2014

Taming a Potent Mutator: Identification and Characterization of Novel Mechanisms of Regulating Antibody Diversification in B-Lymphocytes

Rebecca Kyle Delker

TAMING A POTENT MUTATOR: IDENTIFICATION AND CHARACTERIZATION OF NOVEL MECHANISMS OF REGULATING ANTIBODY DIVERSIFICATION 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

Rebecca Kyle Delker

June 2014

TAMING A POTENT MUTATOR: IDENTIFICATION AND CHARACTERIZATION OF NOVEL MECHANISMS OF REGULATING ANTIBODY DIVERSIFICATION IN B-LYMPHOCYTES

Rebecca Kyle Delker, Ph.D.
The Rockefeller University 2014

Often the needs of an organism exceed the number of genes in the genome. Thus, modification of the genes, themselves, or of the gene products is necessary. This becomes particularly important in cells of the immune system, which have to combat a virtually infinite array of foreign pathogens. B-lymphocytes, the mediators of humoral immunity, have developed extensive mechanisms of gene diversification collectively known as antibody diversification. Antibody diversification, a set of processes necessary for an organism to mount a specific and robust immune response, relies on Activation Induced Cytidine Deaminase (AID) to initiate two of such processes: Somatic Hypermutation (SHM) and Class Switch Recombination (CSR). AID-dependent deamination of cytidine bases within the variable region (SHM) and switch region (CSR) of the immunoglobulin locus (Ig) results in the modification of the antigen binding domain and diversity within effector function of the antibody, respectively. Though the activity of AID is known, the regulation of AID during the different stages of antibody diversification is less well understood. This question has been particularly challenging to address because of the difficulty of working with AID, which becomes insoluble when expressed in non B-cells.

This thesis presents the development of a screen, which searches for interacting partners for poorly soluble proteins. This screen relies on the

insolubility of the protein of interest and the ability of interacting proteins to induce solubilization via binding and masking of exposed hydrophobic domains.

After validation of this screen using representative soluble and insoluble proteins, it was applied to AID and thirty putative AID binding partners were identified. A handful of these proteins were uncovered in prior interaction screens, thus underscoring the validity of this new screening approach.

In addition, this thesis presents a comprehensive analysis, utilizing both *in vitro* and *in vivo* approaches, of one of the putative AID cofactors discovered in the screen, RING Finger Protein 126 (RNF126). *In vitro* studies revealed that RNF126 is a *bona fide* AID binding partner and, in addition, acts as an E3 ubiquitin ligase, modifying AID with the addition of a single ubiquitin moiety. Further, a conditional knockout model of RNF126 was generated and used to determine that RNF126 plays a role *in vivo* in fine-tuning AID activity during SHM and CSR.

The findings presented here demonstrate the utility of a novel screening technique to search for interacting partners for insoluble proteins and, through its use, expands the list of putative AID cofactors. In addition, through a thorough analysis of a single AID binding partner, this thesis puts forth a novel mode of regulation of the potent mutating enzyme, paving the way for future research to uncover the role of mono-ubiquitinated AID during SHM and/or CSR.

ACKNOWLEDGEMENTS

I would like to take this space to thank all of the people who have supported me throughout my PhD. First, my advisor, Nina Papavasiliou. Nina has been supportive both of this project and my career, for which I am incredibly grateful.

In addition, all of the members of the Papavasiliou Lab, both past and present, have provided me with tremendous support even if they were not directly connected to the project. These wonderful people include Claire Hamilton, Monica Mugnier, Linda Molla, Pete Stavropoulos, Dewi Harjanto, Galadriel Hovel-Miner, Jason Pinger, Violeta Rayon Estrada, Danae Schulz, Maryam Zaringhalam, Anita Ramnarain, Eric Fritz, Paul Hakimpour, Alex Strikoudis, Yanjiao Zhou, Jan Davidson-Moncada, Brad Rosenberg, Catharine Boothroyd, Tanya Leonova, and Peter Alff. A special thanks goes to Alex Strikoudis and Yanjiao Zhou, who conducted experiments necessary to provide the foundation for my thesis project.

In addition to the support I received from within the laboratory, I was lucky enough to have a committed Faculty Thesis Committee to help guide the project forward. I would like to thank Howard Hang and Agata Smogorzewska for their helpful suggestions throughout this process. I would also like to thank my external committee member, Jayanta Chaudhuri, for his involvement in this process.

I would also like to thank my family and friends for their support throughout the past six years. The PhD process is akin to the most terrifying (and exhilarating?) roller coaster ride you experienced as a child. For me it was “The Ride” at Universal Studios Hollywood—a water ride, which contains a terrifying, 85-foot, near-vertical drop toward the end. Much like this ride, the PhD begins with hopeful anticipation; the initial twists and turns are exciting, but as with any challenging endeavor, become more stressful as time passes. Fearfulness of what is to come begins to build, until you have hit the top the 85-foot drop. The ride slows and it seems like it will take forever before it is all over (4th year, ‘nuf said) and then, before you even have a chance to process it, you are summoned to exit. As you walk away, your heartbeat begins to slow back down to normal and you begin your search for the next adventure.

I would like to thank all of those who eagerly awaited my arrival at the end and cheered me on during the many ups and downs and twists and turns of the PhD experience. And last, but not least, a very special thanks goes to the one sitting beside me who experienced all of the highs and lows along with me, and who will be there years from now to look back and reminisce about how scary it all seemed and how far we have come.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

LIST OF TABLES

CHAPTER 1

CHAPTER 3

CHAPTER 6

CHAPTER 1: INTRODUCTION

The blueprint of an organism is its genome; that is to say that the genome is the most fundamental code necessary for life. However, much of the diversity of life and the diversity of processes within a cell necessitate modifications to this basic code. Thus the “Central Dogma” that one gene produces one RNA, which produces one protein, has proven to be too simplistic. Mechanisms of diversification have been described at each level of the Central Dogma. Protein structure and function is altered with the use of post-translational modifications (Prabakaran et al., 2012); the composition of, and thus the information contained within, RNA is modified through processes such as alternative splicing and RNA-editing (Hamilton et al., 2010); and lastly, modifications to the DNA, itself, have been instrumental in generating diversity on a cell-to-cell basis. While manipulation, or mutation, of the genome can be detrimental to the organism by promoting genomic instability, complete fidelity in this regard would abrogate evolution at the organismal and cellular level. Thus, modification at the DNA level must be tightly regulated to generate a balance between diversity (or genomic instability) and genomic stability.

One of the most fundamental questions in biology is that of how an organism, or more simply, a cell, is able to respond to a virtually infinite and unknown array of environmental factors given only a limited genome to work with. In many cases, organisms solve this problem by hard-coding many possible surface-bound receptor proteins into the genome in gene families. In these instances,

epigenetic mechanisms are used to regulate mono-allelic expression to ensure that each cell only expresses one of several possible receptors. Complexity in the organism's response, then, is derived from the presence of a population of cells, each expressing a unique receptor. This sort of biological approach can be seen in receptors of sensory systems, such as olfactory receptors for smell (Serizawa et al., 2004) and rhodopsin expression for sight (Johnston and Desplan, 2008; 2010).

In instances where much greater diversity is demanded by selective pressure, additional possibilities that go beyond the scope of the genome are necessary. This idea is particularly prevalent in the immune systems of higher vertebrates, where organisms must combat an unknown and virtually unlimited set of external pathogens. Thus, this system must be anticipatory, that is, it must be prepared to mount a response even before this information is known. Despite the human genome containing approximately 20,000 genes, the two antigen-recognition cell types of the vertebrate adaptive immune system, B and T Lymphocytes, are able to generate greater than 10^7 different receptors against potential pathogens (Market and Papavasiliou, 2003). B Lymphocytes, in particular, make use of additional diversification mechanisms to generate a potentially limitless number of receptors, increasing the host's chance of survival. An exploration of pathogen-defense mechanisms at different branches of evolution reveals a progression from hard-coded immune receptors to receptors with flexibility in their specificity due to various mechanisms of genomic

modification.

1.1 The Evolution of Adaptive Immunity

At the heart of immunity is an organism's ability to differentiate between self and non-self. Organisms of all taxonomical clades have developed mechanisms to do just this, but arguably the most complex systems developed can be found within the adaptive immune systems of vertebrates. The journey to this sort of complexity began with the onset of Innate Immunity, a system alive and well in complex vertebrates.

During evolution, innate immune defenses emerged prior to adaptive immunity and innate immune receptors can be found throughout the animal kingdom and in plants (Flajnik and Kasahara, 2010; Medzhitov and Janeway, 2000). Innate immune receptors are hard-coded in the genome, such that they are passed on to subsequent generations. Because of this, only a limited number of receptors can exist, so as not to overwhelm the genome. Thus, rather than encoding receptors that are specific for unique pathogens, innate immune receptors have evolved to recognize patterns, or motifs, that are common to a number of invading pathogens. For this reason, innate immune receptors are often referred to as Pattern Recognition Receptors (PRRs). Common pathogen-associated motifs (PAMs) detected by these receptors include lipopolysaccharide (LPS), peptidoglycan, bacterial DNA, and viral dsRNA (Medzhitov and Janeway, 2000). Despite being limited in number, because all innate immune cells express

the same repertoire of PRRs, the innate immune response is both rapid and robust.

Because of the constant challenge posed by new types of pathogens, as well as the ability of a single pathogen to evolve to evade immune defenses, even organisms that contain only innate immune systems are under selective pressure to diversify their receptors. Similar to how sensory receptors are encoded as large gene families, many organisms have expanded and diversified gene families encoding innate immune receptors, thus allowing for the expression of receptors of varying specificities (Litman et al., 2005b). This is exemplified in the sea urchin, which contains over 200 Toll-Like Receptors (TLRs) as compared to between 1 and 20 in other animals (Buckley et al., 2008b; Roach et al., 2005). Additional mechanisms observed include: combinatorial use of encoded receptors and gene recombination (Buckley et al., 2008a), inclusion of nucleotide polymorphisms (Lazzaro et al., 2004) and/or alternative splicing of immune receptor genes and gene families (Danilova, 2012). For example, the drosophila immune cells have been shown to express more than 18,000 isoforms of the immunoglobulin-structured receptor, Dscam, through alternative splicing of variable exons with constant exons (Watson et al., 2005). As an interesting aside, even single-celled organisms such as bacteria have developed fairly sophisticated forms of defense. Bacteria have developed defense mechanisms that are both general, much like innate receptors, which involves the digestion of foreign DNA by the recognition of palindromic sequences, as well as a more

specific defense mechanism known as CRISPR, which involves RNAi-like machinery (Fineran and Charpentier, 2012).

