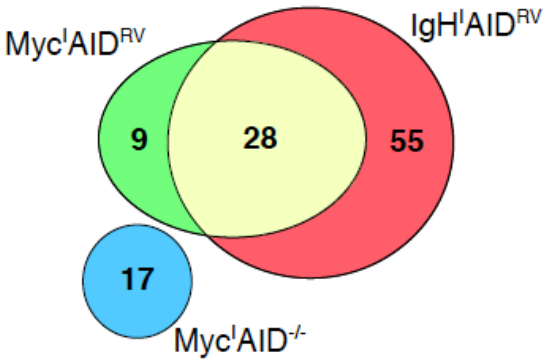


Figure 5.4. Overlap of hotspots in IgH^{I}AID^{RV}, Myc^{I}AID^{RV} and Myc^{I}AID^{KO} samples. Venn diagram showing the number of shared and exclusive AID-dependent hotspot-bearing genes in the three experiments.

We found that 28 of the frequently translocated targets were shared. In contrast, there were a number of unique intra-chromosomal AID-dependent hotspots. For example, rearrangements to *Inf2* on chromosome 12 (~850 kb from IgH^{I}) were only found by IgH^{I} capture while rearrangements near *Pvt1* on chromosome 15 (up to ~350 kb from Myc^{I}) were only found by Myc^{I} capture (Tables 5.2 and 5.3). Thus, there is a bias towards recombination between I-SceI breaks and AID

hotspots within the same chromosome. Additionally, the finding that some hotspots are only captured in *cis* indicates that TC-Seq underestimates the number of AID mediated DSBs in the genome and suggests that we have not reached saturation. Additionally, we note that AID-independent hotspots do not arise in the AID-sufficient sample. We propose that this is due to a higher availability of AID-generated DSBs, which saturate the rearrangement profile to I-SceI generated ends.

Combined analysis of the IgH^{I} and Myc^{I} TC-Seq data sets and mRNA-Seq data shows that AID dependent hotspots are primarily found in transcribed genes (Figure 5.5A). However, while nearly all of the translocated genes are actively transcribed, there is no clear correlation between transcript abundance and rearrangement frequency (Figure 5.5B). Furthermore, ~2000 highly transcribed genes are not rearranged (Figure 5.5A, shaded area). Therefore transcription is necessary but not sufficient for AID targeting, and transcription levels alone cannot account for AID-dependent DSBs.

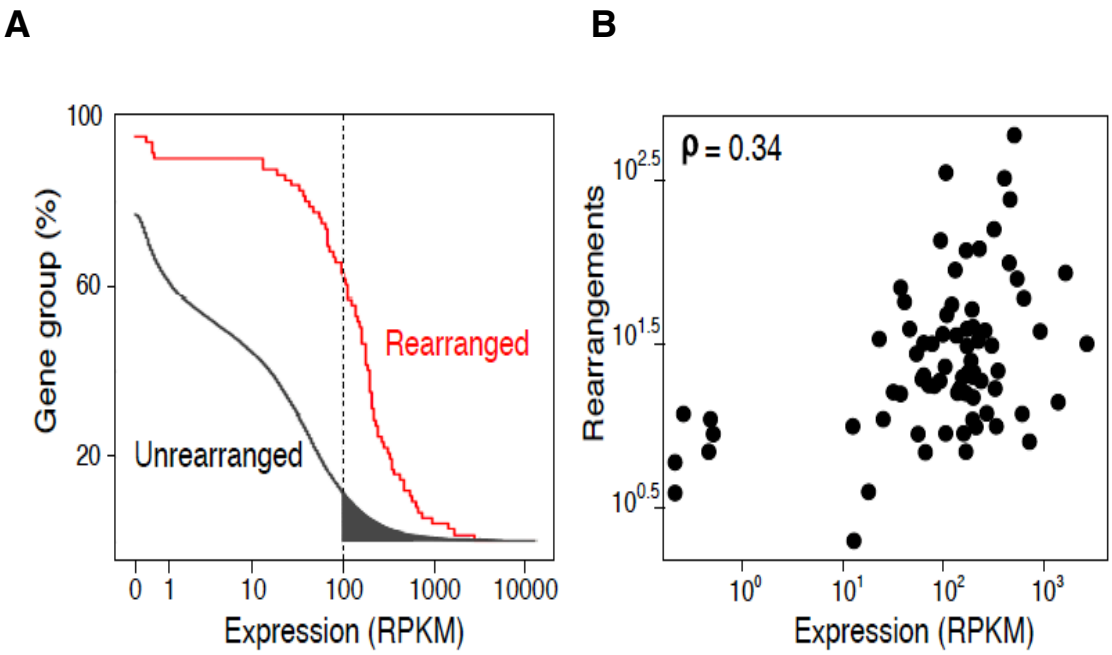


Figure 5.5. Correlating AID dependent hotspots with transcription.

(A) Empirical cumulative distribution analysis showing transcript abundance in genes displaying (red) or lacking (black) AID dependent rearrangement hotspots. Filled-in gray slice represents ~2000 highly transcribed yet unrearranged genes. (B) Total rearrangements as a function of gene expression (mRNA-Seq) (Yamane et al., 2011) in genes bearing AID dependent rearrangement hotspots. Each gene is plotted as one point.

Additionally, AID-dependent hotspots are biased to the region around the transcription start site (Figure 5.6A). This finding is consistent with the accumulation of AID and Spt5 around the promoters of stalled genes and the distribution of somatic hypermutation (Pavri et al., 2010; Yamane et al., 2011). Indeed, AID-dependent TC-Seq hotspots overlap with regions of AID and Spt5 accumulation (Figures 5.6B and 5.7).

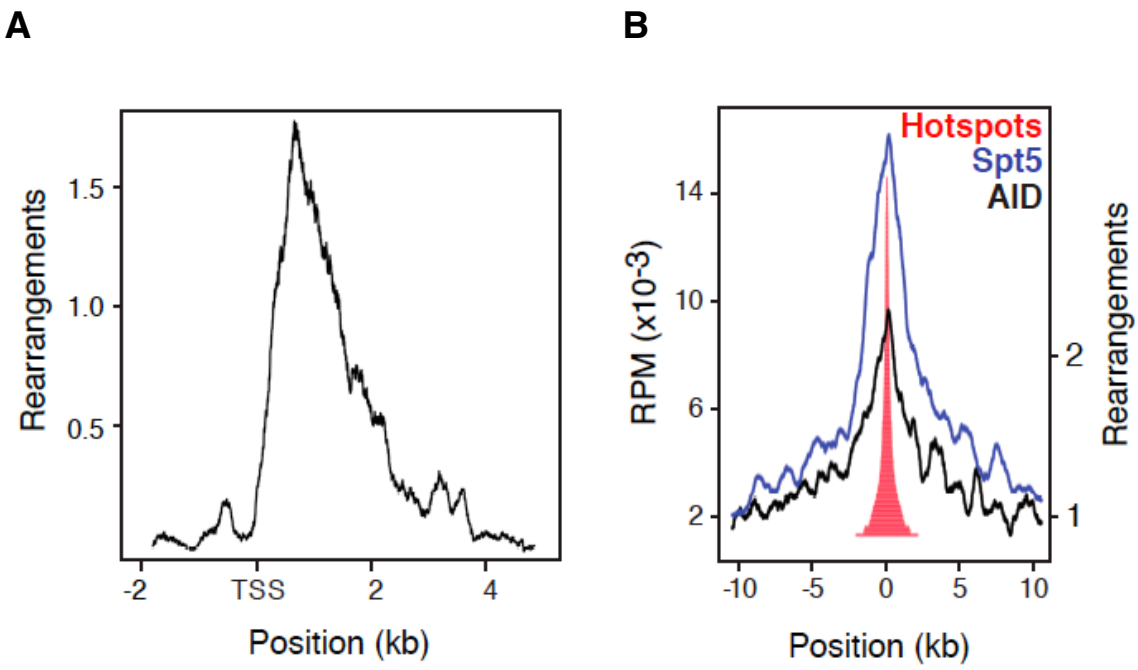


Figure 5.6. Genomic features at AID-dependent hotspots. (A) Composite density graph showing the distribution of rearrangements in genes associated with AID dependent hotspots relative to the TSS. (B) Spt5 and AID recruitment at genomic sites associated with translocation hotspots (Pavri et al., 2010; Yamane et al., 2011).

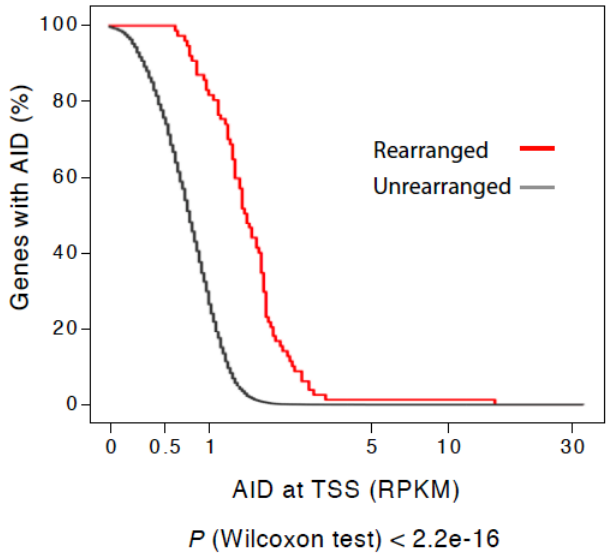


Figure 5.7. AID loading at AID-dependent rearrangement hotspots.

Empirical cumulative distribution function showing AID accumulation at TSSs of genes displaying (red) or lacking (black) AID-dependent TC-Seq hotspots. AID association data was derived from AID ChIP-Seq experiments (Yamane et al., 2011).

This correlation prompted us to explore the relationship between AID activity and accumulation of chromosomal translocations by measuring somatic hypermutation at TC-Seq captured AID targets (Yamane et al., 2011) (Table 5.4). To increase sensitivity we examined AID mediated mutations in *IgK*AID/*UNG*$^{-/-}$ B cells (AID over expression by virtue of a *VKappa* promoter increases AID levels and therefore, mutation rates). In the absence of UNG mutation rates are further increased because U:G mismatches are processed into mutations rather than

DSBs (Schrader et al., 2005; Xue et al., 2006). We found a positive correlation (Spearman coefficient = 0.84) between hypermutation and rearrangement frequency (Figure 5.8). All genes analyzed with a mutation frequency over $10x10^{-5}$ bear rearrangements, and all genes with AID-dependent TC-Seq hotspots show mutations (Figure 5.8). Rearrangements were only seen rarely in genes with lower rates of mutation. This suggests that the frequency of hypermutation and the frequency of AID-induced DSBs are directly proportional. We conclude that AID-dependent TC-Seq hotspots occur on stalled genes that can accumulate Spt5, AID, and may suffer hypermutation relative to background levels.

Table 5.4. Hypermutation analysis. (Pages 118-120)

Gene	Strain	bp sequenced	Mutations	*Igk*AID-*Ung-/-* Mutation frequency (x 10-5)	*Aicda-/-* Mutation frequency (x 10-5)	*P* value
IgSm	*Igk*AID-*Ung-/-*	26101	521	**1996**		**1.00E-08**
	Aicda-/-	26369	0		**<1**	
IgSg1	*Igk*AID-*Ung-/-*	18255	286	**1567**		**1.00E-08**
	Aicda-/-	28270	0		**<1**	
Pim1	*Igk*AID-*Ung-/-*	20453	41	**200.5**		**1.00E-08**
	Aicda-/-	17219	1		**5.8**	
Mir142	*Igk*AID-*Ung-/-*	21551	43	**199.5**		**1.00E-08**
	Aicda-/-	23676	1		**4.22**	
Myc	*Igk*AID-*Ung-/-*	103818	171	**164.7**		**2.17E-06**
	Aicda-/-	19136	3		**15.7**	
Jh4 intron	*Igk*AID-*Ung-/-*	17223	26	**151.0**		**1.00E-08**
	Aicda-/-	14640	2		**13.7**	
Il4ra	*Igk*AID-*Ung-/-*	29062	35	**120**		**5.36E-08**
	Aicda-/-	25257	2		**7.92**	
Nfkb1	*Igk*AID-*Ung-/-*	22734	26	**114.37**		**1.28E-05**
	Aicda-/-	12838	0		**0**	
Mir21	*Igk*AID-*Ung-/-*	25583	27	**105.53**		**2.60E-08**
	Aicda-/-	23454	0		**0**	
Grap	*Igk*AID-*Ung-/-*	19084	19	**99.6**		**1.55E-04**
	Aicda-/-	12402	0		**<1**	
Mir155	*Igk*AID-*Ung-/-*	27677	26	**93.94**		**1.00E-08**
	Aicda-/-	31614	0		**0**	
Csk	*Igk*AID-*Ung-/-*	37807	31	**82.00**		**3.92E-05**
	Aicda-/-	33267	5		**15.03**	
Pax5	*Igk*AID-*Ung-/-*	114476	81	**70.70**		**9.00E-04**
	Aicda-/-	28035	4		**14.27**	
Hist1h1c	*Igk*AID-*Ung-/-*	42680	34	**79.7**		**2.52E-06**
	Aicda-/-	34266	3		**8.8**	
Ly6e	*Igk*AID-*Ung-/-*	20583	13	**63.2**		**2.61E-03**
	Aicda-/-	13546	0		**<1**	
Cd83	*Igk*AID-*Ung-/-*	11577	7	**60.5**		**3.59E-05**
	Aicda-/-	38378	0		**<1**	
H2afx	*Igk*AID-*Ung-/-*	23611	12	**55.1**		**1.81E-03**
	Aicda-/-	18630	0		**<1**	
Gadd45g	*Igk*AID-*Ung-/-*	23728	13	**54.8**		**5.25E-04**
	Aicda-/-	27482	1		**3.6**	
Apobec3	*Igk*AID-*Ung-/-*	37505	18	**47.99**		**3.10E-03**
	Aicda-/-	16440	0		**0**	
Clec2d	*Igk*AID-*Ung-/-*	50996	22	**43.14**		**1.20E-07**