Importantly, evolution of innate immune responses occurs at the level of the organism and on a Darwinian evolution time-scale, improving defense systems from one generation to the next. In contrast, evolution and adaptation within the adaptive immune system occurs at the cellular level, within the lifetime of a single organism. Although it was initially thought that conventional adaptive immunity, consisting of B and T lymphocytes, appeared only in jawed vertebrates in a sort of “big bang” mechanism, evidence now exists to suggest that even jawless vertebrates have evolved an adaptive immune system that displays similar principles of design (Boehm, 2011; Danilova, 2012; Flajnik and Kasahara, 2010; Litman et al., 2005a; 2010). The occurrence of convergent evolution within these two branches of vertebrates suggests that it is more likely that the evolution of conventional adaptive immunity occurred incrementally, with the stepwise acquisition of additional diversifying mechanisms. The main commonalities of all vertebrate adaptive immune systems are the presence of a two-cell system and the expression of receptors, which have been built from gene segments to generate diversity. Conventional adaptive immunity in jawed vertebrates relies on two cell types: B lymphocytes (B cells) and T lymphocytes (T cells). B cells, which act in the humoral immune response, recognize native antigen, and T cells, which participate in cell-mediated immunity, recognize antigen in the context of an antigen presenting cell (APC). Similarly, jawless vertebrates (e.g.

hagfish and lamprey) have been found to possess two distinct lymphocyte-like lineages, which function in humoral and cellular immunity, respectively. This fact suggests that either an extinct common vertebrate ancestor also possessed lymphocyte-like cells, which provided the foundation for the evolution of both forms of adaptive immunity, or that a lymphocyte-precursor cell existed in invertebrates (Boehm, 2011). Notably, other necessary features of conventional adaptive immunity can be found in extant invertebrate species. Most interestingly, the enzymes RAG1/2, which are essential for jawed-vertebrate antibody diversification, were found in the genome of the sea urchin (Fugmann et al., 2006).

The receptors of jawless and jawed adaptive immune systems are structurally unique. Jawless vertebrates express receptors called Variable Lymphocyte Receptors (VLRs), which are structurally similar to PRRs of the innate immune system and contain leucine-rich modules. In contrast, B cells and T cells produce receptors known as the B-cell Receptor (BCR, or antibody) and the T-cell Receptor (TCR); the structure of both receptors consist of characteristic immunoglobulin folds. Despite these differences, the receptors derived from both branches of vertebrates can be highly diversified through the recombination of gene segments. In both cases, it is believed that multiple and diverse gene segments evolved from several duplication events paired with transposition (Boehm, 2011). The mechanistic details of how these gene segments are recombined are slightly different between jawless and jawed vertebrates,

however both likely evolved from the same selective pressure. Jawless vertebrates rely on gene conversion recombination events, placing one of many leucine-rich modules into an incomplete VLR locus. Interestingly, this gene conversion event occurs alongside expression of putative cytosine deaminases, which show homology to the vertebrate antibody diversification enzyme, Activation Induced Cytidine Deaminase (AID) (Rogozin et al., 2007).

Most jawed-vertebrates recombine gene segments using recombination enzymes likely derived from transposon insertion, in a process known as V(D)J recombination. However, in addition to V(D)J recombination, birds utilize cytidine-deaminase-dependent gene conversion events to further diversify their antibodies, reminiscent of the system that evolved in jawless vertebrates. Again, these similarities suggest that complex adaptive immune systems have developed by co-opting enzymes and mechanisms that existed in ancient organisms (Litman et al., 2010). An overview of immunity throughout evolution is presented in Figure 1.1.

Within the branch of jawed vertebrates it is clear that further mechanisms of diversification evolved, as represented by the fact that B cells have developed mechanisms of secondary diversification that do not exist in T cells. This is likely due to the fact that extensive diversification in T cells, which must recognize antigen in the context of self, may be deleterious due to the loss of recognition of antigen-presenting molecules. B cells, on the other hand, recognize antigen in its

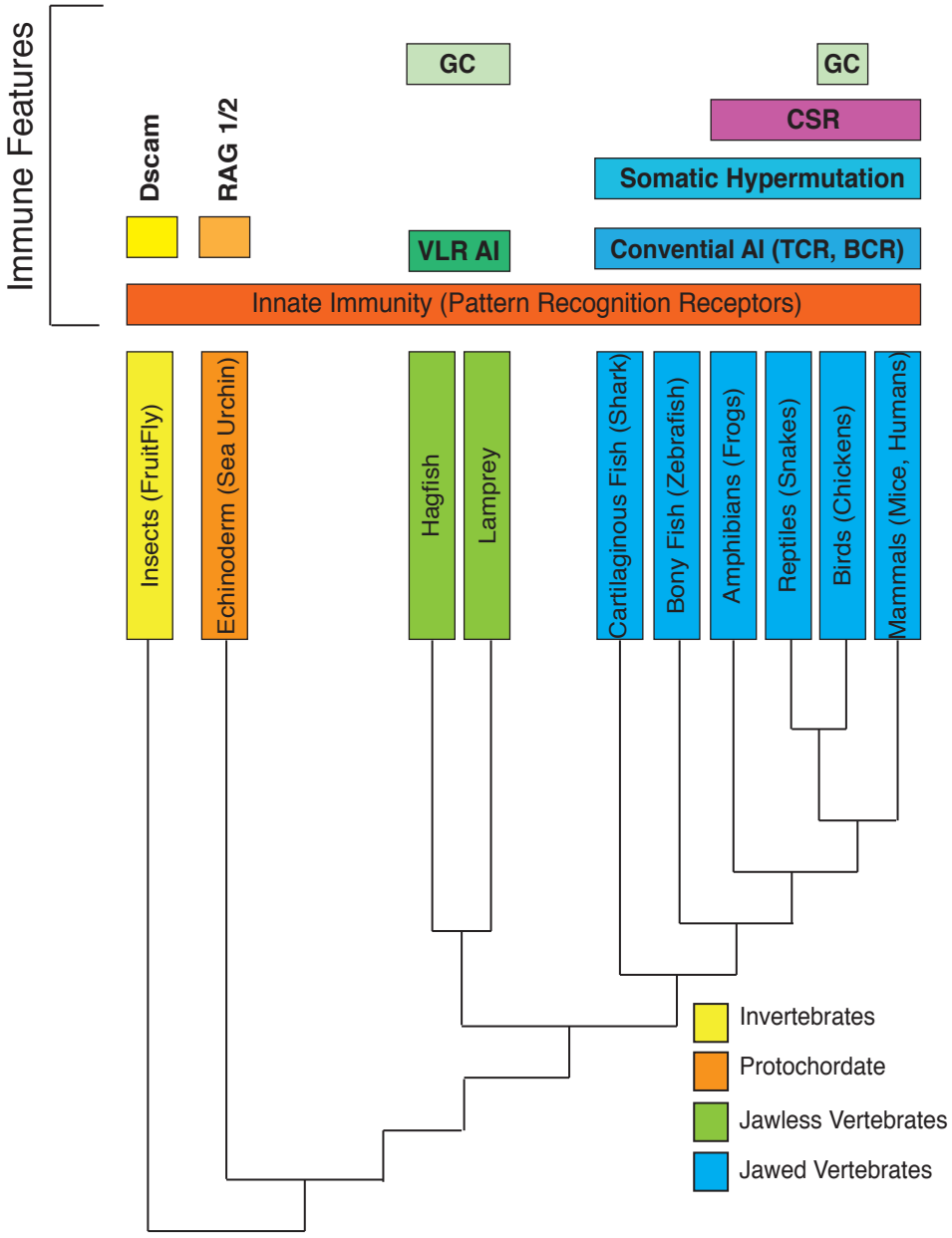


Figure 1.1. An Overview of Immunity throughout Evolution. Select organisms representative of invertebrates, protochordates, jawless vertebrates, and jawed vertebrates are shown in the phylogentic tree. Immune features of each type are depicted above the tree. GC = Gene Conversion,
CSR = Class Switch Recombination, AI = Adaptive Immunity.

native context and thus have much more flexibility in the extent of diversification allowed.

The ability of B cells to diversify their receptors relies on the adoption of mechanisms typically associated with genomic instability, such as double-strand break formation and mutation. While genomic instability is necessary for evolution, when conducted at the level of a single living organism, introduces the possibility of diseases, such as cancer and autoimmunity. Very exquisitely, B cells have been able to harness the power of genomic instability at the immunoglobulin locus, while still protecting genomic stability elsewhere. The mechanisms involved in this complex balancing act are discussed, with a focus on secondary diversification.

1.2 B Lymphocytes and Antibodies

B cells are the mediators of humoral immunity because of their ability to produce both membrane-bound and secreted antigen receptors, known as antibodies. B cell development in mammals begins in the fetal liver and continues in the bone marrow, as a result of differentiation from precursor hematopoietic stem cells. During this development process, the first stage of antibody diversification occurs, producing a mature B cell with a single, non-autoreactive, surface-bound B cell receptor. Once this occurs, the B cell leaves the bone marrow and enters the periphery where it takes up residence in secondary lymphoid organs, such as the spleen and other lymph nodes. It is here that the B cell encounters foreign antigen and is stimulated to undergo secondary

diversification and selected for enhanced affinity for antigen. A schematic of B cell development is presented in Figure 1.2.

Antibodies, themselves, are a complex of approximately 150 kDa, composed of two heavy chain proteins and two light chain proteins. Each of these proteins is first assembled from gene segments contained within the heavy chain immunoglobulin locus (IgH) and either the κ or λ light chain immunoglobulin locus (Igλ or Igκ), respectively. The N-terminus of both the heavy and light chain proteins constitutes the antigen-binding site of the antibody (Fab), the specificity of which can be altered through the variation of amino acid sequence in this region. This portion of the protein is derived from the variable region of the Ig genes. The C-terminus, on the other hand, constitutes the constant domain (Fc) of the antibody and plays no role in antigen-specificity. Rather, the constant domain of the heavy chain decides the isotype of the antibody, which determines the functionality of the antibody by dictating the immune response downstream of antibody-antigen complex formation.

1.3 Antibody Diversification

Over 50 years ago, at a time when the scope of the human genome was not yet known, scientists recognized that it was unlikely that the diversity of antibody generated was encoded in the genome. Most notably, F. MacFarlane Burnet, with no experimental evidence invoked a “randomization” process that would result in the alteration and variation of the coding of immunoglobulin molecules (Burnet, 1976). Prior to his proposal, the only biological precedent for

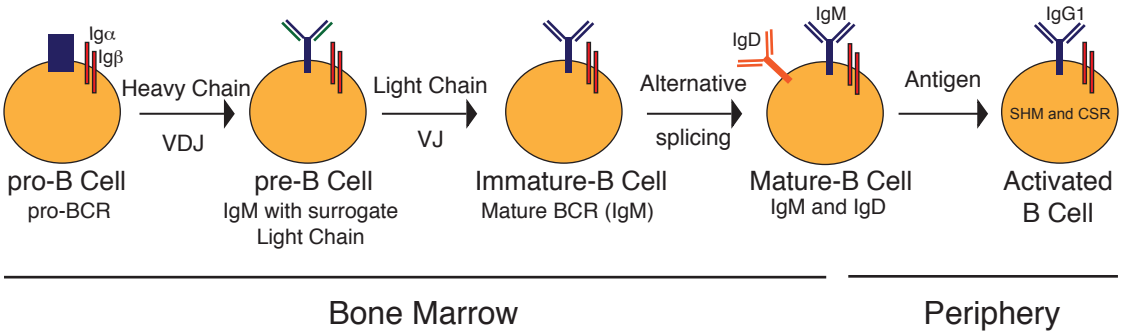


Figure 1.2. An Overview of B-Cell Development. B-Cells develop in the bone marrow. pro-B cells express a pro-BCR composed of glycoprotein and surrogate light chain. It is at this stage that VDJ recombination of the heavy chain locus occurs. The D and J gene segments are rearranged first, followed by the rearrangement of V to D. At the pre-B cell stage, the heavy chain is recombined and expressed as a surface bound IgM. This is complexed with a surrogate light chain. At the pre-B cell stage, VJ rearrangement of the light chain locus occurs. Once V(D)J recombination has occurred on both heavy chain and light chain genes, a mature IgM BCR is expressed on an Immature B-cell. Alternative splicing of the heavy chain locus produces a mature B cell that expresses surface bound IgM and IgD. This mature B cell leaves the bone marrow and enters the periphery where it contacts foreign antigen. Binding of antigen to the mature BCR signals secondary diversification in the germinal centers, including Somatic Hypermutation and Class Switch Recombination. Igα and Igβ are transmembrane proteins that associate with the BCR and generate a signal in response to antigen-BCR complex formation.

such a process was Lederberg's studies on mutation in phage adaptation (LEDERBERG, 1959). The first experimental evidence that such a process occurs was provided by Weigert and Cohn when they showed that immunization altered the amino acid sequence of immunoglobulin lambda light chains, introducing single amino-acid changes (Weigert et al., 1970). Half a decade later, following the advent of recombinant DNA technology, S. Tonegawa showed that, in addition to mutation, somatic gene rearrangement occurs to piece together the immunoglobulin gene from several, variable gene segments (Tonegawa, 1983). These two discoveries began the movement to provide mechanistic insight to Burnet's original hypothesis of randomization (Neuberger, 2008).

Today we have a far better understanding of the mechanisms involved in immunoglobulin diversification. Initial diversification occurs in an antigen-independent fashion through the genomic rearrangement of one of several gene segments, termed variable (V), diversity (D) and joining (J), respectively. This process, known as V(D)J recombination, occurs in both B and T cells.

Encounter of a B cell with antigen initiates secondary diversification reactions at the Immunoglobulin locus. These reactions, Somatic Hypermutation (SHM) and Class Switch Recombination (CSR), provide greater variability within the antigen-binding domain and alter the effector function of the antibody by changing domains within the C-terminus of the protein complex. Each of the methods of diversification share many similarities, however there exist distinct

differences that account for their divergent outcomes. An overview of the three steps of antibody diversification is presented in Figure 1.3.

1.3.1 Primary Diversification—V(D)J Recombination

Initial diversification occurs in an antigen-independent fashion through the genomic rearrangement of one of several V, D and J gene segments. Diversity is attained through the random selection of each gene segment as well as the variability that arises at the junctions upon DNA ligation and repair. The beginning steps toward working out the mechanism of V(D)J recombination began with the discovery of the two essential proteins, Recombinase Activating Gene 1 and 2 (RAG1 and RAG2). The discovery of these enzymes came from a series of well-planned, but also incredibly lucky, experiments conducted in the 1980s in the lab of David Baltimore (Schatz and Baltimore, 2004). Using an artificial V(D)J recombination substrate, it was found that lymphocytes, but not fibroblasts, were able to induce recombination, suggesting the presence of a lymphocyte-specific master regulator (Schatz and Baltimore, 1988). Through transfection and expression of large tracts of genomic sequence in fibroblasts, they were able to rescue V(D)J recombination, leading to the discovery of RAG1 (Schatz et al., 1989). However, it took many more years to discover RAG2, which is co-expressed with RAG1 and necessary for its function (Oettinger et al., 1990).

Since the discovery of the RAG proteins, many of the basic mechanistic details of V(D)J recombination have been worked out. The heavy chain immunoglobulin is composed of V, D and J gene segments, which must be

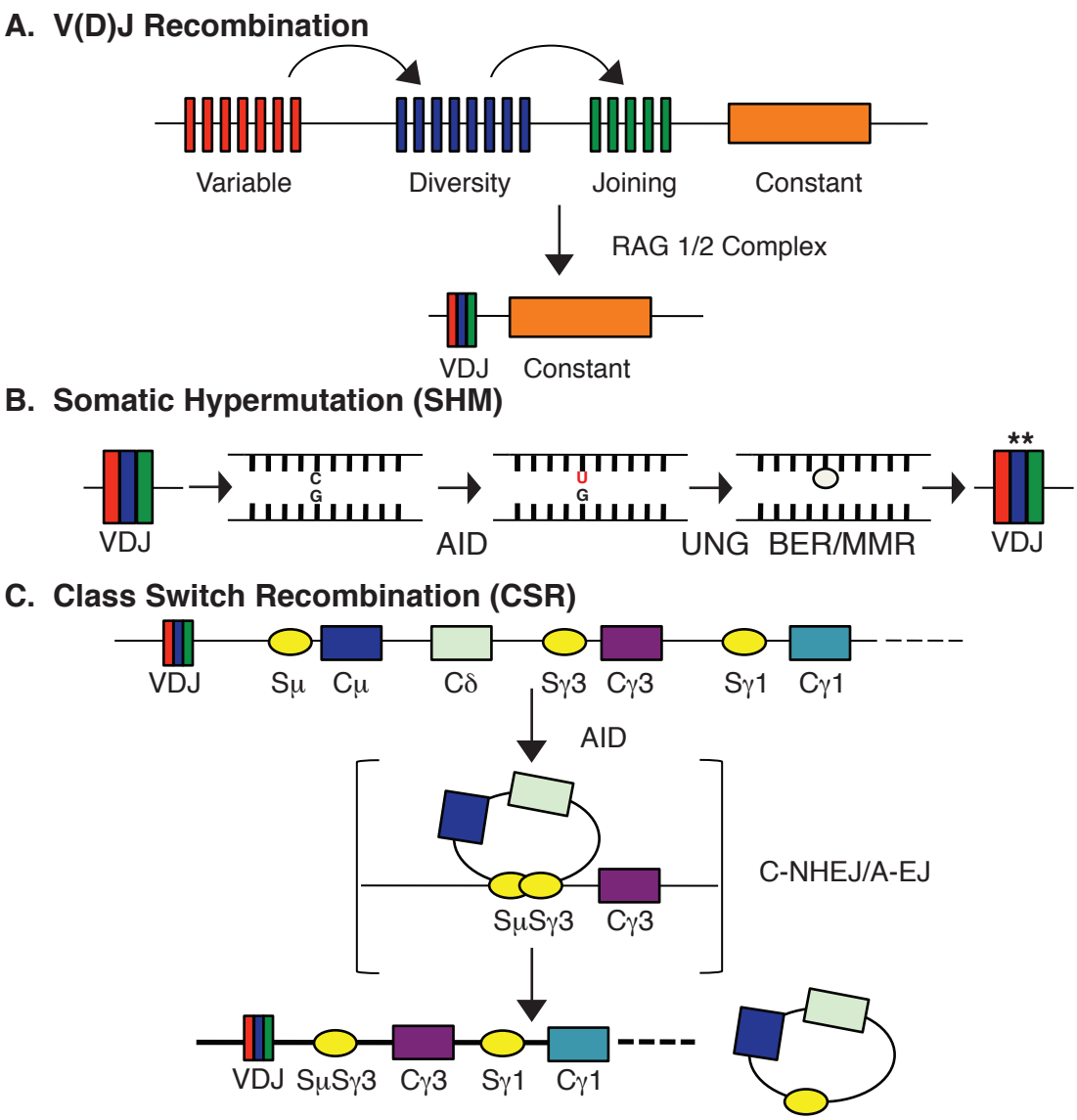


Figure 1.3 An Overview of Antibody Diversification. **A**) V(D)J Recombination is a deletional recombination event between one of many variable, diversity, and joining segments to create the variable region of the immunoglobulin gene. This is catalyzed by the RAG1/2 recombinase complex and occurs in an antigen-independent fashion. **B**) Somatic Hypermutation results in the accumulation of point mutations in the recombined variable region. AID initiates this process through the deamination of cytidine to uridine, followed by removal of the uracil base by Uracil DNA Glycosylase (UNG) and repair by several Base Excision Repair (BER) and MisMatch Repair (MMR) enzymes. * denotes the mutated variable region. **C**) Class Switch Recombination results in the exchange of the default constant region, μ (IgM), for one of many downstream regions. AID initiates this process through deamination of bases in the switch region (yellow circle) upstream of each constant region, resulting in the formation of double strand breaks and repair by non-homologous end joining pathways (NHEJ/A-EJ).

recombined, while the light chain gene recombines only V and J segments. Humans have 65 possible V segments, 27 possible D segments, and 6 possible J segments which can be recombined to produce the heavy chain Ig gene (Market and Papavasiliou, 2003), accounting for much of the diversity in antibody produced. Each of these coding segments is flanked by the presence of very specific sequences known as recombination signal sequences (RSS), which guide recombination. RSS sequences are composed of a stretch of seven highly conserved, palindromic nucleotides (heptamer), and a stretch of nine conserved nucleotides (nonamer) separated by a stretch of either 12 or 23 nucleotides, referred to as 12RSSs and 23RSSs, respectively. The correct recombination of V to D (and not V to V or V to J), as well as D to J (and not D to D or J to J) is helped in part by the fact that 12RSS and 23RSS sequences can only recombine with one another and not themselves, otherwise referred to as the "12/23 rule." V segments are flanked by only 23RSSs, D segments by only 12RSSs and J segments by only 23RSSs. Thus, this excludes the possibility of eliminating the D segment from the full IgH gene (Schatz and Swanson, 2011).

The mechanism of V(D)J recombination can be broken down into two main components: (1) DNA break formation and (2) DNA Repair. The 12 and 23RSSs are brought into close proximity by binding of the RAG proteins, which then induce DNA nicks within each of the RSS's. A trans-esterification reaction between the strands of DNA occurs, resolving the nick by producing a hairpin at the coding end, and leaving the break on the signal sequence end blunt. The

signal sequences will be resolved into a circular piece of DNA (sj), while the hairpins between the coding segments are nicked, and the DNA breaks repaired utilizing proteins from the non-homologous end joining pathway (NHEJ). Additional nucleotides are added before break repair occurs by the non-templated DNA polymerase, TdT, thus providing additional junctional diversity (reviewed in (Alt et al., 2013; Schatz and Swanson, 2011)).

Many more details of this reaction, ranging from epigenetic modifications of the locus (Matheson and Corcoran, 2012), the structure of the RAG1 protein interaction with DNA (Yin et al., 2009), the role RAG2 plays in the recognition of specific histone modifications (Matthews et al., 2007), as well as a number of other findings, have been elucidated. These mechanistic details go beyond the scope of this thesis, but many reviews can be referenced for up-to-date findings (Alt et al., 2013; Nishana and Raghavan, 2012; Schatz and Swanson, 2011). The importance of all of these findings in the context of this thesis, though, is in the recognition of (1) the adaptation of a transposon-like element in evolution to take advantage of genomic instability to produce antibody diversity, and (2) the complexity of the mechanisms developed in order to reign in this genomic instability and protect the genome from extensive damage. Similar themes will be seen in a discussion of the mechanistic details of secondary diversification in B cells.