	Aicda-/-	67695	1		**1.47**	
Ikzf1	*lgk*AID-*Ung-/-*	34994	15	**42.86**		**1.05E-02**
	Aicda-/-	36302	4		**11.02**	
Il4i1	*lgk*AID-*Ung-/-*	41757	14	**38.3**		**1.14E-03**
	Aicda-/-	55503	3		**5.4**	
Adrbk1	*lgk*AID-*Ung-/-*	23333	8	**34.3**		**2.45E-02**
	Aicda-/-	17277	0		**<1**	
Hvcn1	*lgk*AID-*Ung-/-*	24912	8	**32.1**		**2.45E-01**
	Aicda-/-	21033	3		**14.3**	
Bcl11a	*lgk*AID-*Ung-/-*	38265	12	**31.36**		**1.50E-02**
	Aicda-/-	34092	2		**5.87**	
Actb	*lgk*AID-*Ung-/-*	25984	7	**26.9**		**1.62E-02**
	Aicda-/-	23858	0		**<1**	
Ucp2	*lgk*AID-*Ung-/-*	22139	5	**22.6**		**1.69E-02**
	Aicda-/-	27937	0		**<1**	
Gpr15	*lgk*AID-*Ung-/-*	37789	7	**18.52**		**3.15E-01**
	Aicda-/-	29540	2		**6.77**	
Pip5k1c	*lgk*AID-*Ung-/-*	36102	6	**16.6**		**3.16E-02**
	Aicda-/-	56600	1		**1.8**	
Dpagt1	*lgk*AID-*Ung-/-*	26430	4	**15.1**		**5.48E-02**
	Aicda-/-	48791	1		**2.1**	
Lman1	*lgk*AID-*Ung-/-*	29569	3	**10.14**		**3.31E-01**
	Aicda-/-	36662	1		**2.72**	
Sfrs9	*lgk*AID-*Ung-/-*	32871	3	**9.12**		**6.69E-01**
	Aicda-/-	37790	2		**5.29**	
Trex1	*lgk*AID-*Ung-/-*	34574	3	**8.68**		**2.59E-01**
	Aicda-/-	27625	0		**<1**	
Blr1	*lgk*AID-*Ung-/-*	36549	3	**8.21**		**5.73E-01**
	Aicda-/-	38924	1		**2.57**	
IgCm	*lgk*AID-*Ung-/-*	45906	4	**8.71**		**3.29E-01**
	Aicda-/-	25726	0		**<1**	
Cdc42	*lgk*AID-*Ung-/-*	28015	2	**7.14**		**5.05E-01**
	Aicda-/-	22838	0		**<1**	
Hmbs	*lgk*AID-*Ung-/-*	28566	2	**7.00**		**2.08E-01**
	Aicda-/-	34058	0		**<1**	
Eif3h	*lgk*AID-*Ung-/-*	33048	2	**6.05**		**5.01E-01**
	Aicda-/-	30948	0		**<1**	
Rpl4	*lgk*AID-*Ung-/-*	34352	2	**5.82**		**5.02E-01**
	Aicda-/-	30600	0		**<1**	
Rbbp7	*lgk*AID-*Ung-/-*	22472	1	**4.44**		**4.58E-01**
	Aicda-/-	26602	0		**<1**	
Crk	*lgk*AID-*Ung-/-*	27511	1	**3.63**		**1.00E+00**
	Aicda-/-	28215	1		**3.54**	
Spib	*lgk*AID-*Ung-/-*	28941	1	**3.45**		**1.00E+00**
	Aicda-/-	27684	0		**0**	
Fam50a	*lgk*AID-*Ung-/-*	34086	1	**2.93**		**6.14E-01**
	Aicda-/-	32154	2		**6.22**	
Aicda	*lgk*AID-*Ung-/-*	35479	1	**2.82**		**1.00E+00**
	Aicda-/-	25717	0		**0**	
Gdi2	*lgk*AID-*Ung-/-*	31044	0	**<1**		**2.48E-01**
	Aicda-/-	30772	2		**6.5**	

N-myc	*Igk*AID-*Ung*-/- *Aicda*-/-	45462 31251	0 0	**<1**	**<1**	**1.00E+00**
Repeat 1*	*Igk*AID-*Ung*-/- *Aicda*-/-	34173 na	3 _	**8.73**	_	_ _
Repeat 2	*Igk*AID-*Ung*-/- *Aicda*-/-	39636 na	0 _	**0.00**	_	_
Repeat 3*	*Igk*AID-*Ung*-/- *Aicda*-/-	48809 na	1 _	**2.05**	_	_
Serinc3	*Igk*AID-*Ung*-/- *Aicda*-/-	48237 na	2 _	**4.15**	_	_
Mir715	*Igk*AID-*Ung*-/- *Aicda*-/-	46081 44285	11 14	**23.87**	**31.61**	**5.51E-01**

*chr19:28,942,810-28,943,877

**chr11:27,022,268-27,023,001

***chr2:98,506,321-98,507,533

Table 5.4. Hypermutation analysis. Genes were analyzed using DNA isolated from *Igk*AID/*Ung*$^{-/-}$ or *Aicda*$^{-/-}$ LPS + IL-4activated B cells (Robbiani et al., 2009). Genes marked in yellow were analyzed de novo and added to a previous hypermutation survey (Yamane et al., 2011). bp: total number of bp sequenced. Mutations: total number of mutations detected. p value was calculated between the mutation frequency obtained in *Igk*AID/*Ung*$^{-/-}$ and *Aicda*$^{-/-}$.

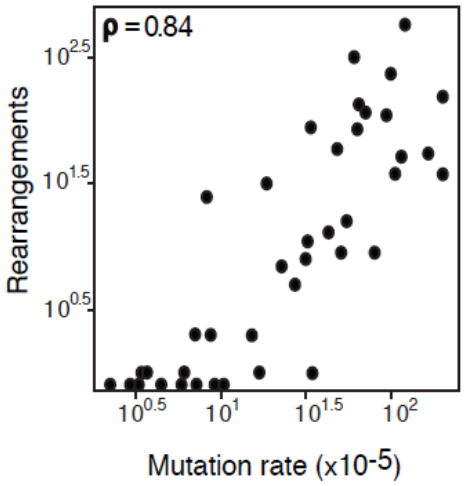


Figure 5.8. Translocation frequency versus mutation frequency in AID targeted genes. Each gene was plotted as a function of its mutation rate in *IgK*AID/*Ung*$^{-/-}$ B cells and its rearrangement frequency as determined by TC-Seq.

Among AID-dependent hotspot containing genes we find several that are translocated or deleted in mature B cell lymphoma. These include *Pax5/IgH*, *Pim1/Bcl6*, *Il21r/Bcl6*, *Gas5/Bcl6* and *Ddx6/IgH* translocations and *Junb* and *Socs1* deletions in diffuse large B cell lymphoma, *Birc3/Malt1* translocation in MALT lymphoma, *Ccnd2/IgK* translocation and *Bcl2l11* deletion in mantle cell lymphoma, *Aff3/Bcl2* and *Grhpr/Bcl6* translocations in follicular lymphoma, *mir142/c-myc* translocation in B cell prolymphocytic leukemia as well as *c-myc/IgH* and *Pvt1/IgK* translocations in Burkitt's lymphoma (Table 5.5).

Table 5.5. AID-dependent rearrangements in human B cell lymphoma.

TC-Seq Gene	Translocation in Mature BCL	Mature BCL Type	Reference
Birc3 (Api2)	t(11;18)(q21;q21)	MALT	(Rosebeck et al., 2011)
Il21r	t(3;16)(q27;p11)	DLBCL	(Ueda et al., 2002)
Pax5	t(9;14)(p13;q32)	DBLCL	(Iida et al., 1999)
Pim1	t(3;6)(q27;p21.2)	DLBCL	(Yoshida et al., 1999)
Aff3	t(2;18)(q11.2;q21)	FL	(Impera et al., 2008)
Gas5	t(1;3)(q25;q27)	DLBCL	(Nakamura et al., 2008)
Ccnd2	t(2;12)(p12;p13)	MCL	(Gesk et al., 2006)
c-myc	t(8;14)(q23;q32)	BL	(Kuppers, 2005)
Ddx6 (Rck)	t(11;14)(q23;q32)	DLBCL	(Lu and Yunis, 1992)
Grhpr	t(3;9)(q27;p11)	FL	(Akasaka et al., 2003)
Bcl2l11	Deleted	MCL	(Bea et al., 2009)
Socs1	Deleted	DLBCL	(Mottok et al., 2009)
Junb	Deleted	DLBCL	(Mao et al., 2002)
mir142	t(8;17)	B-PLL	(Gauwerky et al., 1989)
Pvt1	t(2;8)(p11.2;q24.1)	BL	(Einerson et al., 2006)
IgH	several	several	(Kuppers, 2005)
IgK	several	several	(Kuppers, 2005)
IgL	several	several	(Kuppers, 2005)

Table 5.5. AID-dependent rearrangements in human B cell lymphoma. Genes bearing AID-dependent TC-Seq hotspots are shown with the associated translocation or deletion observed in human mature B cell lymphoma (BCL). BCL types are abbreviated as follows: MALT (mucosa-associated lymphoid tissue lymphoma), DLBCL (diffuse large B cell lymphoma), FL (follicular lymphoma), MCL (mantle cell lymphoma), BL (Burkitt's lymphoma), B-PLL (B cell prolymphocytic leukemia).

Interestingly, we find that AID is capable of inducing DSBs in *Fli1* (Table 5.3), which is translocated to *EWS* in 90% of Ewing's sarcomas, a malignant tumor of uncertain origin (Riggi and Stamenkovic, 2007). We conclude that in addition to mutating many genes, AID also initiates DSBs in numerous non-*Ig* genes. These genes serve as substrates for translocations associated with mature B-cell lymphoma, strongly implicating AID as a source of genomic instability in these cancers.

CHAPTER 6:

Discussion

To date, the study of chromosomal aberrations has been primarily limited to events identified in tumors and tumor cell lines. Although we have learned a great deal about the importance of genomic rearrangements in cancer, it has not been possible to develop an understanding of the cellular and molecular forces that govern their genesis. To examine genomic rearrangements in primary cells in short term cultures, we developed a technique to catalog these events by deep sequencing, TC-Seq. Using this technique we have obtained the complete recombinomes of two genes, *c-myc* and *IgH*, in activated B cells in the presence and absence of AID. Our analysis of over 180,000 rearrangements reveal the importance of transcription, chromosome territories and physical proximity in recombinogenesis, and identifies hotspots for AID-mediated translocations in mature B cells. Our data also suggest that non-Ig DSBs induced by AID are an important source of genomic instability in the mature B cell lymphomas. Finally, TC-Seq is a widely applicable technique; it can be adapted to study translocation biology is any tissue or cell type and to discover the location and nature of other DNA damage sources.

Nuclear Proximity and Chromosomal Position

The existence of chromosome territories, regions in which individual chromosomes segregate, has been long proposed (Cremer and Cremer, 2001) and recently shown to be a key feature of genome organization and determinant locus proximity (Lieberman-Aiden et al., 2009). However, until now the degree to which this impacts the genomic rearrangement landscape was unclear. Our analysis provides evidence that physical proximity and chromosome territories are partial determinants for joining of rearrangement partners. First, we find that rearrangements tend to occur intra-chromosomally. For example, when a DSB is introduced in *c-myc* 40% of rearrangements to that site occur on chromosome 15. This observation is consistent with the analysis of rearrangements in breast cancer genomes, which showed that 92% of events occur intra-chromosomally and involve joining of DSBs within 2 Mb of each other (Stephens et al., 2009). Furthermore, since our experiments were performed in primary cells we can conclude that this phenomenon is a basic feature of rearrangement, not a result of cancer cell specific effects or selection.