1.3.2 Secondary Diversification – Somatic Hypermutation and Class-Switch Recombination

Encounter of a mature B cell with antigen initiates secondary diversification reactions at the Ig loci, collectively referred to as the germinal center response. These reactions, Somatic Hypermutation (SHM) and Class Switch Recombination (CSR) share many basic mechanisms but function very differently with respect to antibody diversity. SHM generates greater diversity within the V(D)J-rearranged antigen-binding domain of the antibody and CSR produces antibodies with the capacity to recruit several different types of downstream immune effectors, thus governing the type of immune response. SHM is coupled with a process of selection such that B cells with antibodies that gain affinity toward antigen are selected for survival and B cells with antibodies that have lost affinity toward antigen (or gained affinity for self-derived peptides) are selected for apoptosis. Thus, over time, a population of B cells with increased affinity for antigen is produced. This process is better known as Affinity Maturation.

Because the processes of SHM and CSR seem to have exceedingly different outcomes—it was an astonishing development 10 years ago when AID was identified as a key player in both reactions. In fact, the cytidine deaminase, which enzymatically converts a cytidine base to a uridine (C to U) through a deamination reaction, is essential to initiate both processes. The mechanism of this deamination reaction is depicted in Figure 1.4.

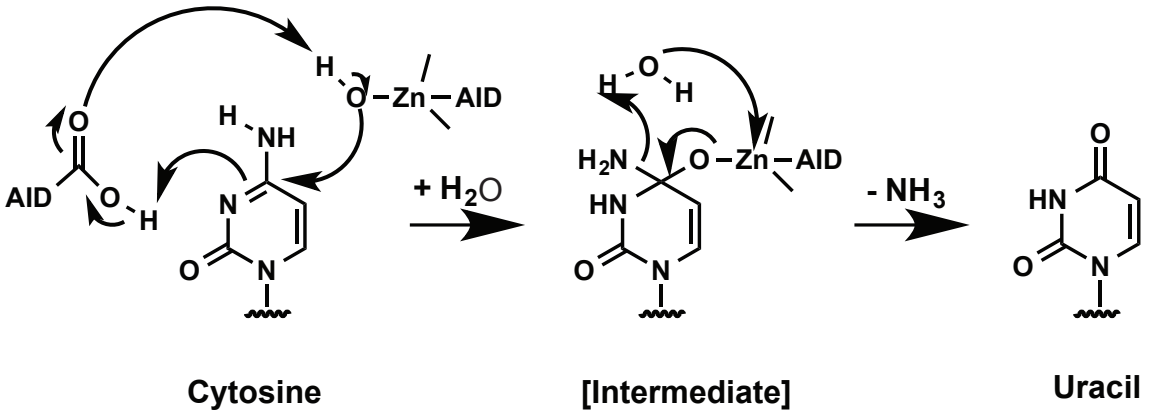


Figure 1.4 Proposed AID-mediated cytidine deamination mechanism.
A proposed mechanism based on bacterial cytidine deaminase enzymes that show homology to AID. A glutamic acid residue (E58), as well as Zn-coordinating residues (H56, C87, C90) play a role in the mechanism.

1.3.2.1 Somatic Hypermutation

During SHM, point mutations accumulate in the rearranged variable region of both the Ig heavy (IgH) and Ig light (Igλ or Igκ) genes. Because this region of the Ig gene makes up the antigen-binding domain of the protein complex, it is these mutations that dictate the change in affinity of the antibody for antigen. There are many characteristic patterns that can be described in terms of the location and types of mutations that occur during SHM. First, mutations begin approximately 50-100 nucleotides downstream of the IgVariable promoter region and extend another 1,000 – 2,000 nucleotides (Lebecque and Gearhart, 1990; Rada and Milstein, 2001). This suggests that the 5' boundary of SHM is the promoter, inferring a connection between mutation and transcription of the locus. In fact, many groups have observed that mutation frequency is intimately linked with the level of transcription through the region (Longerich et al., 2005). Movement of the promoter in front of non-mutated regions induces mutation (Peters and Storb, 1996), deletion of the promoter drastically reduces mutation (Fukita et al., 1998), and regulation of transcriptional activity by use of the Tet operator results in mutation frequencies that are well-correlated with the level of transcription (Bachl et al., 2001).

It is likely that the role of transcription during SHM is highly nuanced. On one level, transcription is necessary to open the DNA duplex, revealing a ssDNA substrate for AID deamination. However, more recent findings suggest that the RNA molecules, themselves, or processing of the RNA play an important role.

First, in addition to sense transcription, anti-sense transcription occurs through the variable coding exons (Perlot et al., 2008). The function of these anti-sense transcripts is still unknown. Further, the RNA Exosome, an RNA processing and degradation complex, has been shown to enhance AID deamination on both the template and non-template strands of DNA—a phenomenon seen *in vivo* (Milstein et al., 1998), but not *in vitro* (Basu et al., 2011). Whether the Exosome degrades the nascent sense transcript, revealing the template strand of DNA, or whether it is involved in the processing of other ncRNAs (either sense or antisense), remains to be seen. However, these findings emphasize the fact that the role of transcription during SHM likely goes beyond the initial unwinding of the locus.

Despite the ability of AID to deaminate only base C, mutations can be found occurring at all four nucleotides. At C:G base pairs, transition mutations are predominant over transversion mutations, implicating direct replication over the mismatched U:G as a repair mechanism. C:G mutations occur in a strand non-specific manner, whereas A:T mutations show a strand bias (Mayorov et al., 2005a). And lastly, though this is not always the case, often deaminated cytidine bases are found in the context of the sequence WRCY, or its complement, RGYW (W = A/T, R = A/G, Y = C/T) (Rogozin and Kolchanov, 1992). These observed characteristics have been instrumental in elucidating the factors and mechanisms involved in the resolution of AID-mediated deamination events to produce point mutations at the variable region of the Ig locus.

Two main repair pathways have been implicated in the SHM reaction: the Base Excision Repair pathway (BER) and the Mismatch Repair pathway (MMR). An overview of these pathways and their involvement in SHM is presented in Figure 1.5. While these pathways are normally used for faithful and error-free repair, they have been co-opted to introduce mutations through the recruitment of error-prone repair proteins.

The BER pathway functions mainly at the deaminated C. Uracil Deglycosylase (UNG) recognizes the mismatched U:G basepair and excises the U to generate an abasic site (Li et al., 2012; Rada et al., 2002b). The endonuclease, APE1, and the Mre11-Rad50-Nbs1 (MRN) complex have been implicated in nicking the DNA to generate a single-strand break (SSB) (Larson et al., 2005; Masani et al., 2013; Yabuki et al., 2005). The formation of the SSB provides DNA polymerases access to the DNA. During error-free repair, the high fidelity polymerase, pol β, is recruited; however, during SHM, error-prone polymerases are recruited. The deoxycytidyl transferase, Rev1, which forms a complex with the DNA polymerase ζ, specifically incorporates cytidine bases opposite uridine bases or abasic sites. Rev1 has specifically been found to play a role in SHM, particularly in the generation of transversion mutations at the deaminated C base (Jansen et al., 2006; Simpson and Sale, 2003). A similar loss of C and G transversions is seen in mice, which lack the ability to ubiquitinate the polymerase processivity factor, PCNA. Mice that contain either a lysine-to-arginine mutation at lysine 164 or that lack the E3 ubiquitin ligase, Rad18, which

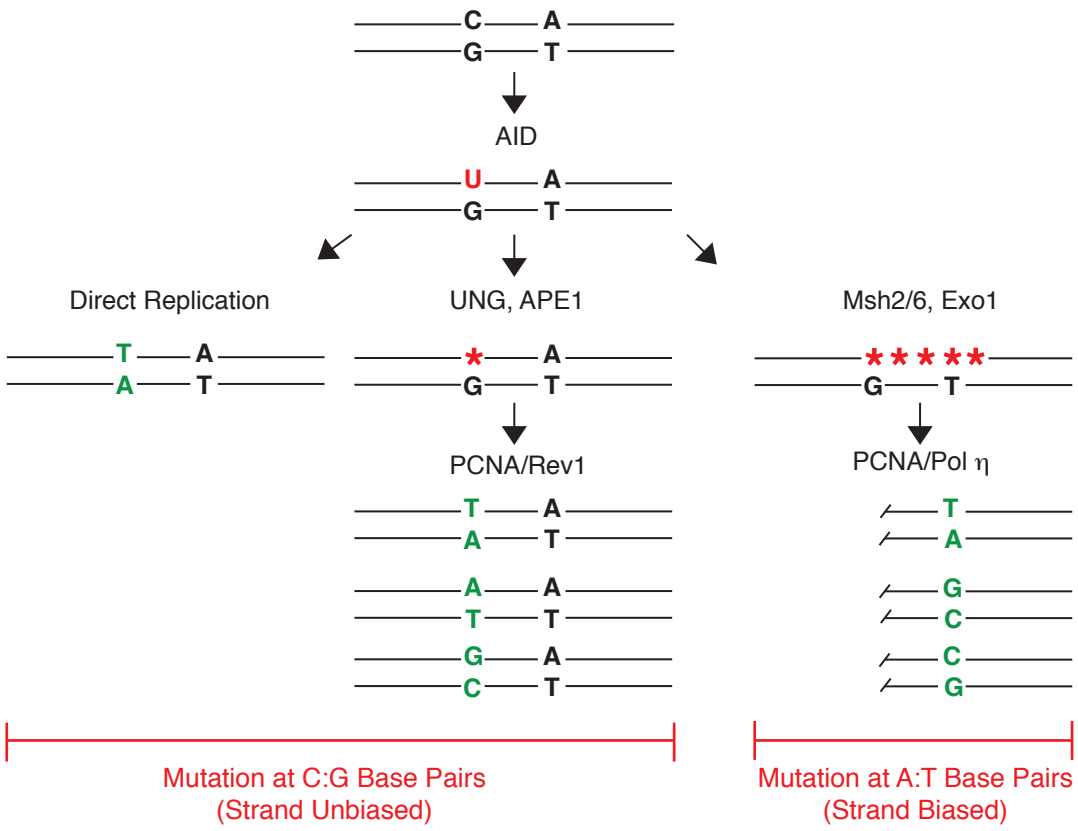


Figure 1.5. Mechanisms of AID-mediated mutation during SHM. Initial AID deamination events are processed by various repair pathways to result in the accumulation of point mutations in the variable region of the Ig locus. Direct replication over a U:G mistmatch results in a transition mutation. The involvement of Base Excision Repair enzymes results in all possible mutations at the deaminated C. Uracil Deglycosylase (UNG) excises the mismatched U, resulting in an abasic site. The endonuclease, APE1, generates a SSB. The DNA-encircling processivity factor, PCNA, recruits the Rev1-containing error prone polymerase complex to generate mutations at the deaminated C. The involvement of Mismatch Repair enzymes results in the mutation of A:T base pairs in proximity to the deaminated C. Recognition of the mismatch by the Msh2/6 complex and resection of DNA by the exonuclease, Exo1, results in a ssDNA gap. This gap is filled in by the error prone polymerase, pol η, which is again recruited by PCNA. This results in all possible mutations at A:T basepairs. Mutations at C:G base pairs tend to be strand unbiased, suggesting that AID accesses both the non-template and template strands of DNA equally. However, mutations at the A:T base pairs are strand biased and show greater mutation of A bases on the non-template strand. This is likely due to the lack of access of error prone repair proteins to the template strand. The star represents an abasic site.

ubiquitinates PCNA, have altered SHM mutation patterns (Arakawa et al., 2006; Bachl et al., 2006; Langerak et al., 2009; Roa et al., 2008). Thus, ubiquitinated PCNA plays a role in dictating whether the BER repair pathway will be error-free or error-prone.

In a parallel, non-competitive fashion, the MMR pathway serves the purpose of generating mutations at basepairs adjacent to the deaminated C. The complex of Msh2-Msh6 recognizes the U:G mismatch and, through a currently unknown mechanism, a SSB is generated (Krijger et al., 2009; Li et al., 2006; Martomo et al., 2004; Roa et al., 2010). During MMR, this SSB allows the exonuclease, ExoI, to enter the DNA and resect the DNA neighboring to the mismatched basepair (Bardwell et al., 2004), creating a single-stranded gap. Just as was seen for the BER mechanism, either error-free or error-prone polymerases can be recruited to fill in the gap. Again, ubiquitinated PCNA recruits the error-prone polymerase, pol η, which erroneously polymerizes frequently opposite of templated T and A bases (Delbos et al., 2005; Zeng et al., 2001; 2004). Thus, this accounts for the formation of mutations at A:T basepairs that are not directly the result of AID deamination. In fact, mice that lack both Msh2 and pol η produce a mutation pattern that completely lacks mutations at A:T basepairs (Delbos et al., 2007).

The combined effort of (1) direct replication over the U:G mismatch, (2) the error-prone BER pathway and (3) the error-prone MMR pathway can account for all mutations observed during SHM. However, many mechanistic details, such

as how error-prone repair is targeted and restricted to the Ig locus, remain unknown. In fact, experiments using *ung*$^{-/-}$ mice suggest that there is a direct competition between error-free and error-prone repair during SHM. The mutation pattern seen in *ung*$^{-/-}$ mice, which cannot excise the mismatched U base, shows a loss of transversion mutations and an increase in transition mutations at C:G basepairs. This suggests that, in the absence of UNG, transition mutations are generated through replication over the mismatched base pair; however, when UNG is present and the U excised, error-free BER occurs to faithfully correct at least some of the deaminated C bases (Rada et al., 2002b). Another perplexing observation is that mutations at C:G basepairs are strand-unbiased whereas mutations at A:T basepairs are strand biased. This suggests that, while AID-mediated deamination is targeted to both strands, the subsequent repair processes are not. In fact, both pol η and the Msh2-Msh6 complex appear to be biased toward the non-template strand, thus accounting for the bias at A:T basepairs, but not C:G basepairs (Mayorov et al., 2005a; Unniraman and Schatz, 2007). How this specific targeting occurs is unclear. It is possible that the non-transcribed strand, which is not associated with the nascent pre-mRNA, is more exposed to these protein complexes. However other mechanisms, such as the involvement of the RNA Exosome, may be at play here.

Just as was seen during the discussion of V(D)J recombination, mechanisms that promote genomic instability are co-opted to generate diversity at the Ig loci. During SHM, the introduction of point mutations occurs at the

variable regions of the IgH and IgL genes. Introduction of mutations outside of the Ig loci, though, would be detrimental to the organism. It has been shown in mice deficient in UNG and Msh2 that AID-mediated mutation can be detected outside of the Ig locus, however not throughout the entire genome. Thus, it appears that at least two levels of regulation occur—one that targets AID to a subset of the genome, and the second, which targets error-free versus error-prone repair enzymes (Liu et al., 2008). Thus, a complex set of mechanisms, which are now only being revealed, has evolved to allow genomic instability to exist along side genomic stability within a single cell—quite an amazing accomplishment.

1.3.2.2 Class Switch Recombination

Class-Switch Recombination (CSR) is a very unique form of antibody diversification because rather than altering the antigen-binding capabilities of the receptor, it alters the function of the receptor within the immune response. The IgM isotype is the most ancient form of antibody and defines the B cell lineage. Its presence emerged in evolution beginning in cartilaginous fish and can be found in all jawed vertebrates. While primitive vertebrates appear to have a small variety of other isotypes of antibodies, it was not until the emergence of amphibians that CSR evolved, allowing single cells to switch between different isotypes and thus different effector functions as needed (Flajnik and Kasahara, 2010). Besides IgM and IgD (which is a spliced isoform of the antibody), mammals have three main isotypes of antibody (with additional subsets): IgG,

which is involved in high affinity memory responses and makes up the bulk of the antibody response to pathogens, IgE, which functions mainly in the inflammatory and allergy responses, and IgA, which is found at mucosal surfaces (Flajnik and Kasahara, 2010). The C-terminal region of the Ig heavy chain gene and protein defines each of these isotypes.

Just like SHM, CSR is initiated by, and absolutely dependent upon, AID-mediated deamination of deoxycytidine residues in the Ig locus. However, rather than occurring in the variable region, deamination occurs in highly repetitive "switch" regions (S) upstream of each of the constant region gene segments (Figure 1.3). For example, Sμ lies upstream of Cμ, which encodes the constant region for IgM and Sγ1 lies upstream of Cγ1, which encodes the constant region for IgG1. S regions range in size from 1-12 kilobases and contain a highly G-rich non-template strand sequence (Chaudhuri and Alt, 2004; Dunnick et al., 1993). Though mutations do accumulate within the S regions, the purpose of AID-mediated deamination during CSR is to generate DSBs within two switch regions (Sμ and a downstream S region), which can then be recombined. This functions to delete the IgM constant region and replace it with a downstream constant region of a different isotype. The excised DNA is ligated to form a circular piece of DNA (Iwasato et al., 1990) (Figure 1.3). Thus, while the initiating step is the same between SHM and CSR, the outcomes are quite different. This difference is largely due to how the cell resolves the mutations generated.

Similar to SHM, CSR requires transcription through the locus. Specifically, transcription initiates at intervening (I) exons upstream of each of the switch regions and continues through each of the constant region exons producing a non-coding, “sterile”, transcript known as the germline transcript (GLT). Transcription occurs constitutively at the μ locus and in a stimulus- and switch-dependent fashion at each of the other constant loci (Lee et al., 2001; Stavnezer, 1996; Stavnezer-Nordgren and Sirlin, 1986). Transcription from I promoters is regulated by the distal 3’IgH regulatory region (RR), as deletions in this region result in decreased transcription and defective CSR (Pinaud et al., 2001; 2011; Vincent-Fabert et al., 2010). Again, transcription has been proposed to open the locus, revealing a ssDNA substrate for AID activity. Because of the highly repetitive, C/G rich sequence of S regions, transcription through this region is capable of producing stable RNA:DNA hybrids known as R-loops, exposing the G-rich non-template strand for AID activity (Chaudhuri et al., 2003; Daniels and Lieber, 1995; Yu et al., 2003). Experiments have demonstrated that the presence of a G-rich non-template strand aids CSR (Shinkura et al., 2003), however other studies suggest that R-loop formation may not be the only mechanism at play here. Most notably, replacement of the murine G-rich S region with the A/T rich region from *xenopus laevis*, which cannot form R loops, was still able to support CSR (Tashiro et al., 2001; Zarrin et al., 2004). This data suggests that either the formation of stem-loop structures at palindromic sequences or the presence of the conserved sequence, 5’-AGCT-3’, is sufficient to target AID.

In addition to making the locus more accessible to AID, experimental evidence suggests that the RNA molecule, itself, is participating in the mechanism of CSR. It is well-established that transcription is not sufficient for CSR; rather, splicing of the sterile RNA is necessary (Hein et al., 1998; Lorenz et al., 1995). The association of AID with splice-associated protein factors such as PTBP2 and CTNNBL1 further suggests the importance of RNA splicing for CSR (Conticello et al., 2008; Nowak et al., 2011). It remains unclear at present if it is the act of splicing or the product of splicing, the RNA molecule, which is important for CSR. However, just as in the variable region, high levels of sense and antisense transcription occur at the constant regions to produce a number of non-coding RNAs (Perlot et al., 2008).

Recent work has focused on understanding the potential role of these Ig-associated non-coding RNAs in the regulation of CSR. These RNA molecules could act as a guide RNAs, targeting AID to the correct region of the Ig locus (unpublished data, Chaudhuri Lab), or even recruit necessary DNA repair factors, as has been seen in DSB repair mechanisms in other systems (Wei et al., 2012). The discovery that the RNA exosome is required for efficient CSR further stresses the importance of RNA and RNA processing for CSR (Basu et al., 2011).

As a side note, transcription generated at Iμ occurs after recombination, generating a post-switch transcript, which contains Iμ spliced to the constant region exons of the switched isotype (Li et al., 1994). While this is currently

thought to be a by-product of CSR, possible functions for this transcript have simply not been investigated.

Similar repair pathways used for SHM are implicated in the formation of double strand breaks (DSB) downstream of AID-mediated deamination at S regions. Mismatched deoxyuridine bases are excised by UNG to generate an abasic site, allowing APE1 to generate a SSB (Masani et al., 2013; Rada et al., 2002b). SSBs sufficiently close to one another on opposite strands are easily processed into DSBs, generating the substrate for recombination. Thus, in the core of the S region where there is a high density of 5'-AGCT-3' deamination hotspots, AID-mediated deamination and SSB break formation is likely to occur at high rates (Min et al., 2003). It should be noted that, though UNG is clearly necessary for CSR, evidence exists to raise questions regarding its specific function. Reconstitution of *ung*$^{-/-}$ B cells with a catalytically inactive form of UNG restores CSR. In addition, inhibition of the activity of UNG with a peptide inhibitor does not affect CSR levels (Begum et al., 2009). Thus, it is possible that UNG plays a non-canonical role during CSR, acting as a scaffold for the recruitment of downstream DSB repair factors. It is also possible that UNG plays a dual role and, in its absence, the generation of DNA breaks is compensated for by the presence of the MMR repair pathway proteins.

In collaboration with the BER pathway, the MMR pathway also plays a role in generating DSBs in S regions during CSR. Recognition of a mismatched basepair by the Msh2-Msh6 complex leads to the recruitment of Mlh1 and Pms2,

which in turn recruit Exo1 to resect a patch of DNA neighboring to the original mismatched basepair. Mice that contain deficiencies in each of these proteins have severely defective CSR (Bardwell et al., 2004; Ehrenstein and Neuberger, 1999; Ehrenstein et al., 2001; Martin et al., 2003; Martomo et al., 2004; Schrader et al., 1999; 2007; Stavnezer and Schrader, 2006). The importance of both pathways for CSR is demonstrated by the complete loss of CSR in Msh2/UNG double knockout mice (Xue et al., 2006). In these mice, Uracil insertion into DNA is resolved via replication to produce characteristic C-to-T mutations. It is likely that in physiological settings, the BER pathway functions within the core of the S regions where a high density of SSBs form, whereas the MMR pathway is used to process DNA to generate DSBs from non-proximal SSBs. This likely occurs in the extremities of the S region where AID-hotspots are less frequent. This model is supported by two key pieces of data. First, in the absence of Msh2, the Sμ tandem repeat core region is necessary for CSR to occur; second, CSR events to the isotype, IgG2b, which contains a switch region least speckled with AID-hotspots, is most affected by the loss of mismatch repair enzymes (Min et al., 2003; Schrader et al., 1999).