We find that the effects of physical proximity are most evident in the 350 kb region around a DSB. In both the absence and presence of AID, the plurality of rearrangements fall in this region. Interestingly, this marked preference for DSB repair within 350 kb matches the range of γH2AX spreading from a DSB

(Bothmer et al., 2010; Savic et al., 2009). γH2AX is a marker of the DNA damage response, which has been shown to spread widely around the site of a DSB, perhaps marking the extent of the DNA damage focus. It has been hypothesized that that the DNA damage response facilitates proximal rearrangement, a phenomenon most prominent at the *IgH* locus during CSR. (Bothmer et al., 2011). Our observed preference for local repair might reflect this fact.

The magnitude of the effect of chromosome territories on rearrangement is far less prominent than proximal joining, but is consistent with recent genome mapping data obtained by high-throughput chromosome conformation capture (Hi-C) (Lieberman-Aiden et al., 2009). When compared to trans-chromosomal joining, we find that a bias to intra-chromosomal rearrangements is evident even when as much as 50 Mb separates DSBs. In mouse, the mean autosome size is ~130 Mb, so a 50 Mb preference for intra-chromosomal joining on either side of a DSB will encompass nearly the entire average chromosome. An intra-chromosomal joining bias is evident in the preferential joining of AID hotspots and non-hotspots on chromosome 12 and chromosome 15 with their respective I-SceI breaks. When a break occurs in *IgH* the AID target *inf2* ~1 Mb from the break appears as a robust hotspot, but when a break occurs in *c-myc*, *inf2* is not captured. The inverse occurs for *pvt1* exon 5, an AID target on chromosome 15 that is captured by *c-myc* but not by *IgH*. We conclude that intra-chromosomal joining is preferred to trans-chromosomal joining.

Hi-C data shows that intra-chromosomal contact probability is inversely proportional to distance and obeys a power law scaling with a slope of –1.08 (Lieberman-Aiden et al., 2009). We find that this closely matches the behavior of a DSB. Rearrangement density decreases as a power function of the distance from a break. The exponent of this power law is -1.29, close to the Hi-C estimate of -1.08. Since rearrangement density diminishes with distance in a manner similar to contact probability, it is mediated by proximity, a likely consequence of local chromosome packing and nuclear chromosomal territories. We conclude that the nuclear organization of chromosome territories impacts the rearrangement landscape and is a determinant of rearrangement propensity. Functionally, a strong preference for proximal intra-chromosomal rearrangement minimizes gross genomic alterations. We propose that this may be an important feature of DSB repair regulation that maintains genomic integrity.

Transcription

Transcription is associated with increased rates of DNA damage and genome instability; these effects are likely mediated by a number of different mechanisms (Gottipati and Helleday, 2009). Transcription may expose ssDNA, which is susceptible to chemical or oxidative damage (Aguilera, 2002). Additionally, head-on collision of the replication and transcription machinery has been implicated in fork stalling and genomic instability (Takeuchi et al., 2003). While these phenomena have been well documented, it remains unclear which is the primary contributor to transcription associated instability and if this is a major mechanism of instability genome-wide.

TC-Seq data suggests that transcription associated instability plays a key role in genome instability. First, in the case of the *c-myc* locus, transcription increases the size of the local area around a DSB that is available for recombination from 50 kb to 300 kb. Rearrangements appear throughout the transcribed *pvt1* gene; this suggests that transcription is indeed recombinogenic. Perhaps more striking, I-SceI breaks rearrange predominantly to transcribed genes genome-wide. We find that in the absence of AID, ~50% of rearrangements occur to genes; whereas only 40% of the genome is genic, this represents a significant enrichment. Rearrangements also tend to occur in transcribed genes, and appear to increase proportionally with transcription levels; significantly more

rearrangements than expected occur to highly transcribed genes while significantly less than expected occur to silent genes. This data is further supported by the finding that activating histone mark association and PolII loading tends to occur on rearranged genes, but not on genes lacking rearrangements. A bias for rearrangement between genic regions was also reported in recent studies on tumors; in sequenced breast cancer genomes 50% of rearrangements occur to genes (Stephens et al., 2009). However, in these studies the role of transcription, transformation or selection in these events could not be evaluated. Our experiments demonstrate that transcribed genic regions are over-represented in chromosomal rearrangements in primary cells in short-term cultures. Thus, this likely represents a general principle underlying recombinogenesis genome-wide.

There are several non-mutually exclusive explanations for our results. They may be a consequence of increased instability at transcribed genes, increased recombinogenicity of transcribed genes or simple physical proximity. Composite density analysis of the genic rearrangement profile implicates ssDNA instability as a major contributor to transcription associated recombination. The plurality of genic rearrangements occur to the transcription start site (TSS). The TSS is a site of substantial $KMnO_4$ sensitivity along the gene body, and thus the primary site of ssDNA through the gene (Nechaev et al., 2010; Zhang and Gralla, 1989). And, as discussed in the introduction, ssDNA has documented sensitivity to damage

as compared to dsDNA (Frederico et al., 1990; Zhang and Gralla, 1989). Thus, we suggest that exposed ssDNA serves as a major source of genomic instability that predisposes transcribed genes to damage and rearrangement.

In addition to being more susceptible to damage, this effect may be due to proximity. *c-myc* is transcribed in activated B cells, and, Hi-C experiments show that transcribed regions are spatially associated in the nucleus (Lieberman-Aiden et al., 2009). Therefore, the proximity of *c-myc* to other transcribed genes might contribute to a bias for genic rearrangements. Additionally, we find that *c-myc* and *IgH* translocate to both AID-dependent and -independent extra-chromosomal targets at similar rates. FISH experiments in cytogenetically normal lymphoma cell lines demonstrate that *IgH* and *c-myc* are in relatively close proximity (Roix et al., 2003). Thus, similar rearrangement profiles of *c-myc* and *IgH* might be the result of a shared nuclear neighborhood, the inhabitants of which are determined by transcriptional activity.

Whatever the origin, we speculate that this phenomenon may have consequences for tumorigenesis. The rearrangement of proto-oncogenes to transcribed regions may lead to their dysregulation by placing them next to constitutively active promoters. Alternatively, rearrangements of coding regions to one another might produce fusion proteins that alter cellular metabolism, as is the case in CML.

AID and Chromosome Translocation

AID initiates SHM and CSR by deaminating cytosine residues in ssDNA exposed by transcription (Chaudhuri and Alt, 2004; Di Noia and Neuberger, 2007; Nussenzweig and Nussenzweig, 2010; Peled et al., 2008; Stavnezer et al., 2008). A primary site of AID activity is the *IgH* locus, where it generates DSBs (Petersen et al., 2001). In addition to participating in the physiological recombination reactions, these breaks may also partner in the generation of chromosomal rearrangements. Indeed, in many of the mature B cell lymphomas *IgH* is translocated to proto-oncogenes thereby dysregulating their expression and driving tumorigenesis (Kuppers, 2005). Until recently, the source of breaks in non-*Ig* genes translocated in mature B cell lymphomas was largely unknown. In 2008, Robbiani and colleagues demonstrated that AID generates DSBs in c-myc, which is translocated in Burkitt's lymphoma (Robbiani et al., 2008). They showed one year later that AID deregulation can affect tumor growth in a mouse model of B cell lymphomagenesis, in some cases targeting *mir-142* for translocation (Robbiani et al., 2009). Thus, the retinue of non-Ig genes conclusively shown to suffer AID mediated DSB was 2, *c-myc* and *mir-142.*

Other studies have suggested additional DSB targets of AID. Liu and colleagues sequenced 118 genes from germinal center B cells and showed that AID was capable of mutating 25% of them (Liu et al., 2008). This study suggested that AID

may act promiscuously, but it examined a very limited number of loci and only analyzed mutations, not DSBs. Subsequent genome-wide studies examined the locus association of certain factors by ChIP-Seq, elucidating AID's targeting mechanism and extending the catalog of potential AID targets. Pavri and colleagues showed that the *IgH* locus and the TSSs of stalled genes load Spt5, a PolII stalling factor (Pavri et al., 2010). AID interacts with Spt5, which associates with thousands of loci, specifically at the TSS of stalled genes. AID ChIP-Seq confirmed the presence of AID on genes that load Spt5 (Yamane et al., 2011). Both the Spt5 and AID ChIP-Seq studies included mutational analysis of a few loci in a fixed position of the gene body, but lacked a direct measurement of AID activity genome-wide (Pavri et al., 2010; Yamane et al., 2011). Without such a measure, these studies remain suggestive. Moreover, to date no method exists to perform an unbiased examination of AID-mediated DSBs or translocations. Therefore, the precise relationships between AID and Spt5 occupancy, mutation, and translocations have not been investigated.

Here, we provide the first genome-wide direct measurement of AID activity and the first unbiased analysis of AID-mediated DSBs. First, we show that AID targets discrete sites for DSBs. These sites are predominantly genic; 90% of the AID-dependent rearrangement partners of *IgH* are in genes. We can detect rearrangements to the transcription start sites of nearly 100 different genes that are actively transcribed. This is consistent with many studies showing that AID is

associated with transcription (Jung et al., 1993; Stavnezer-Nordgren and Sirlin, 1986; Yoshikawa et al., 2002), and with the positions of sequenced *c-myc/IgH* translocations in primary B cells (Robbiani et al., 2008).

The TSS is a site enriched in ssDNA (Nechaev et al., 2010; Zhang and Gralla, 1989); this appears to be the primary site of AID activity genome-wide, which is consistent with studies demonstrating AID activity on ssDNA *in vitro* and *in vivo* (Bransteitter et al., 2003; Chaudhuri et al., 2003; Dickerson et al., 2003; Petersen-Mahrt et al., 2002; Pham et al., 2003; Ramiro et al., 2003; Sohail et al., 2003). While genes rearranged by AID are transcribed, expression (as measured by mRNA-Seq) does not correlate directly with rearrangement frequency. This suggests that transcription is necessary but not rate-limiting for rearrangement. Rather, PolII stalling is the likely mechanism for AID targeting. This is reflected in the distribution of AID and its co-factor Spt5 in the genome (Pavri et al., 2010; Yamane et al., 2011), and that AID-dependent rearrangements occur mainly on transcription start sites of stalled genes that carry high levels of the PolII stalling factor Spt5.

Interestingly, The AID-dependent rearrangement partners of *c-myc* and *IgH* appear largely similar. We observe a bias for intra-chromosomal joining of AID targets to I-SceI sites. But, for extra-chromosomal AID targets, we observe no such effect. This might be due to a shared nuclear compartment of *c-myc*, *IgH*

and most AID-targeted genes (which are transcribed). Their mutual physical proximity may be facilitating similar trans-chromosomal rearrangement profiles. Given the demonstrated physical proximity of *IgH* and *c-myc* (Roix et al., 2003) and the tendency of transcribed regions to interact (Lieberman-Aiden et al., 2009), transcription-associated genomic organization may mediate this phenomenon.

A recent study using Nbs1 ChIP-on-ChIP as a surrogate for DNA damage suggested that AID targets repeat rich sequences (Staszewski et al., 2011). These were proposed to occur specifically at CA repeats, non-CA tandem repeats and SINEs. In contrast, we find no AID-dependent increase in rearrangements to these, or any other, non-*Ig* repeats. This is consistent with AID and Spt5 ChIP-Seq studies, which failed to note any loading at these sites. Moreover, AID-dependent rearrangement hotspots predominantly occur in genes, not in or near repeat regions that are not transcribed.

In addition, we find a strong correlation between hypermutation and rearrangements, suggesting that genes susceptible to AID mediated DSBs are a subset of the most highly mutated genes in the genome. Consistent with this notion, we show that *Pax5, Il21r, Gas5, Ddx6*, *Birc3*, *Ccnd2*, *Aff3, Grhpr, c-myc, Pvt1*, *Bcl2l11, Socs1, mir142, Junb* and *Pim1*, which are translocated or deleted in mature B cell lymphomas, and are among the more highly mutated AID

targets, bear AID-dependent translocation hotspots. This suggests that in addition to hypermutation, AID is a major source of genomic instability in these cancers. Many of the mature B cell lymphomas are thought to originate in the germinal center, which is the primary site of AID expression and activity (Kuppers, 2005; Muramatsu et al., 1999). And, many of the AID-initiated translocations we describe are recurrent and clonal in these tumors. In line with data showing that AID is oncogenic *in vivo* (Robbiani et al., 2009), we suggest that AID can initiate genomic rearrangements essential for mature B cell lymphomagenesis. It should be noted that our experiments were performed on *in vitro* stimulated B cells. Germinal center B cells may have an alternate gene expression profile that influences the number and position of AID target sites. This may account for the absence of some genes from our hotspot lists known to be translocated in mature B cell lymphoma. For example, we do not detect the *Bcl6/IgH* translocation observed in diffuse large B cell lymphoma (Ye et al., 1995). As such, to further extend these studies and capture the entire complement of IgH translocation partners, we propose that TC-Seq might be adapted for *in vivo* use in germinal centers.