Even though it has been shown that the formation of two DSBs is sufficient to induce a recombination event by engineering ISce-I restriction sites in place of switch regions, these recombination events occur at a rate much lower than wildtype (Zarrin et al., 2007). Thus, it appears that the efficiency of CSR depends on the formation of an excess of deamination events, mutations, SSBs and DSBs

within the S regions. In fact, it has been shown that removal of the faithful repair polymerase, Pol β, actually results in an increase in mutation and DSB formation at the S regions, resulting in slightly increased levels of CSR (Wu and Stavnezer, 2007). Again, there is a battle between error-free and error-prone repair at the Ig locus. Some deamination events are correctly repaired; however, because the number of mutations exceeds the capacity of error-free repair, error-prone repair gains access and DSBs form.

The last step of CSR requires the resolution of DSBs through long-range recombination of breaks formed in two S regions. This occurs primarily through the use of the classical non-homologous-end-joining (C-NHEJ) pathway resulting in blunt DSB junctions, and through the alternative-end-joining pathway (A-EJ), which results in break joints with greater homology. Often the use of the A-EJ pathway is revealed in the absence of a critical NHEJ protein factor.

Initially, the DSB is sensed and marked by a variety of protein factors. The Mre11-Rad50-Nbs1 (MRN) complex recognizes the presence of DSB and activates the kinase, ATM (Dinkelmann et al., 2009). However, because the severity of CSR defect is greater in MRN-deficient mice as compared to ATM-deficient mice, it is likely that the MRN complex plays additional roles (Dinkelmann et al., 2009; Lumsden et al., 2004). In fact, the MRN complex plays a role in funneling DSBs into both the NHEJ and A-EJ pathways for repair. Phosphorylation of ATM initiates the ATM-dependent DNA damage response, which in turn induces the phosphorylation of a number of proteins essential for

DSB repair including the histone variant, H2AX (to form phospho-H2AX or γ-H2AX) and 53BP1. These proteins, together, form foci that span over 100 Kb around the DSB and are essential for CSR (Reina San-Martin et al., 2003; Ward et al., 2004).

During C-NHEJ, the complex Ku70/Ku80 binds broken DNA ends and recruits XRCC4 and Ligase4 (Lig4) to promote end-joining. The kinases, DNA-PKcs are also recruited, which activate DNA end processing by the endonuclease, Artemis. Deficiencies in each of these factors results in defects in CSR to varying degrees, typically ranging from about 20%-40% of wildtype levels (Boboila et al., 2010; Han and Yu, 2008; Li et al., 2008; Rivera-Munoz and Soulas-Sprauel, 2009; Soulas-Sprauel et al., 2007). However, the fact that CSR is not completely abolished in the absence of necessary NHEJ protein factors suggests the existence of a parallel pathway. In addition, analysis of break joints in the absence of the C-NHEJ pathway reveals that a greater level of microhomology is used in repair.

The A-EJ pathway is less well understood, but involves additional end-processing of DNA by the MRN complex and the DNA end processing factor, CtIP. Both factors promote a more microhomology-mediated form of DNA repair (Lee-Theilen et al., 2011). In addition, the DNA damage sensor, Parp1, has been implicated in this pathway (Robert et al., 2009); however, the bulk of the factors involved, including the necessary DNA ligase, remain unknown (Boboila et al., 2012). It also remains unclear what factors dictate the choice of pathway. It has

been shown that 53BP1 in complex with its binding partner, Rif1, prevents DNA end-resection, thereby promoting C-NHEJ over A-EJ (Bothmer et al., 2010; 2011; Di Virgilio et al., 2013); however, additional mechanisms are likely to be revealed in the coming years. A summary of the mechanisms used to resolve DSBs during CSR are presented in Figure 1.6 (adapted from (Xu et al., 2012).

1.4 Activation Induced Cytidine Deaminase (AID)

Much of SHM and CSR can be carried out by ubiquitous DNA repair factors. In fact, the only known B-cell specific factor is AID. Because of the potential dangers of AID activity, including off-target mutations and the formation of deleterious translocations, great interest has focused on how AID is regulated to specifically induce genomic instability at the Ig locus, protecting the rest of the genome from harm.

1.4.1 The discovery of AID

The discovery of AID and the elucidation of its mechanism were greatly facilitated by the formation of the B lymphocyte cell line, CH12F3 in 1996 by the Honjo laboratory. Derived from the CH12.LX lymphoma cell line, CH12F3s were selected to undergo class switch recombination exclusively, and at high frequency, to the isotype IgA upon stimulation with IL-4, TGFβ, and CD40L. Despite the presence of other B cell lines available at the time, the CH12F3 line provided a robust inducible system in which to study the mechanics of class switch recombination (Nakamura et al., 1996).

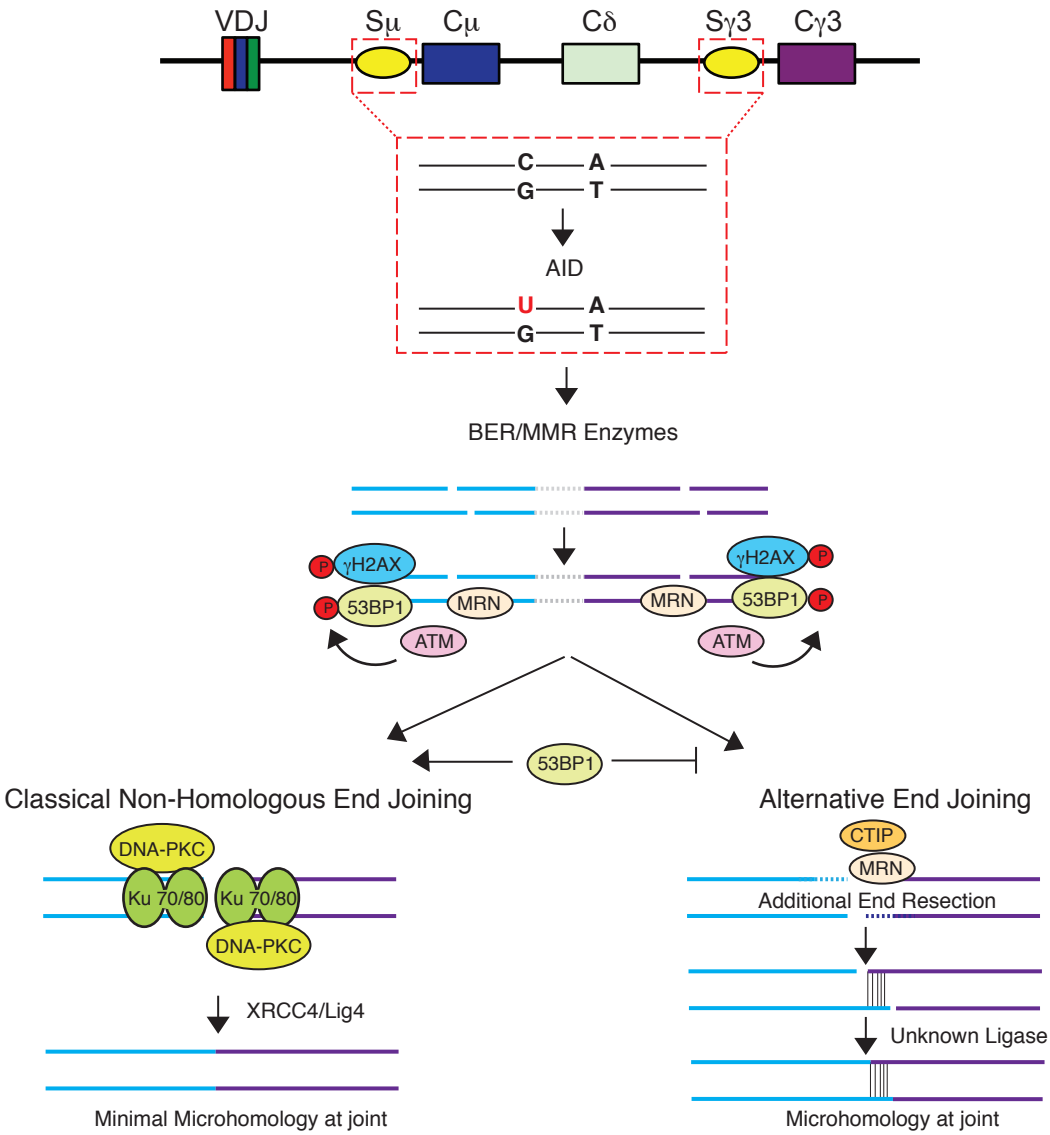


Figure 1.6 Overview of CSR Repair Mechanisms. Within two switch regions (yellow circles, blue/purple lines), AID deamination coupled with DNA processing by BER and MMR enzymes results in the formation of DSBs. Recognition of DSBs by the MRN complex (Mre11, Rad50, Nbs1) activates the ATM-dependent repair pathway, resulting in the phosphorylation of 53BP1 and H2AX. Breaks are repaired through two pathways: Classical non-homologous end joining (C-NHEJ) and alternative end-joining (A-EJ). The proteins known to take part in each pathway are depicted. C-NHEJ results in blunt joints or joints with minimal microhomology. A-EJ results in break joints with larger stretches of homology. 53BP1 has been shown to prevent excess resection, thus preventing A-EJ.

Predicting that a specific recombinase was responsible for CSR, Muramatsu and Honjo applied a PCR-based subtraction method to screen genes upregulated upon stimulation of CH12F3 cells for CSR. Among the four novel genes discovered, Activation-Induced-Cytidine-Deaminase (AID) proved to be interesting because of its (1) germinal center B cell restriction, (2) homology to the APOBEC family of RNA cytidine deaminases and (3) *in vitro* deaminase activity, unique from that of APOBEC1 (Muramatsu et al., 1999).

Confirmation of the necessity of AID in somatic hypermutation and class switch recombination came with the generation of an *aicda*$^{-/-}$ animal by Muramatsu *et al.* in 2000. Though no gross abnormalities were observed, these animals suffered a striking inability to undergo CSR (upon stimulation *in vitro* and antigenic challenge *in vivo)*, and surprisingly, also SHM (*in vivo*) (Muramatsu et al., 2000).

Concurrent with the mouse work conducted in the laboratory of T. Honjo, Anne Durandy was studying human hyper-IgM syndrome (HIGM) patients who completely lack class switch recombination and over-produce IgM. At the time, the known cause of HIGM was X-linked, resulting in a deficiency in CD40 signaling (Reviewed in (Durandy et al., 2004)). However, the Durandy lab also identified groups of patients with an autosomal form of hyper-IgM syndrome and using standard human genetics identified AID as the gene responsible for a subset of these patients (henceforth known as type II, or HIGM2 patients). Interestingly, these patients also lacked somatically mutated immunoglobulin

genes (Revy et al., 2000). The Honjo and Durandy papers were published back-to-back and together provided incontrovertible evidence that AID plays a central and initiating role in both CSR and SHM.

1.4.2 AID is a DNA mutator

Initial characterization of AID revealed that it possessed a cytidine deaminase domain with homology to several unknown genes and also to the known RNA deaminase, Apolipoprotein-B mRNA editing catalytic polypeptide –1 (APOBEC1) (Muramatsu et al., 1999). Due to the homology of AID to the mRNA editing enzyme, APOBEC1, it was initially hypothesized that AID also edited mRNA. In order to account for the involvement of AID in both SHM and CSR, it was imagined that AID targets either (1) distinct mRNAS, a DNA mutator in the context of SHM and a region-specific recombinase/nuclease in the context of CSR, or (2) a single mRNA that functions in both SHM and CSR, facilitated by task-specific targeting co-factors, similar to ACF, the targeting co-factor for ApoB editing (Mehta et al., 2000; Muramatsu et al., 2000).

Alternatively, it was proposed that AID may edit DNA directly, providing substance to earlier theories that postulated the existence of a mutating factor that is directly targeted to the Ig locus (Neuberger et al., 1998; Peters and Storb, 1996; Wiesendanger et al., 1998). Genetic evidence for this scenario was provided by a number of studies that demonstrated that ectopic expression of AID was able to mutate mammalian cells, as well as bacteria and yeast (Martin and Scharff, 2002; Mayorov et al., 2005b; Petersen-Mahrt et al., 2002; Yoshikawa et al., 2002). As it is unlikely that AID would edit the same mRNA in

all of these cells to result in the generation of a novel DNA mutator, the simplest interpretation of these data is that AID is a DNA mutator, and as such, the first member of a family of polynucleotide deaminases that acts on host DNA. In support of these cellular studies, *in vitro* studies have confirmed that AID exhibits activity on DNA substrates, but not RNA substrates, despite its ability to bind both DNA and RNA (Bransteitter et al., 2003; Chaudhuri et al., 2003; Dickerson et al., 2003; Nonaka et al., 2009) Several studies have raised questions regarding the DNA editing hypothesis including evidence that AID can edit the RNA of the hepatitis B virus, catalytically dead mutants of UNG can rescue CSR, as well as correlative studies suggesting that AID downregulates Topoisomerase I through an RNA-editing mechanism, leading to DSB formation during CSR (Begum et al., 2009; Kobayashi et al., 2009; Liang et al., 2013). Despite these studies, though, the preponderance of evidence suggests that AID directly targets DNA. Not only has AID been observed to be localized at the Ig locus using chromatin immunoprecipitation (ChIP), but, through a number of indirect assays, the existence of uracils in DNA has been shown to occur in an AID-dependent manner (Maul et al., 2011; Pavri et al., 2010). Lastly, a comprehensive study using RNA-seq of both small RNAs and poly-A+ mRNAs from stimulated B cells shows no evidence of RNA editing upon stimulation for CSR (Fritz et al., 2013). Thus, proof for the alternative RNA-deamination model will require the discovery of an AID-mutated RNA molecule.

1.4.3 The regulation of AID

Regardless of the controversy concerning the molecular mode of action of AID, the mutagenic strength of AID and thus its threat to genomic stability is not disputed. As such, it is expected that AID is tightly regulated at multiple levels within the cell. In fact, several different modes of regulation have begun to be uncovered, ranging from transcription of AID through regulation of the AID protein. The various methods of AID regulation are presented below and summarized in Figure 1.8 (adapted from (Xu et al., 2012) and (Alt et al., 2013)).

1.4.3.1 Transcriptional control of AID expression

Because of the mutagenic potential of AID, its expression must be tightly regulated. AID transcript is undetectable in naïve B cells, but greatly up-regulated in response to stimulation. AID expression is induced by classic CSR stimuli, such as lipopolysaccharide, CD40-receptor engagement, interleukin-4 (IL-4), and TGFβ, as well as IgM-CD19-CD21 crosslinking in the context of SHM (Dedeoglu et al., 2004; Faili et al., 2002; He et al., 2004; Pone et al., 2012; Rawlings et al., 2012). These stimuli function to up-regulate both the canonical and non-canonical NF-κB pathways, which work together to induce and maintain AID expression (Xu et al., 2012).

In addition to NF-κB, several activating and inhibitory factors have been identified that help to regulate proper AID expression. These proteins have been found to bind in four main cis-regulatory elements throughout the AID locus: Region I lies directly upstream of the promoter, Region II lies downstream of the promoter in the first intron, Region III lies 17 kb downstream of the promoter, and

Region IV lies 9 kb upstream of the promoter (Crouch et al., 2007; Gonda et al., 2003; Xu et al., 2012; Yadav et al., 2006). The activating factors, including HOXC4, SMAD3, SMAD4, PAX5, E2A proteins, and BATF bind throughout these four regions and cooperate with NF-κB at the promoter to enhance transcription of AID (Betz et al., 2010; Ise et al., 2011; Park et al., 2009; Sayegh et al., 2003; Tran et al., 2009). To keep these activating forces in check, Region II also contains binding sites for inhibitory factors, such as MYB and E2F (Tran et al., 2009). These factors are important for maintaining AID in the off-state in naïve B cells (Xu et al., 2012).

Expression of AID has been documented in other cell types, though at much lower levels than in germinal center B cells. These cell types include oocytes, primordial germ cells (PGCs), ES cells, breast tissue, and prostate epithelial cells (Fritz and Papavasiliou, 2010; Lin et al., 2009; Morgan et al., 2004; Pauklin et al., 2009). Because of the danger of AID expression, it has been suggested that AID has a separate function, particularly in DNA demethylation, in these cell types. While good evidence exists to suggest that this is the case during zebrafish development, future studies are necessary to determine if AID functions during mammalian development (Abdouni et al., 2013; Rai et al., 2008). In addition, thorough DNA methylation studies have been conducted in murine splenic B cells and it does not appear that AID plays a role in DNA methylation status in this context (Fritz et al., 2013).

In addition, AID expression can be induced in conditions associated with cellular transformation (Gourzi et al., 2006) and has been implicated in the generation of translocations common in B cell lymphomas (Klein et al., 2011; Klein and Dalla-Favera, 2008; Robbiani et al., 2008; 2009).

1.4.3.2 Post-transcriptional control of AID expression

miR-155, a miRNA previously shown to play a role in the proper activation of B lymphocytes, directly regulates AID levels in response to activating stimuli (Dorsett et al., 2008; Teng et al., 2008). Disruption of the miR-155 target site in the 3' UTR of the AID locus produces an up-regulation of AID protein levels upon activation of B cells; this results in increased levels of CSR. Furthermore, AID expression in miR-155 mutant mice was found to be temporally deregulated, as shown by AID expression in peripheral B cells. This suggests that miR-155 plays a role in switching off AID expression in post-GC B cells (Teng et al., 2008).

Furthermore, both reports noted that in the absence of miR-155 control excess AID protein led to mutation of off-target sites. For example, the gene *bcl6* was found mutated at a rate three times higher than in WT mice; importantly, the occurrence of cMyc-IgH translocations, a known side effect of aberrant AID activity, was increased 15-fold (Dorsett et al., 2008; Teng et al., 2008). Thus, via the regulation of AID protein levels, miR155 acts as a tumor suppressor by silencing potentially harmful mutations and translocations.

In addition to miR-155, Teng *et al*. cloned three other microRNAs that were regulated during the transition from resting to activated B cell upon CSR

stimulation. One, miR-181a, was significantly down-regulated during the time course of the reaction (Teng et al., 2008). At the same time, De Yebenes *et al.* undertook a screen for microRNAs that suppressed CSR and found that ectopic overexpression of miR-181b resulted in a ~50% reduction in the levels of CSR to IgG1 (de Yébenes et al., 2008). The miR-181 family comprises four distinct transcripts with identical seed regions (a, b, c and d) so it is likely that both groups identified the same functional seed region of mir-181 as functionally down-regulated in the context of CSR and, in addition, as perturbing CSR when ectopically expressed. Standard luciferase assays with mutant and wild-type versions of the AID 3'UTR suggest that miR-181b may directly control levels of AID expression; however, whether this control in the proper cellular context is direct or indirect remains to be determined.

1.4.3.3 Post-translational control of AID protein

Post-translational modifications of proteins (PTMs) are one of the most well known mechanisms used to regulate the activity of proteins, as well as provide diversity to their function. This is one mechanism by which a cell can overcome the limitations of the genome (Hunter, 2007). Thus far, much of the focus on PTMs with reference to AID regulation has been in the realm of phosphorylation, but recently investigation has spread to ubiquitination. What is known about AID in the context of these two modifications is presented below. In time, it is likely that there will be a greater understanding of the regulation of AID in terms of these, and other, powerful PTMs.

Phosphorylation:

Thus far, AID has been found to be phosphorylated at 5 different residues: serine 3 (S3), threonine 27 (T27), serine 38 (S38), threonine 140 (T140) and tyrosine 184 (T184) (Basu et al., 2005; Gazumyan et al., 2011; McBride et al., 2006; 2008; Pasqualucci et al., 2006).

The first occurrence of AID phosphorylation observed, and that which has been studied the most, is the phosphorylation of S38. Carried out by the ubiquitous, cAMP-dependent kinase, PKA, this phosphorylation event has proven to be necessary for both efficient SHM and CSR (Basu et al., 2005; Cheng et al., 2009; McBride et al., 2006; Pasqualucci et al., 2006). The finding that S38 phosphorylation is necessary for the interaction of AID with the ssDNA binding protein, RPA, provided some insight into the importance of this modification (Basu et al., 2008; Chaudhuri et al., 2004; Rada, 2009; Vuong et al., 2009). Initially, RPA was proposed to stabilize ssDNA in transcription bubbles to maintain a substrate for AID; however, this would not be necessary during CSR due to ability of switch regions to form stable R-loops. Based on this, it would be expected that loss of S38 phosphorylation would only result in a defect in SHM, not CSR. More recent work, though, has provided convincing evidence that RPA plays distinct roles during SHM and CSR. The S38A, phospho-null mutant of AID is unable to deaminate a transcribed SHM substrate *in vitro*, while still maintaining the ability to deaminate a transcribed CSR substrate *in vitro* (Vuong et al., 2009). Thus, it appears that RPA does play a role in increasing access of

ssDNA substrates for AID during SHM, when R-loops do not form, but not during CSR. The defect in CSR observed upon loss of S38 phosphorylation appears to occur downstream of AID-mediated deamination. It has been shown that AID can access the S regions in the absence of RPA, and in fact, that phosphorylation of AID at the Ig locus is necessary for the recruitment of RPA to the locus (Vuong et al., 2009). Thus, during CSR, it appears that RPA has been co-opted to play a role in DSB repair. Recent work has shown that RPA binds ssDNA exposed by 5' -to- 3' resection and is predicted to allow for alternative, microhomology-based end-joining to occur if C-NHEJ fails (Yamane et al., 2013). Whether the recruitment of RPA plays other roles during CSR, such as the recruitment of downstream repair factors, remains to be seen.