Future Directions

Finally, we propose that TC-Seq can be adapted for use in other cell types and systems to study translocation biology, nuclear architecture and the mechanics of other DNA damaging enzyme, chemicals and conditions.

Translocation profiles of other cell types. Recurrent translocations appear in tumors other than mature B cell lymphomas. For example, most mucoepidermoid carcinomas harbor the *MECT1-MAML2* gene rearrangement and in many prostate cancers *TMPRSS2* is found fused to the *ERG* gene (Clark and Cooper, 2009; Hunt, 2011). In this study we have examined the rearrangement profile of *IgH*, and thereby have revealed the mechanisms of its rearrangement. We propose that examining the recombination profile of DSBs in rearrangement-prone genes of other cell types might reveal the forces underlying their behavior. TC-Seq can be performed in any cell culture, so these experiments would simply require that the cell type of interest be isolated and grown.

Examining junctional microanatomy. TC-Seq relies on paired-end Illumina sequencing to identify the I-SceI containing locus and a sequence displaced from the partner sequence breakpoint; 36 nucleotides are read from each end of each molecule. This sequencing modality has the advantage of extremely high throughput, using latest-generation chemistry each run can yield greater than 240 million reads. Such a quantity of data allows us to assay tens of million of cells

concurrently while ensuring complete saturation, thereby facilitating highly powered studies of rare events. However, Illumina paired-end sequencing does not allow for sequencing through breakpoints. High throughput breakpoint analysis would yield information about junctional microanatomy (deletion, microhomology utilization and insertion) thereby allowing us to distinguish between NHEJ and alt-NHEJ mediated events. First, one could explore the relative contributions of these two pathways to the unselected rearrangement landscape in primary cells. Second, these experiments could be performed in the absence of DNA repair factors to investigate their contribution to aberrant joining in these two pathways.

To obtain such data one needs at least 300 bp continuous sequencing reads, through the entire fragmented DNA molecule. An alternative sequencing technology, 454 sequencing, performs 400 bp continuous reads that would allow such an analysis. Unfortunately, this technology is merely medium-throughput (1 million reads per run), so much less power can be achieved in these analyses. We have successfully adapted TC-Seq for use with a 454 sequencer and are exploring the genome-wide landscape of junctional microanatomy.

Mapping common fragile sites. Common fragile sites are areas of the genome damaged during partial replicative stress. While about 80 of these sites are currently known, they are generally coarsely mapped. While a sequence feature

is likely to underlie this phenomenon the exact cause of their emergence is incompletely defined (Durkin and Glover, 2007). These sites are documented to partner in rearrangements (Burrow et al., 2009). We propose that TC-Seq can be employed to map rearrangements that arise after treatment with aphidicolin, a drug that induces DSBs at fragile sites (Glover et al., 1984). In this way, the exact nature and location of fragile sites can be determined.

A silent gene recombinome. TC-Seq data has revealed that *c-myc* and *IgH*, both transcribed genes, tend to rearrange to other transcribed genes. Since the composite density analysis of genic rearrangements reveals a bias to the transcription start site, ssDNA fragility may be a partial determinant of this bias. However, what role genomic architecture plays in this trend is unknown. We propose that obtaining the recombinome for a DSB in a silent gene could reveal the relative roles of ssDNA instability and spatial proximity in this phenomenon. We expect that if differential localization of transcribed and non-transcribed genes impacts rearrangement profiles, a DSB in a silent gene should recombine less frequently than a DSB in an active gene to other transcribed genes. Such an experiment could be performed by generating a knockout mouse with an I-SceI site in a gene silent in B cells, or by performing TC-Seq on *IgH* in a cell type in which it is unexpressed.

Exploring RAG-induced non-Ig DSBs. RAGs may catalyze non-*Ig* DSBs. These DSBs may participate in rearrangement observed in B cell lymphomas (Vaandrager et al., 2000). However, the spectrum of RAG targeting is unknown. TC-Seq may be adapted to discover these targets. We propose that pre-B cells bearing I-SceI sites may be isolated from mouse bone marrow and infected with Ablson virus, thereby immortalizing and freezing them in an early large pre-B cell stage in which RAGs are not expressed (Muljo and Schlissel, 2003). Administration with STI-571 (Gleevec) will allow cells to differentiate into the late pre-B cell stage and enforce RAG expression (Muljo and Schlissel, 2003). At this time cells can be infected with I-SceI to generate RAG-mediated rearrangements to that site.

RAG function differs substantially from AID. While AID has no known end protection function, RAG2 shepherds DSBs to faithful joining by the NHEJ pathway (Corneo et al., 2007). Thereby, RAG-generated ends may be less amenable to rearrangement generation than AID generated ends. coreRAG2 released DNA ends may be more available to aberrant joining (Corneo et al., 2007; Deriano et al., 2011); we propose that using this RAG2 mutant in TC-Seq might facilitate non-*Ig* RAG target discovery.

CHAPTER 7:

Methods

B Cell Cultures, Infections and Sorting

Resting B lymphocytes were isolated from mouse spleens by immunomagnetic depletion with anti-CD43 MicroBeads (Miltenyi Biotech) and cultured at 0.5×10^6 cells/ml in RPMI supplemented with L-glutamine, sodium pyruvate, antibiotic/antimycotic, HEPES, 50 μM 2-mercaptoethanol (all from GIBCO-BRL), and 10% fetal calf serum (Hyclone). B cells were stimulated in the presence of 500ng/ml RP105 (BD Pharmingen), 25 μg/ml lipopolysaccharide (LPS) (Sigma) and 5 ng/ml mouse recombinant IL-4 (Sigma). Retroviral supernatants were prepared by cotransfection of BOSC23 cells with pCL-Eco and pMX-IRES-GFP-derived plasmids encoding for I-SceI-mCherry or AID-GFP with Fugene 6, 72 hr before infection. At 20 and 44 hr of lymphocyte culture, retroviral supernatants were added, and B cells were spinoculated at 1150 g for 1.5 hr in the presence of 10 μg/ml polybrene. For dual infection, separately prepared retroviral supernatants were added simultaneously on both days. After 4 hr at 37°C, supernatants were replaced with LPS and IL-4 in supplemented RPMI. At 96 hr from the beginning of their culture, singly infected B cells were collected and

frozen in 10 million cell pellets at -80C. Dually infected B cells were sorted for double positive cells with a FACSAria instrument (Becton Dickson) then frozen down.

Translocation Capture Sequencing (TC-Seq)

Genomic DNA Extraction and Sonication

5x10 million B cell aliquots were thawed on ice. Cells were resuspended in 100ul of Phosphate buffered saline (PBS) and transferred to a 15ml conical tube containing 5mL Proteinase K buffer (100mM Tris pH8, 0.2% SDS, 200mM NaCl, 5mM EDTA) and 50ul of 20mg/ml Proteinase K. Cells were digested overnight at 55C. To extract DNA, addition of 5mL of phenol-chloroform was followed by 30s of gentle inversion and 20 minutes of centrifugation at 4000rpm. The aqueous phase was transferred to a 50ml conical tube containing 12.5mL of 100% ethanol. The mixture was centrifuged for 15 minutes at 4000rpm at 4C. The pellet was washed twice with 5mL 70% ethanol, transferred to a microcentrifuge tube, air dried for 5 minutes and resuspended in 100uL 10mM Tris pH8 with 1uL of 0.5mg/ml DNAse-free RNAse (Roche). Genomic DNA concentration was adjusted to 167ng/uL with 10mM Tris pH8 and divided into microcentrifuge tubes containing 300uL each. Genomic DNA was fragmented by sonication at low power for 7 cycles (30” on/30” off) in a Bioruptor (Diagenode) to yield a 500-1350bp distribution of DNA fragments. DNA was further divided into 50x30ul (5ug) aliquots in 1.5mL eppendorf tubes.

Genomic DNA Fragment Preparation, Linker Ligation and Elimination of Unrearranged Loci

Each experiment consisted of genomic DNA from 50 million B cells in 50 x 5ug aliquots for a total of 250ug of fragmented genomic DNA. The following reactions were performed individually on 5ug aliquots. DNA was blunted by End-It DNA Repair Kit (Epicentre) then purified by Qiagen PCR purification column and eluted in 43uL of Buffer EB. Blunted DNA was adenosine-tailed by addition of 5ul 10x NEB2 buffer, 1uL 10mM dATP and 2uL Klenow fragment 3->5' exo$^-$ (NEB) followed by incubation at 37C for 1hr. Each reaction was purified by Qiagen PCR purification kit and eluted in 40uL of Buffer EB. Each aliquot of blunted, A-tailed DNA fragments was ligated to 200pmol of linkers by addition of 4uL linker, 5uL NEB T4 DNA ligase buffer and 1uL high concentration T4 DNA ligase (NEB) followed by incubation overnight at 16C. Ligase was inactivated by incubation at 70C for 20 minutes, reactions were purified by Qiagen PCR purification column and eluted in 44uL Buffer EB. Native loci were eliminated by adding 5uL 10x NEB Isce1 buffer, 0.5uL 100x BSA and 1uL I-Sce1 then incubating the reaction at 37C for 2 hours. Each reaction was purified by Qiagen PCR purification column and eluted in 50uL Buffer EB. Finally, all 50 reactions were pooled.

PCR Round 1

Pooled linker-ligated DNA was divided into 2 equal parts for semi-nested ligation-mediated PCR using either forward or reverse primers. Forward and reverse enrichment streams are kept separate for the entire remainder of the protocol. DNA was divided into 1ug aliquots and mixed with 20ul 5x Phusion HF buffer, 3ul 10mM dNTPs, 1ul 20uM biotinylated pMycF1, pMycR1, pIgHF1 or pIgHR1, 1uL of Phusion Taq (Finnzymes) and H2O to 100uL. Single-primer PCR reactions were run – 1x(98C-1min) 12x(98C-15sec, 65C-30sec, 72C-45sec) 1x(72C-1min) 1x(4C-forever) - then each tube was spiked with 1ul 20uM pLinker and subjected to additional cycles of PCR - 1x(98C-1min) 35x(98C-15sec, 65C-30sec, 72C-45sec) 1x(72C-5min) 1x(4C-forever). Forward and reverse PCR reactions were pooled separately, and 400uL of each was purified in 4 Qiagen PCR purification columns, each eluted in 50uL EB. All 200ul of each sample was run on a 2% agarose gel until well separated and appropriately size fragments [for IgH^{I}: F (350bp-1kb) R (300-1kb), for Myc^{I}: F (300bp-1kb) R (300-1kb)] were excised. DNA was purified in Qiagen Gel purification columns and gel-based size selection was repeated once. 100uL of washed magnetic T1 magenetic streptavidin beads (Invitrogen) were resuspended in 400ul 2x B&W buffer (10mM Tris pH7.5, 1mM EDTA, 2M NaCl), 200ul was added to each forward and reverse PCR1 products and the mixture incubated for 1hr with gentle nutation at room temperature. Beads were isolated, washed 3x in 500uL 1x B&W buffer, 1x in

ddH2O and resuspended in 87ul H2O. Unrearranged loci were further eliminated by adding 10ul 10x NEB Isce1 buffer, 1ul 100x BSA and 2uL Isce1, then digesting DNA on beads for 2 hours at 37C. Beads were washed 3 times in 1x B&W, once in H2O and resuspended in 50uL H2O.

Round 2 PCR

20ul of beads from each of the forward and reverse PCR1 was separately mixed with 40ul 5x Phusion HF buffer, 6ul 10mM dNTPs, 2ul 20uM pIgHF2, pIgHR2, pMycF2 or pMycR2, 2uL 20uM pLinker, 2ul Phusion taq and 128ul H2O then separated into 50ul aliquots and subjected to PCR - 1x(98C-1min) 35x(98C-10sec, 65C-30sec, 72C-40sec) 1x(72C-5min) 1x(4C-forever). PCR reactions were separated from beads and each tube was run on one lane of a 2% agarose gel until appropriate separation was achieved. Desired fragment sizes were excised [for IgH^{I}: F (300bp-1kb) R (250-1kb), for Myc^{I}: F (270bp-1kb) R (270-1kb)] and purified by Qiagen gel purification column. Size selection was repeated once more in 4 Qiagen gel purification columns for each forward and reverse; each column was eluted with 30uL EB.

Preparing Enriched DNA for Paired-end High Throughput Sequencing

120ul of isolated DNA was mixed with 13uL 10x NEB Buffer 4, 2uL of Asc1 and

H2O to 150ul then incubated at 37C for 2 hours to cleave the linker off enriched products. Restriction digests were purified by Qiagen PCR purification column and redigested with Asc1 to ensure complete linker removal. Desired fragment sizes [for IgH^{I}: F (250bp-1kb) R (200-1kb). for Myc^{I}: F (220bp-1kb) R (220-1kb)] were isolated by eletrophoresis on a 2% agarose gel and isolated by Qiagen gel purification column. DNA was blunted by End-It DNA Repair Kit (Epicentre), column purified, A-tailed as described above and ligated to Illumina paired end adapters. Ligation reactions were column purified then run on a 2% agarose gel and desired fragment sizes isolated [for IgH^{I}: F (320bp-1kb) R (270-1kb), for Myc^{I}: F (300bp-1kb) R (300-1kb)] (Note, at this point the product may not be visible on the gel). Adapter ligated fragments were enriched by 25 cycles of PCR with Illumina PCR primers PE1.0 and PE2.0 and desired fragment sizes isolated on a 2% agarose gel [for IgH^{I}: F (400bp-1kb) R (350-1kb), for Myc^{I}: F (380bp-1kb) R (380-1kb)]. Samples were confirmed to be enriched for rearranged loci by TOPO cloning and screening 10 forward- and 10 reverse-enrichment colonies. Finally, appropriate size distribution was confirmed by Agilent Bioanalyzer. Forward and reverse libraries for the same sample were mixed in equimolar ratios and sequenced by 36x36 or 54x54 paired-end deep sequencing on an Illumina GAII.

Mapping of Translocation Hotspots

A translocation hotspot was defined as a localized enrichment of translocation events above what is expected from the null hypothesis of uniform distribution of translocation events along the genome. To identify such hotspots, candidate regions were defined as locations containing consecutive translocations with distances shorter than expected from the mappable size of the mm9 genome assembly ($P < 0.01$ each as determined by a negative binomial test). For a candidate region to be called a hotspot it had to (1) have more than 3 translocations and (2) have at least one read from each of the two sides of the bait and (3) have at least 10% of the translocations come from each side of the bait and (4) have a combined P value less than 10^{-9} given the number of translocations and length of the region as determined by a negative binomial test. Hotspots with a large degree (>80%) of overlap with repeat regions or small footprints (<100nt) were removed. For AID-dependent hotspot lists hotspots with less than 10-fold enrichment over the $AID^{-/-}$ control were removed. Analyses of RNA-Seq, chromatin modifications, AID-, PolII-, and Spt5-ChIP as well as the identification of cryptic I-SceI sites, TSSs, genic and intergenic domains were carried out in R (http://www.R-project.org).

Hypermutation Analysis

CD43^{-} splenocytes from *Igk*AID-*Ung*$^{-/-}$ or *Aicda*$^{-/-}$ mice were cultured at 0.1 x 10^6 cells/ml with LPS+IL-4, and 0.5 mg/ml of aCD180 (RP105) antibody (RP/14, BD Pharmingen). At 72 hrs cells were diluted 1:4 and cultured for another 48 hs. 50 ng of genomic DNA was amplified for 30 cycles with Phusion DNA polymerase (New England Biolabs) and specific primers. For nested PCR, two-20 cycle amplifications were performed with DMSO. The amplicon was cloned using PCR Zero blunt (Invitrogen) and sequenced.

Primers

50mM double stranded asymmetric linker (dsaLinker) was generated by mixing equimolar parts of pLT and pLB in annealing buffer (10 mM Tris, pH 8.0, 50 mM NaCl, 1 mM EDTA), heating the mixture to boiling, then incubating it at room temperature for 2 hours.

Table 7.1. Primers for TC-Seq.

Name	5' modification	Sequence	3' modification
pLT		GCAGCGGATAACAATTTCACACAGGACGTACTGTGGCGCGCCT	
pLB	Phosphorylated	GGCGCGCCACAGTACTTGACTGAGCTTTA	dideoxycytosine
pLinker		GCAGCGGATAACAATTTCACACAGGAC	
pMycF1	Biotinylated	CAAAATTGGGACAGGGATGTGACC	
pMycR1	Biotinylated	GGTGTCAAATAATAAGAGACACCTCCC	
pIgHF1	Biotinylated	GCTTATTGTTGAATGGGCCAAAGGTC	
pIgHR1	Biotinylated	CTAAAGCAATGACTGAAGACTCAGTCCC	
pMycF2		CTTGGGGGAAACCAGAGGGAATC	
pMycR2		TACACTCTAAACCGCGACGCC	
pIgHF2		GGTAGGCCTGGACTTTGGGTC	
pIgHR2		ACTGTGGCTGCCTCTGGCTTAC	

Table 7.2. Primers for hypermutation analysis. (Pages 150-151)

Gene	Primer_F	Primer_R	Annealing Temp	Nested_ Primer_F	Nested_ Primer_R	Annealing temp
Nfkb1	CTCTCCG CACCACG GTGCAG	GAGGTCAG GCACAAAA CATG	62c	GAAGC GGAGA GGAACT CTTG	GAGGTC AGGCAC AAAACA TG	64c
mir21	CTTTTAAA TAAAGCAA GGCTTTTA CCACAG	CATGGAGTT AATAATAGA CAGGACAA AG	60c			
mir-155	CAAAGCT GACAGGA AAGTCCA G	CAGCTGAC CCCACTCT GC	64c			
Csk	ACTGTCCT GCAGGCC ACTGAT	CAGATGGC CCACACGA TTAG	60c	CAGAGA GGTGA GTGCTG G	GAGAG GAAGTC TTATAA GCAATG C	60c
Apobec3	TGAACTCC TGACCTTC AGAAG	TGTTCTACC TCCAAGTCC TAC	60c			
Clec2d	CTTCCCTA CCTATGCT TAGTC	TCAAGAGCT CAAAACAGA AAGG	60c	AGGAG GTAGAA GTGGGT AAG	AAAGGA TTTGGT CCTGAG TAG	60c
Ikzf1	CATTCTTG GCCCCCA AAGC	CCCTAACAC CAGTTTGGT TCTG	60c			
bcl11a	ATGTGAA CCGAGCC GTCGT	GGGGTGAG GGTGCGAA CTAAT	60c			
Gpr15	AATCCAG CCAACCA GACAAC	ACTGGACT GCTGTGCT GTTC	60c			
Blr1	CTGGTCC TGCTTGG AACTTTA	ACGTCTGG CCTCATGCT TTC	60c			
Spib	GCTCTGT	TGAGGCCA	60c			

	CTTCCTAG ACAAG	GAAGTAGAT GATC				
aicda	CCCTTGA GTCTGGTT TGCATTG	GAGGCAGG TGGATCATT CTG	60c			
Repeat1 (chr19)	TTGTCCAG ATGAACTT GACTGTG	TGGTTTTCA TTTTCGGTG TCTTCTG	60c			
Repeat2 (chr11)	GTCTGTA GATGTAG CTCTCTTC	GTATTCCAC CAAGATCTG CTTAG	60c			
Repeat3 (chr2)	CTGGCTAT GTCCCAG TTCTTC	CATGTGCA GGAACATG TAACTAAG	62c			
Serinc3	ATACCTTT TAGAGGA GGCTGTA C	GTAACTCAG GTGATTAG GTCTTG	60c			
mir715	CCTAGGT GCCTGCT TCTGAG	AGGGAACG ACACAGCA GAAC	60c			

CHAPTER 8:

References

Aguilera, A. (2002). The connection between transcription and genomic instability. EMBO J *21*, 195-201.

Akasaka, T., Lossos, I.S., and Levy, R. (2003). BCL6 gene translocation in follicular lymphoma: a harbinger of eventual transformation to diffuse aggressive lymphoma. Blood *102*, 1443-1448.

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X.*, et al.* (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature *403*, 503-511.

Allen, C., Ashley, A.K., Hromas, R., and Nickoloff, J.A. (2011). More forks on the road to replication stress recovery. J Mol Cell Biol *3*, 4-12.

Allen, C., Miller, C.A., and Nickoloff, J.A. (2003). The mutagenic potential of a single DNA double-strand break in a mammalian chromosome is not influenced by transcription. DNA Repair (Amst) *2*, 1147-1156.

Ammazzalorso, F., Pirzio, L.M., Bignami, M., Franchitto, A., and Pichierri, P. (2010). ATR and ATM differently regulate WRN to prevent DSBs at stalled replication forks and promote replication fork recovery. EMBO J *29*, 3156-3169.

Bahler, D.W., and Levy, R. (1992). Clonal evolution of a follicular lymphoma: evidence for antigen selection. Proc Natl Acad Sci U S A *89*, 6770-6774.

Bardwell, P.D., Woo, C.J., Wei, K., Li, Z., Martin, A., Sack, S.Z., Parris, T., Edelmann, W., and Scharff, M.D. (2004). Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice. Nat Immunol *5*, 224-229.

Barzilai, A., Rotman, G., and Shiloh, Y. (2002). ATM deficiency and oxidative stress: a new dimension of defective response to DNA damage. DNA Repair (Amst) *1*, 3-25.

Basu, U., Meng, F.L., Keim, C., Grinstein, V., Pefanis, E., Eccleston, J., Zhang, T., Myers, D., Wasserman, C.R., Wesemann, D.R.*, et al.* (2011). The RNA

exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. Cell *144*, 353-363.

Bea, S., Salaverria, I., Armengol, L., Pinyol, M., Fernandez, V., Hartmann, E.M., Jares, P., Amador, V., Hernandez, L., Navarro, A.*, et al.* (2009). Uniparental disomies, homozygous deletions, amplifications, and target genes in mantle cell lymphoma revealed by integrative high-resolution whole-genome profiling. Blood *113*, 3059-3069.

Beletskii, A., Grigoriev, A., Joyce, S., and Bhagwat, A.S. (2000). Mutations induced by bacteriophage T7 RNA polymerase and their effects on the composition of the T7 genome. J Mol Biol *300*, 1057-1065.

Bellan, C., Stefano, L., Giulia de, F., Rogena, E.A., and Lorenzo, L. (2010). Burkitt lymphoma versus diffuse large B-cell lymphoma: a practical approach. Hematol Oncol *28*, 53-56.

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

Betz, A.G., Milstein, C., Gonzalez-Fernandez, A., Pannell, R., Larson, T., and Neuberger, M.S. (1994). Elements regulating somatic hypermutation of an immunoglobulin kappa gene: critical role for the intron enhancer/matrix attachment region. Cell *77*, 239-248.

Blanco, J.G., Dervieux, T., Edick, M.J., Mehta, P.K., Rubnitz, J.E., Shurtleff, S., Raimondi, S.C., Behm, F.G., Pui, C.H., and Relling, M.V. (2001). Molecular emergence of acute myeloid leukemia during treatment for acute lymphoblastic leukemia. Proc Natl Acad Sci U S A *98*, 10338-10343.

Boboila, C., Jankovic, M., Yan, C.T., Wang, J.H., Wesemann, D.R., Zhang, T., Fazeli, A., Feldman, L., Nussenzweig, A., Nussenzweig, M.*, et al.* (2010a). Alternative end-joining catalyzes robust IgH locus deletions and translocations in the combined absence of ligase 4 and Ku70. Proc Natl Acad Sci U S A *107*, 3034-3039.

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

Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., Muller, S., Eils, R., Cremer, C., Speicher, M.R.*, et al.* (2005). Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. PLoS Biol *3*, e157.

Bothmer, A., Robbiani, D.F., Di Virgilio, M., Bunting, S.F., Klein, I.A., Feldhahn, N., Barlow, J., Chen, H.T., Bosque, D., Callen, E.*, et al.* (2011). Regulation of DNA end joining, resection, and immunoglobulin class switch recombination by 53BP1. Mol Cell *42*, 319-329.

Bothmer, A., Robbiani, D.F., Feldhahn, N., Gazumyan, A., Nussenzweig, A., and Nussenzweig, M.C. (2010). 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. J Exp Med *207*, 855-865.

Brandt, V.L., and Roth, D.B. (2004). V(D)J recombination: how to tame a transposase. Immunol Rev *200*, 249-260.

Bransteitter, R., Pham, P., Scharff, M.D., and Goodman, M.F. (2003). Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. Proc Natl Acad Sci U S A *100*, 4102-4107.

Branzei, D., and Foiani, M. (2010). Maintaining genome stability at the replication fork. Nat Rev Mol Cell Biol *11*, 208-219.

Brock, R.D. (1971). Differential mutation of the beta-galactosidase gene of Escherichia coli. Mutat Res *11*, 181-186.

Brunet, E., Simsek, D., Tomishima, M., DeKelver, R., Choi, V.M., Gregory, P., Urnov, F., Weinstock, D.M., and Jasin, M. (2009). Chromosomal translocations induced at specified loci in human stem cells. Proc Natl Acad Sci U S A *106*, 10620-10625.

Burrow, A.A., Williams, L.E., Pierce, L.C., and Wang, Y.H. (2009). Over half of breakpoints in gene pairs involved in cancer-specific recurrent translocations are mapped to human chromosomal fragile sites. BMC Genomics *10*, 59.

Callen, E., Jankovic, M., Wong, N., Zha, S., Chen, H.T., Difilippantonio, S., Di Virgilio, M., Heidkamp, G., Alt, F.W., Nussenzweig, A.*, et al.* (2009). Essential role for DNA-PKcs in DNA double-strand break repair and apoptosis in ATM-deficient lymphocytes. Mol Cell *34*, 285-297.

Camacho, E., Hernandez, L., Hernandez, S., Tort, F., Bellosillo, B., Bea, S., Bosch, F., Montserrat, E., Cardesa, A., Fernandez, P.L.*, et al.* (2002). ATM gene inactivation in mantle cell lymphoma mainly occurs by truncating mutations and missense mutations involving the phosphatidylinositol-3 kinase domain and is associated with increasing numbers of chromosomal imbalances. Blood *99*, 238-244.

Campbell, P.J., Stephens, P.J., Pleasance, E.D., O'Meara, S., Li, H., Santarius, T., Stebbings, L.A., Leroy, C., Edkins, S., Hardy, C.*, et al.* (2008). Identification of

somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nat Genet *40*, 722-729.

Campbell, P.J., Yachida, S., Mudie, L.J., Stephens, P.J., Pleasance, E.D., Stebbings, L.A., Morsberger, L.A., Latimer, C., McLaren, S., Lin, M.L.*, et al.* (2010). The patterns and dynamics of genomic instability in metastatic pancreatic cancer. Nature *467*, 1109-1113.

Carbone, A., Gloghini, A., Cabras, A., and Elia, G. (2009). The Germinal centre-derived lymphomas seen through their cellular microenvironment. Br J Haematol *145*, 468-480.

Caudill, C.M., Zhu, Z., Ciampi, R., Stringer, J.R., and Nikiforov, Y.E. (2005). Dose-dependent generation of RET/PTC in human thyroid cells after in vitro exposure to gamma-radiation: a model of carcinogenic chromosomal rearrangement induced by ionizing radiation. J Clin Endocrinol Metab *90*, 2364-2369.

Chaudhuri, J., and Alt, F.W. (2004). Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. Nat Rev Immunol *4*, 541-552.

Chaudhuri, J., Khuong, C., and Alt, F.W. (2004). Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. Nature *430*, 992-998.

Chaudhuri, J., Tian, M., Khuong, C., Chua, K., Pinaud, E., and Alt, F.W. (2003). Transcription-targeted DNA deamination by the AID antibody diversification enzyme. Nature *422*, 726-730.

Clark, J.P., and Cooper, C.S. (2009). ETS gene fusions in prostate cancer. Nat Rev Urol *6*, 429-439.

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

Cory, S., Graham, M., Webb, E., Corcoran, L., and Adams, J.M. (1985). Variant (6;15) translocations in murine plasmacytomas involve a chromosome 15 locus at least 72 kb from the c-myc oncogene. EMBO J *4*, 675-681.

Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet *2*, 292-301.

Cremer, T., and Cremer, M. (2010). Chromosome territories. Cold Spring Harb Perspect Biol *2*, a003889.

Dalla-Favera, R., Martinotti, S., Gallo, R.C., Erikson, J., and Croce, C.M. (1983). Translocation and rearrangements of the c-myc oncogene locus in human undifferentiated B-cell lymphomas. Science *219*, 963-967.

Derheimer, F.A., and Kastan, M.B. (2010). Multiple roles of ATM in monitoring and maintaining DNA integrity. FEBS Lett *584*, 3675-3681.

Deriano, L., Chaumeil, J., Coussens, M., Multani, A., Chou, Y., Alekseyenko, A.V., Chang, S., Skok, J.A., and Roth, D.B. (2011). The RAG2 C terminus suppresses genomic instability and lymphomagenesis. Nature *471*, 119-123.

Deriano, L., Stracker, T.H., Baker, A., Petrini, J.H., and Roth, D.B. (2009). Roles for NBS1 in alternative nonhomologous end-joining of V(D)J recombination intermediates. Mol Cell *34*, 13-25.

Di Noia, J., and Neuberger, M.S. (2002). Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. Nature *419*, 43-48.

Di Noia, J.M., and Neuberger, M.S. (2007). Molecular mechanisms of antibody somatic hypermutation. Annu Rev Biochem *76*, 1-22.

Dickerson, S.K., Market, E., Besmer, E., and Papavasiliou, F.N. (2003). AID mediates hypermutation by deaminating single stranded DNA. J Exp Med *197*, 1291-1296.

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

Difilippantonio, S., Gapud, E., Wong, N., Huang, C.Y., Mahowald, G., Chen, H.T., Kruhlak, M.J., Callen, E., Livak, F., Nussenzweig, M.C.*, et al.* (2008). 53BP1 facilitates long-range DNA end-joining during V(D)J recombination. Nature *456*, 529-533.

Dillon, L.W., Burrow, A.A., and Wang, Y.H. (2010). DNA instability at chromosomal fragile sites in cancer. Curr Genomics *11*, 326-337.

Dimitrova, N., Chen, Y.C., Spector, D.L., and de Lange, T. (2008). 53BP1 promotes non-homologous end joining of telomeres by increasing chromatin mobility. Nature *456*, 524-528.

Druker, B.J., Talpaz, M., Resta, D.J., Peng, B., Buchdunger, E., Ford, J.M., Lydon, N.B., Kantarjian, H., Capdeville, R., Ohno-Jones, S.*, et al.* (2001). Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. N Engl J Med *344*, 1031-1037.

Durandy, A. (2003). Activation-induced cytidine deaminase: a dual role in class-switch recombination and somatic hypermutation. Eur J Immunol *33*, 2069-2073.

Durkin, S.G., and Glover, T.W. (2007). Chromosome fragile sites. Annu Rev Genet *41*, 169-192.

Einerson, R.R., Law, M.E., Blair, H.E., Kurtin, P.J., McClure, R.F., Ketterling, R.P., Flynn, H.C., Dogan, A., and Remstein, E.D. (2006). Novel FISH probes designed to detect IGK-MYC and IGL-MYC rearrangements in B-cell lineage malignancy identify a new breakpoint cluster region designated BVR2. Leukemia *20*, 1790-1799.

Falck, J., Coates, J., and Jackson, S.P. (2005). Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. Nature *434*, 605-611.

Fix, D.F., and Glickman, B.W. (1987). Asymmetric cytosine deamination revealed by spontaneous mutational specificity in an Ung- strain of Escherichia coli. Mol Gen Genet *209*, 78-82.

Ford, D., Easton, D.F., Bishop, D.T., Narod, S.A., and Goldgar, D.E. (1994). Risks of cancer in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. Lancet *343*, 692-695.

Frank, K.M., Sekiguchi, J.M., Seidl, K.J., Swat, W., Rathbun, G.A., Cheng, H.L., Davidson, L., Kangaloo, L., and Alt, F.W. (1998). Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV. Nature *396*, 173-177.

Frederico, L.A., Kunkel, T.A., and Shaw, B.R. (1990). A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. Biochemistry *29*, 2532-2537.

Fugazzola, L., Pilotti, S., Pinchera, A., Vorontsova, T.V., Mondellini, P., Bongarzone, I., Greco, A., Astakhova, L., Butti, M.G., Demidchik, E.P.*, et al.* (1995). Oncogenic rearrangements of the RET proto-oncogene in papillary thyroid carcinomas from children exposed to the Chernobyl nuclear accident. Cancer Res *55*, 5617-5620.

Gaidano, G., Ballerini, P., Gong, J.Z., Inghirami, G., Neri, A., Newcomb, E.W., Magrath, I.T., Knowles, D.M., and Dalla-Favera, R. (1991). p53 mutations in human lymphoid malignancies: association with Burkitt lymphoma and chronic lymphocytic leukemia. Proc Natl Acad Sci U S A *88*, 5413-5417.

Garcia-Rubio, M., Huertas, P., Gonzalez-Barrera, S., and Aguilera, A. (2003). Recombinogenic effects of DNA-damaging agents are synergistically increased

by transcription in Saccharomyces cerevisiae. New insights into transcription-associated recombination. Genetics *165*, 457-466.

Gauwerky, C.E., Huebner, K., Isobe, M., Nowell, P.C., and Croce, C.M. (1989). Activation of MYC in a masked t(8;17) translocation results in an aggressive B-cell leukemia. Proc Natl Acad Sci U S A *86*, 8867-8871.

Gesk, S., Klapper, W., Martin-Subero, J.I., Nagel, I., Harder, L., Fu, K., Bernd, H.W., Weisenburger, D.D., Parwaresch, R., and Siebert, R. (2006). A chromosomal translocation in cyclin D1-negative/cyclin D2-positive mantle cell lymphoma fuses the CCND2 gene to the IGK locus. Blood *108*, 1109-1110.

Glover, T.W., Berger, C., Coyle, J., and Echo, B. (1984). DNA polymerase alpha inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. Human genetics *67*, 136-142.

Gordon, M.S., Kanegai, C.M., Doerr, J.R., and Wall, R. (2003). Somatic hypermutation of the B cell receptor genes B29 (Igbeta, CD79b) and mb1 (Igalpha, CD79a). Proc Natl Acad Sci U S A *100*, 4126-4131.

Gostissa, M., Yan, C.T., Bianco, J.M., Cogne, M., Pinaud, E., and Alt, F.W. (2009). Long-range oncogenic activation of Igh-c-myc translocations by the Igh 3' regulatory region. Nature *462*, 803-807.

Gottipati, P., Cassel, T.N., Savolainen, L., and Helleday, T. (2008). Transcription-associated recombination is dependent on replication in Mammalian cells. Mol Cell Biol *28*, 154-164.

Gottipati, P., and Helleday, T. (2009). Transcription-associated recombination in eukaryotes: link between transcription, replication and recombination. Mutagenesis *24*, 203-210.

Gu, Y., Seidl, K.J., Rathbun, G.A., Zhu, C., Manis, J.P., van der Stoep, N., Davidson, L., Cheng, H.L., Sekiguchi, J.M., Frank, K.*, et al.* (1997). Growth retardation and leaky SCID phenotype of Ku70-deficient mice. Immunity *7*, 653-665.

Guikema, J.E., Linehan, E.K., Tsuchimoto, D., Nakabeppu, Y., Strauss, P.R., Stavnezer, J., and Schrader, C.E. (2007). APE1- and APE2-dependent DNA breaks in immunoglobulin class switch recombination. J Exp Med *204*, 3017-3026.

Hastings, P.J., Ira, G., and Lupski, J.R. (2009). A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS Genet *5*, e1000327.

Helgeson, B.E., Tomlins, S.A., Shah, N., Laxman, B., Cao, Q., Prensner, J.R., Cao, X., Singla, N., Montie, J.E., Varambally, S.*, et al.* (2008). Characterization of TMPRSS2:ETV5 and SLC45A3:ETV5 gene fusions in prostate cancer. Cancer Res *68*, 73-80.

Herman, R.K., and Dworkin, N.B. (1971). Effect of gene induction on the rate of mutagenesis by ICR-191 in Escherichia coli. J Bacteriol *106*, 543-550.

Huang, M.E., and Kolodner, R.D. (2005). A biological network in Saccharomyces cerevisiae prevents the deleterious effects of endogenous oxidative DNA damage. Mol Cell *17*, 709-720.

Huebner, K., and Croce, C.M. (2001). FRA3B and other common fragile sites: the weakest links. Nat Rev Cancer *1*, 214-221.

Hunt, J.L. (2011). An update on molecular diagnostics of squamous and salivary gland tumors of the head and neck. Arch Pathol Lab Med *135*, 602-609.

Huppi, K., Siwarski, D., Skurla, R., Klinman, D., and Mushinski, J.F. (1990). Pvt-1 transcripts are found in normal tissues and are altered by reciprocal(6;15) translocations in mouse plasmacytomas. Proc Natl Acad Sci U S A *87*, 6964-6968.

Iida, S., Rao, P.H., Ueda, R., Chaganti, R.S., and Dalla-Favera, R. (1999). Chromosomal rearrangement of the PAX-5 locus in lymphoplasmacytic lymphoma with t(9;14)(p13;q32). Leuk Lymphoma *34*, 25-33.

Imai, K., Slupphaug, G., Lee, W.I., Revy, P., Nonoyama, S., Catalan, N., Yel, L., Forveille, M., Kavli, B., Krokan, H.E.*, et al.* (2003). Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. Nat Immunol *4*, 1023-1028.

Impera, L., Albano, F., Lo Cunsolo, C., Funes, S., Iuzzolino, P., Laveder, F., Panagopoulos, I., Rocchi, M., and Storlazzi, C.T. (2008). A novel fusion 5'AFF3/3'BCL2 originated from a t(2;18)(q11.2;q21.33) translocation in follicular lymphoma. Oncogene *27*, 6187-6190.

Jager, U., Bocskor, S., Le, T., Mitterbauer, G., Bolz, I., Chott, A., Kneba, M., Mannhalter, C., and Nadel, B. (2000). Follicular lymphomas' BCL-2/IgH junctions contain templated nucleotide insertions: novel insights into the mechanism of t(14;18) translocation. Blood *95*, 3520-3529.

Jung, S., Rajewsky, K., and Radbruch, A. (1993). Shutdown of class switch recombination by deletion of a switch region control element. Science *259*, 984-987.

Keil, R.L., and Roeder, G.S. (1984). Cis-acting, recombination-stimulating activity in a fragment of the ribosomal DNA of S. cerevisiae. Cell *39*, 377-386.

Kim, J.W., Chang, E.M., Song, S.H., Park, S.H., Yoon, T.K., and Shim, S.H. (2011). Complex chromosomal rearrangements in infertile males: complexity of rearrangement affects spermatogenesis. Fertil Steril *95*, 349-352, 352 e341-345.

Komori, T., Okada, A., Stewart, V., and Alt, F.W. (1993). Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes. Science *261*, 1171-1175.

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

Kuppers, R. (2005). Mechanisms of B-cell lymphoma pathogenesis. Nat Rev Cancer *5*, 251-262.

Kuppers, R., Klein, U., Hansmann, M.L., and Rajewsky, K. (1999). Cellular origin of human B-cell lymphomas. N Engl J Med *341*, 1520-1529.

Lawley, P.D., and Phillips, D.H. (1996). DNA adducts from chemotherapeutic agents. Mutat Res *355*, 13-40.

Lebman, D.A., Lee, F.D., and Coffman, R.L. (1990). Mechanism for transforming growth factor beta and IL-2 enhancement of IgA expression in lipopolysaccharide-stimulated B cell cultures. J Immunol *144*, 952-959.

Lee-Theilen, M., Matthews, A.J., Kelly, D., Zheng, S., and Chaudhuri, J. (2011). CtIP promotes microhomology-mediated alternative end joining during class-switch recombination. Nat Struct Mol Biol *18*, 75-79.

Lenz, G., Nagel, I., Siebert, R., Roschke, A.V., Sanger, W., Wright, G.W., Dave, S.S., Tan, B., Zhao, H., Rosenwald, A.*, et al.* (2007). Aberrant immunoglobulin class switch recombination and switch translocations in activated B cell-like diffuse large B cell lymphoma. J Exp Med *204*, 633-643.

Libura, J., Slater, D.J., Felix, C.A., and Richardson, C. (2005). Therapy-related acute myeloid leukemia-like MLL rearrangements are induced by etoposide in primary human CD34+ cells and remain stable after clonal expansion. Blood *105*, 2124-2131.

Lieber, M.R. (2008). The mechanism of human nonhomologous DNA end joining. J Biol Chem *283*, 1-5.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., *et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.

Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B.G., Ohgi, K., Zhang, J., Rose, D.W., Fu, X.D., Glass, C.K., *et al.* (2009). Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell *139*, 1069-1083.

Liu, B., and Alberts, B.M. (1995). Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. Science *267*, 1131-1137.

Liu, J., and Levens, D. (2006). Making myc. Curr Top Microbiol Immunol *302*, 1-32.

Liu, M., Duke, J.L., Richter, D.J., Vinuesa, C.G., Goodnow, C.C., Kleinstein, S.H., and Schatz, D.G. (2008). Two levels of protection for the B cell genome during somatic hypermutation. Nature *451*, 841-845.

Lu, D., and Yunis, J.J. (1992). Cloning, expression and localization of an RNA helicase gene from a human lymphoid cell line with chromosomal breakpoint 11q23.3. Nucleic Acids Res *20*, 1967-1972.

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

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

Mani, R.S., and Chinnaiyan, A.M. (2010). Triggers for genomic rearrangements: insights into genomic, cellular and environmental influences. Nat Rev Genet *11*, 819-829.

Mani, R.S., Tomlins, S.A., Callahan, K., Ghosh, A., Nyati, M.K., Varambally, S., Palanisamy, N., and Chinnaiyan, A.M. (2009). Induced chromosomal proximity and gene fusions in prostate cancer. Science *326*, 1230.

Manis, J.P., Dudley, D., Kaylor, L., and Alt, F.W. (2002). IgH class switch recombination to IgG1 in DNA-PKcs-deficient B cells. Immunity *16*, 607-617.

Manis, J.P., Morales, J.C., Xia, Z., Kutok, J.L., Alt, F.W., and Carpenter, P.B. (2004). 53BP1 links DNA damage-response pathways to immunoglobulin heavy chain class-switch recombination. Nat Immunol *5*, 481-487.

Manuelidis, L. (1985). Individual interphase chromosome domains revealed by in situ hybridization. Hum Genet *71*, 288-293.

Mao, X., Lillington, D., Child, F., Russell-Jones, R., Young, B., and Whittaker, S. (2002). Comparative genomic hybridization analysis of primary cutaneous B-cell lymphomas: identification of common genomic alterations in disease pathogenesis. Genes Chromosomes Cancer *35*, 144-155.

Marcu, K.B., Bossone, S.A., and Patel, A.J. (1992). myc function and regulation. Annu Rev Biochem *61*, 809-860.

Marnett, L.J. (2000). Oxyradicals and DNA damage. Carcinogenesis *21*, 361-370.

Martin, A., Bardwell, P.D., Woo, C.J., Fan, M., Shulman, M.J., and Scharff, M.D. (2002). Activation-induced cytidine deaminase turns on somatic hypermutation in hybridomas. Nature *415*, 802-806.

Martin, A., Li, Z., Lin, D.P., Bardwell, P.D., Iglesias-Ussel, M.D., Edelmann, W., and Scharff, M.D. (2003). Msh2 ATPase activity is essential for somatic hypermutation at a-T basepairs and for efficient class switch recombination. J Exp Med *198*, 1171-1178.

Martomo, S.A., Yang, W.W., and Gearhart, P.J. (2004). A role for Msh6 but not Msh3 in somatic hypermutation and class switch recombination. J Exp Med *200*, 61-68.

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

Mottok, A., Renne, C., Seifert, M., Oppermann, E., Bechstein, W., Hansmann, M.L., Kuppers, R., and Brauninger, A. (2009). Inactivating SOCS1 mutations are caused by aberrant somatic hypermutation and restricted to a subset of B-cell lymphoma entities. Blood *114*, 4503-4506.

Moynahan, M.E., and Jasin, M. (2010). Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. Nat Rev Mol Cell Biol *11*, 196-207.

Mueller, P.R., and Wold, B. (1989). In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. Science (New York, N Y) *246*, 780-786.

Muljo, S.A., and Schlissel, M.S. (2003). A small molecule Abl kinase inhibitor induces differentiation of Abelson virus-transformed pre-B cell lines. Nat Immunol *4*, 31-37.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell *102*, 553-563.

Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O., and Honjo, T. (1999). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J Biol Chem *274*, 18470-18476.

Muschen, M., Re, D., Jungnickel, B., Diehl, V., Rajewsky, K., and Kuppers, R. (2000). Somatic mutation of the CD95 gene in human B cells as a side-effect of the germinal center reaction. J Exp Med *192*, 1833-1840.

Muto, T., Okazaki, I.M., Yamada, S., Tanaka, Y., Kinoshita, K., Muramatsu, M., Nagaoka, H., and Honjo, T. (2006). Negative regulation of activation-induced cytidine deaminase in B cells. Proc Natl Acad Sci U S A *103*, 2752-2757.

Nakamura, Y., Takahashi, N., Kakegawa, E., Yoshida, K., Ito, Y., Kayano, H., Niitsu, N., Jinnai, I., and Bessho, M. (2008). The GAS5 (growth arrest-specific transcript 5) gene fuses to BCL6 as a result of t(1;3)(q25;q27) in a patient with B-cell lymphoma. Cancer Genet Cytogenet *182*, 144-149.

Nambu, Y., Sugai, M., Gonda, H., Lee, C.G., Katakai, T., Agata, Y., Yokota, Y., and Shimizu, A. (2003). Transcription-coupled events associating with immunoglobulin switch region chromatin. Science *302*, 2137-2140.

Nechaev, S., Fargo, D.C., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science *327*, 335-338.

Neves, H., Ramos, C., da Silva, M.G., Parreira, A., and Parreira, L. (1999). The nuclear topography of ABL, BCR, PML, and RARalpha genes: evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation. Blood *93*, 1197-1207.

Nickoloff, J.A. (1992). Transcription enhances intrachromosomal homologous recombination in mammalian cells. Mol Cell Biol *12*, 5311-5318.

Nickoloff, J.A., and Reynolds, R.J. (1990). Transcription stimulates homologous recombination in mammalian cells. Mol Cell Biol *10*, 4837-4845.

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

Nussenzweig, A., and Nussenzweig, M.C. (2010). Origin of chromosomal translocations in lymphoid cancer. Cell *141*, 27-38.

Okazaki, I.M., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K., and Honjo, T. (2003). Constitutive expression of AID leads to tumorigenesis. J Exp Med *197*, 1173-1181.

Papavasiliou, F.N., and Schatz, D.G. (2000). Cell-cycle-regulated DNA double-stranded breaks in somatic hypermutation of immunoglobulin genes. Nature *408*, 216-221.

Pasqualucci, L., Migliazza, A., Fracchiolla, N., William, C., Neri, A., Baldini, L., Chaganti, R.S., Klein, U., Kuppers, R., Rajewsky, K.*, et al.* (1998). BCL-6 mutations in normal germinal center B cells: evidence of somatic hypermutation acting outside Ig loci. Proc Natl Acad Sci U S A *95*, 11816-11821.

Pasqualucci, L., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R.S., Kuppers, R., and Dalla-Favera, R. (2001). Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. Nature *412*, 341-346.

Patel, K.J., Yu, V.P., Lee, H., Corcoran, A., Thistlethwaite, F.C., Evans, M.J., Colledge, W.H., Friedman, L.S., Ponder, B.A., and Venkitaraman, A.R. (1998). Involvement of Brca2 in DNA repair. Mol Cell *1*, 347-357.

Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina San-Martin, B., Barreto, V.*, et al.* (2010). Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. Cell *143*, 122-133.

Peled, J.U., Kuang, F.L., Iglesias-Ussel, M.D., Roa, S., Kalis, S.L., Goodman, M.F., and Scharff, M.D. (2008). The biochemistry of somatic hypermutation. Annu Rev Immunol *26*, 481-511.

Petersen, S., Casellas, R., Reina-San-Martin, B., Chen, H.T., Difilippantonio, M.J., Wilson, P.C., Hanitsch, L., Celeste, A., Muramatsu, M., Pilch, D.R.*, et al.* (2001). AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. Nature *414*, 660-665.

Petersen-Mahrt, S.K., Harris, R.S., and Neuberger, M.S. (2002). AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. Nature *418*, 99-103.

Pfeifer, G.P., Steigerwald, S.D., Mueller, P.R., Wold, B., and Riggs, A.D. (1989). Genomic sequencing and methylation analysis by ligation mediated PCR. Science (New York, N Y) *246*, 810-813.

Pham, P., Bransteitter, R., Petruska, J., and Goodman, M.F. (2003). Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. Nature *424*, 103-107.

Pleasance, E.D., Cheetham, R.K., Stephens, P.J., McBride, D.J., Humphray, S.J., Greenman, C.D., Varela, I., Lin, M.L., Ordonez, G.R., Bignell, G.R.*, et al.* (2010a). A comprehensive catalogue of somatic mutations from a human cancer genome. Nature *463*, 191-196.

Pleasance, E.D., Stephens, P.J., O'Meara, S., McBride, D.J., Meynert, A., Jones, D., Lin, M.L., Beare, D., Lau, K.W., Greenman, C.*, et al.* (2010b). A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature *463*, 184-190.

Qiu, J.X., Kale, S.B., Yarnell Schultz, H., and Roth, D.B. (2001). Separation-of-function mutants reveal critical roles for RAG2 in both the cleavage and joining steps of V(D)J recombination. Mol Cell *7*, 77-87.

Rada, C., Williams, G.T., Nilsen, H., Barnes, D.E., Lindahl, T., and Neuberger, M.S. (2002). Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. Curr Biol *12*, 1748-1755.

Raghavan, S.C., Kirsch, I.R., and Lieber, M.R. (2001). Analysis of the V(D)J recombination efficiency at lymphoid chromosomal translocation breakpoints. J Biol Chem *276*, 29126-29133.

Raghavan, S.C., Swanson, P.C., Wu, X., Hsieh, C.L., and Lieber, M.R. (2004). A non-B-DNA structure at the Bcl-2 major breakpoint region is cleaved by the RAG complex. Nature *428*, 88-93.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432-445.

Rajagopal, D., Maul, R.W., Ghosh, A., Chakraborty, T., Khamlichi, A.A., Sen, R., and Gearhart, P.J. (2009). Immunoglobulin switch mu sequence causes RNA polymerase II accumulation and reduces dA hypermutation. J Exp Med *206*, 1237-1244.

Ramiro, A.R., Jankovic, M., Callen, E., Difilippantonio, S., Chen, H.T., McBride, K.M., Eisenreich, T.R., Chen, J., Dickins, R.A., Lowe, S.W.*, et al.* (2006). Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. Nature *440*, 105-109.

Ramiro, A.R., Jankovic, M., Eisenreich, T., Difilippantonio, S., Chen-Kiang, S., Muramatsu, M., Honjo, T., Nussenzweig, A., and Nussenzweig, M.C. (2004). AID is required for c-myc/IgH chromosome translocations in vivo. Cell *118*, 431-438.

Ramiro, A.R., Stavropoulos, P., Jankovic, M., and Nussenzweig, M.C. (2003). Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. Nat Immunol *4*, 452-456.

Reina-San-Martin, B., Chen, H.T., Nussenzweig, A., and Nussenzweig, M.C. (2004). ATM is required for efficient recombination between immunoglobulin switch regions. J Exp Med *200*, 1103-1110.

Richardson, C., and Jasin, M. (2000). Frequent chromosomal translocations induced by DNA double-strand breaks. Nature *405*, 697-700.

Riggi, N., and Stamenkovic, I. (2007). The Biology of Ewing sarcoma. Cancer Lett *254*, 1-10.

Robbiani, D.F., Bothmer, A., Callen, E., Reina-San-Martin, B., Dorsett, Y., Difilippantonio, S., Bolland, D.J., Chen, H.T., Corcoran, A.E., Nussenzweig, A.*, et al.* (2008). AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. Cell *135*, 1028-1038.

Robbiani, D.F., Bunting, S., Feldhahn, N., Bothmer, A., Camps, J., Deroubaix, S., McBride, K.M., Klein, I.A., Stone, G., Eisenreich, T.R.*, et al.* (2009). AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. Mol Cell *36*, 631-641.

Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J Biol Chem *273*, 5858-5868.

Roix, J.J., McQueen, P.G., Munson, P.J., Parada, L.A., and Misteli, T. (2003). Spatial proximity of translocation-prone gene loci in human lymphomas. Nat Genet *34*, 287-291.

Rosebeck, S., Madden, L., Jin, X., Gu, S., Apel, I.J., Appert, A., Hamoudi, R.A., Noels, H., Sagaert, X., Van Loo, P.*, et al.* (2011). Cleavage of NIK by the API2-MALT1 fusion oncoprotein leads to noncanonical NF-kappaB activation. Science *331*, 468-472.

Roth, D.B. (2003). Restraining the V(D)J recombinase. Nat Rev Immunol *3*, 656-666.

Roth, D.B., and Wilson, J.H. (1986). Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. Mol Cell Biol *6*, 4295-4304.

Rouet, P., Smih, F., and Jasin, M. (1994). Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol Cell Biol *14*, 8096-8106.

Savic, V., Sanborn, K.B., Orange, J.S., and Bassing, C.H. (2009). Chipping away at gamma-H2AX foci. Cell Cycle *8*, 3285-3290.

Sawyers, C.L. (1993). Molecular consequences of the BCR-ABL translocation in chronic myelogenous leukemia. Leuk Lymphoma *11 Suppl 2*, 101-103.

Schildkraut, E., Miller, C.A., and Nickoloff, J.A. (2006). Transcription of a donor enhances its use during double-strand break-induced gene conversion in human cells. Mol Cell Biol *26*, 3098-3105.

Schrader, C.E., Edelmann, W., Kucherlapati, R., and Stavnezer, J. (1999). Reduced isotype switching in splenic B cells from mice deficient in mismatch repair enzymes. J Exp Med *190*, 323-330.

Schrader, C.E., Linehan, E.K., Mochegova, S.N., Woodland, R.T., and Stavnezer, J. (2005). Inducible DNA breaks in Ig S regions are dependent on AID and UNG. J Exp Med *202*, 561-568.

Schroder, A.R.W., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. Cell *110*, 521-529.

Schultz, L.B., Chehab, N.H., Malikzay, A., and Halazonetis, T.D. (2000). p53 binding protein 1 (53BP1) is an early participant in the cellular response to DNA double-strand breaks. J Cell Biol *151*, 1381-1390.

Schwartz, M., Zlotorynski, E., Goldberg, M., Ozeri, E., Rahat, A., le Sage, C., Chen, B.P., Chen, D.J., Agami, R., and Kerem, B. (2005). Homologous recombination and nonhomologous end-joining repair pathways regulate fragile site stability. Genes Dev *19*, 2715-2726.

Shen, H.M., Bozek, G., Pinkert, C.A., McBride, K., Wang, L., Kenter, A., and Storb, U. (2008). Expression of AID transgene is regulated in activated B cells but not in resting B cells and kidney. Mol Immunol *45*, 1883-1892.

Shen, H.M., Peters, A., Baron, B., Zhu, X., and Storb, U. (1998). Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. Science *280*, 1750-1752.

Simsek, D., and Jasin, M. (2010). Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. Nat Struct Mol Biol *17*, 410-416.

Sohail, A., Klapacz, J., Samaranayake, M., Ullah, A., and Bhagwat, A.S. (2003). Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. Nucleic Acids Res *31*, 2990-2994.

Sonoda, E., Sasaki, M.S., Buerstedde, J.M., Bezzubova, O., Shinohara, A., Ogawa, H., Takata, M., Yamaguchi-Iwai, Y., and Takeda, S. (1998). Rad51-deficient vertebrate cells accumulate chromosomal breaks prior to cell death. EMBO J *17*, 598-608.

Stack, S.M., Brown, D.B., and Dewey, W.C. (1977). Visualization of interphase chromosomes. J Cell Sci *26*, 281-299.

Staszewski, O., Baker, R.E., Ucher, A.J., Martier, R., Stavnezer, J., and Guikema, J.E. (2011). Activation-induced cytidine deaminase induces reproducible DNA breaks at many non-Ig Loci in activated B cells. Mol Cell *41*, 232-242.

Stavnezer, J., Guikema, J.E., and Schrader, C.E. (2008). Mechanism and regulation of class switch recombination. Annu Rev Immunol *26*, 261-292.

Stavnezer-Nordgren, J., and Sirlin, S. (1986). Specificity of immunoglobulin heavy chain switch correlates with activity of germline heavy chain genes prior to switching. EMBO J *5*, 95-102.

Stephens, P.J., McBride, D.J., Lin, M.L., Varela, I., Pleasance, E.D., Simpson, J.T., Stebbings, L.A., Leroy, C., Edkins, S., Mudie, L.J*., et al.* (2009). Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature *462*, 1005-1010.

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

Takata, M., Sasaki, M.S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A., and Takeda, S. (1998). Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. EMBO J *17*, 5497-5508.

Takeuchi, Y., Horiuchi, T., and Kobayashi, T. (2003). Transcription-dependent recombination and the role of fork collision in yeast rDNA. Genes Dev *17*, 1497-1506.

Takizawa, T., Gudla, P.R., Guo, L., Lockett, S., and Misteli, T. (2008). Allele-specific nuclear positioning of the monoallelically expressed astrocyte marker GFAP. Genes Dev *22*, 489-498.

Thomas, B.J., and Rothstein, R. (1989). Elevated recombination rates in transcriptionally active DNA. Cell *56*, 619-630.

Tsai, C.L., Drejer, A.H., and Schatz, D.G. (2002). Evidence of a critical architectural function for the RAG proteins in end processing, protection, and joining in V(D)J recombination. Genes Dev *16*, 1934-1949.

Tumas-Brundage, K., and Manser, T. (1997). The transcriptional promoter regulates hypermutation of the antibody heavy chain locus. J Exp Med *185*, 239-250.

Ueda, C., Akasaka, T., Kurata, M., Maesako, Y., Nishikori, M., Ichinohasama, R., Imada, K., Uchiyama, T., and Ohno, H. (2002). The gene for interleukin-21 receptor is the partner of BCL6 in t(3;16)(q27;p11), which is recurrently observed in diffuse large B-cell lymphoma. Oncogene *21*, 368-376.

Vaandrager, J.W., Schuuring, E., Philippo, K., and Kluin, P.M. (2000). V(D)J recombinase-mediated transposition of the BCL2 gene to the IGH locus in follicular lymphoma. Blood *96*, 1947-1952.

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

Vilette, D., Ehrlich, S.D., and Michel, B. (1996). Transcription-induced deletions in plasmid vectors: M13 DNA replication as a source of instability. Mol Gen Genet *252*, 398-403.

Voelkel-Meiman, K., Keil, R.L., and Roeder, G.S. (1987). Recombination-stimulating sequences in yeast ribosomal DNA correspond to sequences regulating transcription by RNA polymerase I. Cell *48*, 1071-1079.

Wang, J.H., Gostissa, M., Yan, C.T., Goff, P., Hickernell, T., Hansen, E., Difilippantonio, S., Wesemann, D.R., Zarrin, A.A., Rajewsky, K.*, et al.* (2009). Mechanisms promoting translocations in editing and switching peripheral B cells. Nature *460*, 231-236.

Ward, I.M., Reina-San-Martin, B., Olaru, A., Minn, K., Tamada, K., Lau, J.S., Cascalho, M., Chen, L., Nussenzweig, A., Livak, F.*, et al.* (2004). 53BP1 is required for class switch recombination. J Cell Biol *165*, 459-464.

Weinstock, D.M., Brunet, E., and Jasin, M. (2007). Formation of NHEJ-derived reciprocal chromosomal translocations does not require Ku70. Nat Cell Biol *9*, 978-981.

Weiss, L.M., Warnke, R.A., Sklar, J., and Cleary, M.L. (1987). Molecular analysis of the t(14;18) chromosomal translocation in malignant lymphomas. N Engl J Med *317*, 1185-1189.

Wierstra, I., and Alves, J. (2008). The c-myc promoter: still MysterY and challenge. Adv Cancer Res *99*, 113-333.

Windbichler, N., Papathanos, P.A., Catteruccia, F., Ranson, H., Burt, A., and Crisanti, A. (2007). Homing endonuclease mediated gene targeting in Anopheles gambiae cells and embryos. Nucleic Acids Res *35*, 5922-5933.

Xie, A., Hartlerode, A., Stucki, M., Odate, S., Puget, N., Kwok, A., Nagaraju, G., Yan, C., Alt, F.W., Chen, J.*, et al.* (2007). Distinct roles of chromatin-associated proteins MDC1 and 53BP1 in mammalian double-strand break repair. Mol Cell *28*, 1045-1057.

Xue, K., Rada, C., and Neuberger, M.S. (2006). The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in msh2-/- ung-/- mice. J Exp Med *203*, 2085-2094.

Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H.W., Robbiani, D.F., McBride, K., Nussenzweig, M.C., and Casellas, R. (2011). Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. Nat Immunol *12*, 62-69.

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

Ye, B.H., Chaganti, S., Chang, C.C., Niu, H., Corradini, P., Chaganti, R.S., and Dalla-Favera, R. (1995). Chromosomal translocations cause deregulated BCL6 expression by promoter substitution in B cell lymphoma. EMBO J *14*, 6209-6217.

Yoshida, S., Kaneita, Y., Aoki, Y., Seto, M., Mori, S., and Moriyama, M. (1999). Identification of heterologous translocation partner genes fused to the BCL6 gene in diffuse large B-cell lymphomas: 5'-RACE and LA - PCR analyses of biopsy samples. Oncogene *18*, 7994-7999.

Yoshikawa, K., Okazaki, I.M., Eto, T., Kinoshita, K., Muramatsu, M., Nagaoka, H., and Honjo, T. (2002). AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. Science *296*, 2033-2036.

You, Z., Chahwan, C., Bailis, J., Hunter, T., and Russell, P. (2005). ATM activation and its recruitment to damaged DNA require binding to the C terminus of Nbs1. Mol Cell Biol *25*, 5363-5379.

Yu, K., Chedin, F., Hsieh, C.L., Wilson, T.E., and Lieber, M.R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol *4*, 442-451.

Zhang, L., and Gralla, J.D. (1989). In situ nucleoprotein structure at the SV40 major late promoter: melted and wrapped DNA flank the start site. Genes Dev *3*, 1814-1822.

Zhao, Q., Caballero, O.L., Levy, S., Stevenson, B.J., Iseli, C., de Souza, S.J., Galante, P.A., Busam, D., Leversha, M.A., Chadalavada, K.*, et al.* (2009). Transcriptome-guided characterization of genomic rearrangements in a breast cancer cell line. Proc Natl Acad Sci U S A *106*, 1886-1891.

Zhu, C., Bogue, M.A., Lim, D.S., Hasty, P., and Roth, D.B. (1996). Ku86-deficient mice exhibit severe combined immunodeficiency and defective processing of V(D)J recombination intermediates. Cell *86*, 379-389.

Zhu, C., and Roth, D.B. (1995). Characterization of coding ends in thymocytes of scid mice: implications for the mechanism of V(D)J recombination. Immunity *2*, 101-112.

Zlotorynski, E., Rahat, A., Skaug, J., Ben-Porat, N., Ozeri, E., Hershberg, R., Levi, A., Scherer, S.W., Margalit, H., and Kerem, B. (2003). Molecular basis for expression of common and rare fragile sites. Mol Cell Biol *23*, 7143-7151.