Similar to S38 phosphorylation, phosphorylation of T140 plays a role in enhancing both SHM and CSR. Mutation of T140 to generate a phospho-null mutant of AID results in a decrease in both CSR and SHM. Interestingly, SHM is affected to a greater extent than CSR, suggesting that phosphorylation of T140 may play a role in governing the interaction of AID with factors that are specific for each process (McBride et al., 2008; Vuong and Chaudhuri, 2012).

Unlike the other phosphorylation sites discussed, phosphorylation of S3 inhibits AID activity. Mutation of the site to prevent phosphorylation leads to an increase in CSR, as well as off-target effects of AID activity, such as cMyc-IgH translocations. This phosphorylation event can be generated *in vitro* by the ubiquitous kinase, PKC and removed by the phosphatase, PP2A. In fact,

inhibition of PP2A results in an increase in AID phosphorylation at this site and a concomitant decrease in CSR (Gazumyan et al., 2011). The mechanism of how S3 phosphorylation inhibits activity is not yet known, but may involve the interaction of AID with proteins that block its activity.

The phosphorylation events on Y184 and T27 do not have any known function at this point. Both fall within predicted PKA phosphorylation motifs and both can be phosphorylated *in vitro* by PKA. While Y184-Phospho AID has been detected in primary B cells by mass-spectrometry analysis, the T27-phospho form of AID has not, raising questions about its physiological relevance. In addition, mutation of the Y184 phosphorylation site does not severely inhibit CSR, suggesting that this phospho-event may not play a role in CSR (Basu et al., 2005; Pasqualucci et al., 2006). Additional experiments are necessary to determine the functionality of these two modifications (Vuong and Chaudhuri, 2012).

Ubiquitination:

Ubiquitin is an approximately 8.5 kDa protein that is covalently attached at its C-terminus to an amino acid, typically a lysine, on a target protein. Ubiquitination can occur in many flavors, ranging from mono-ubiquitination, or the attachment of a single ubiquitin moiety to the target protein, to poly-ubiquitination, or the attachment of ubiquitin chains to the target protein. Ubiquitin chains are formed by the covalent attachment of one ubiquitin protein to one of seven lysine residues on another ubiquitin protein. Each of the different types of chains,

composed of one or many different ubiquitin linkages, are recognized differently by ubiquitin-binding proteins and thus can produce very different downstream events. Lysine-48 (K48) linkages are commonly found in proteins targeted for degradation by the proteasome, lysine-63 (K63) linkages are commonly found in cell signaling complexes, and mono-ubiquitination of histones has been shown to play a role in nucleosome dynamics and the regulation of transcription (Pickart, 2001a; 2001b). Despite these examples, though, it is very likely that each of these types of ubiquitination can play a variety of roles, depending on the cellular context. An overview of the enzymes involved in ubiquitination is presented in Figure 1.7.

Aoufouchi *et. al.* were the first to observe poly-ubiquitination of nuclear AID upon treatment of cells with a proteasome inhibitor, MG132 (Aoufouchi et al., 2008). This provides an explanation for the observation that nuclear AID has a much shorter half-life than cytoplasmic AID and suggests ubiquitin-mediated degradation as a mechanism to regulate AID activity in the nucleus (Geisberger et al., 2009). Cytoplasmic poly-ubiquitination has also been observed upon inhibition of the chaperone, Hsp90 (Orthwein et al., 2010).

Neither of these papers were able to identify (1) the residue(s) on AID that were ubiquitinated and (2) the E3 ubiquitin ligase(s) involved in the ubiquitination event. In addition, while the pattern of ubiquitination observed by Aoufouchi *et. al.* can be interpreted as poly-ubiquitination, the lower molecular weight “laddering” pattern seen is more reminiscent of multiple mono-ubiquitination events

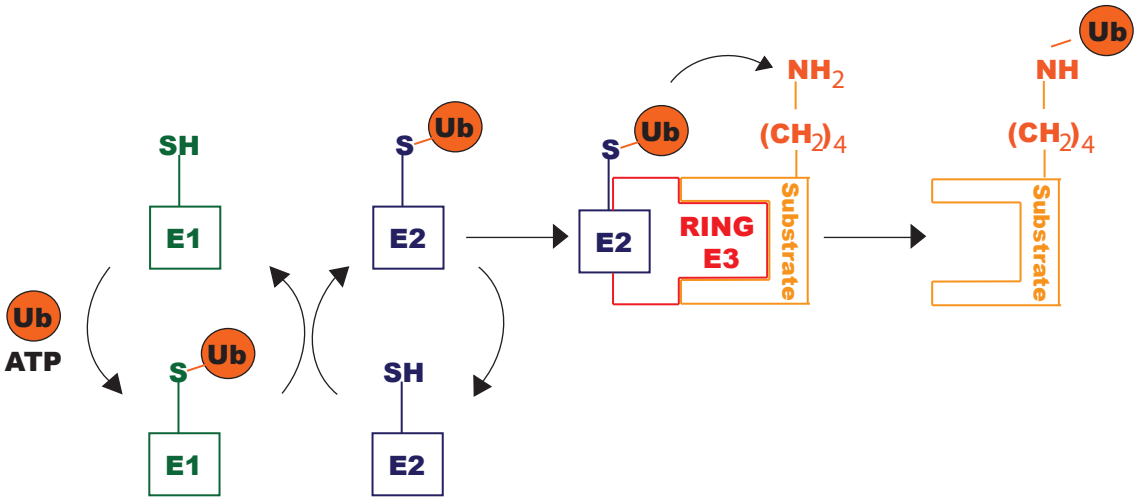


Figure 1.7 Overview of the ubiquitination cascade. Ubiquitination requires a cascade of three enzymes: the ubiquitin activating enzyme (E1), the ubiquitin conjugating enzyme (E2) and the ubiquitin ligase (E3). In an energy-dependent mechanism, ubiquitin is covalently linked at its C-terminus to a cysteine residue on the E1 and then transferred to a cysteine on the E2. E3 proteins, which contain RING domains, act as a bridge between the E2 and the substrate protein and provide much of the specificity for the substrate. This bridge allows for the transfer of the ubiquitin from the E2 to a residue (lysine, shown) on the target protein. A mono-ubiquitination event is shown, however repetition of this same mechanism is used to produce poly-ubiquitin chains.

(Aoufouchi et al., 2008). This opens the exciting possibility that AID is both poly-ubiquitinated for degradation, and mono-ubiquitinated for some other *in vivo* function. This thesis presents the identification of a novel E3 ligase that is able to mono-ubiquitinate AID. It will be very exciting in the future to discern the cellular roles of each of the different ubiquitination events and also determine whether one or multiple ubiquitin ligases are used in these processes.

1.4.3.4 Subcellular localization of AID

One means to avoid excess genomic mutation by a potent DNA mutator, such as AID, is to regulate its cellular localization. Exclusion of excess AID from the nuclear compartment should relieve mutagenic stress. In fact, it has been observed that in cells where AID is known to be active in the nucleus, no nuclear accumulation can be detected (Rada et al., 2002a). Thus, there exist mechanisms to actively import AID into the nucleus when necessary, export it from the nucleus to regulate its activity and maintain it in the cytosol to prevent off-target mutation.

Nuclear export of AID is dependent on the CRM1 export pathway and is thus sensitive to inhibition by the drug, Leptomycin B (LMB). This export has been shown to rely on the last ten amino acids of AID (residues 188-198), which make up the nuclear export sequence (NES, Figure 1.8). Treatment of cells with LMB, or mutation of key hydrophobic residues within the NES, results in the accumulation of nuclear AID (Brar et al., 2004; Ito et al., 2004; McBride et al., 2004). Interestingly, mutation of the C-terminal NES also results in a decrease in

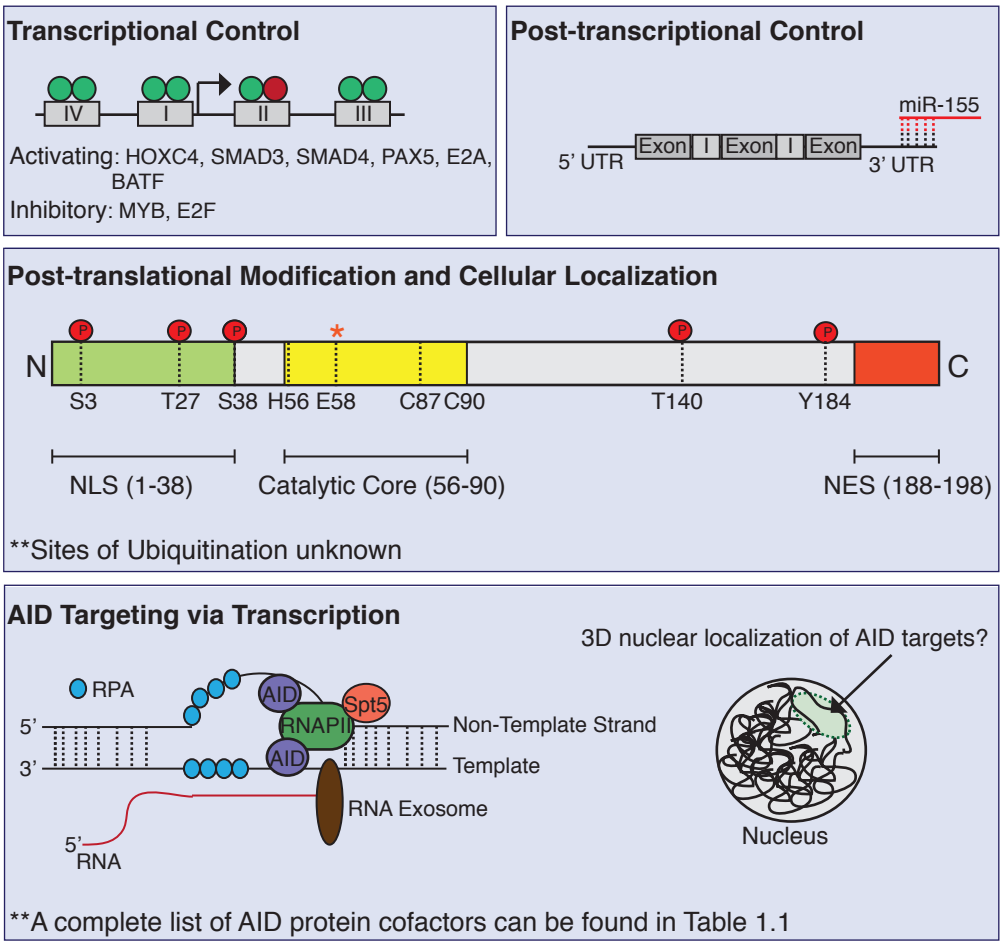


Figure 1.8 Regulation of AID. Various mechanisms of AID regulation are shown. Transcription of AID is regulated by a balance of activating (green) and inhibitory (red) transcription factors. AID expression is further regulated at the level of translation via miRNAs, specifically miR-155. AID protein is regulated by the addition of post-translational modifications. The five known phosphorylation sites are depicted with a "P" enclosed in a red circle. Ubiquitination has been observed, but the site of modification is not known. Catalytic residues of AID are also shown. H56, C87 and C90 participate in the coordination of Zn, and E58 plays a catalytic role in the deamination mechanism. AID subcellular localization is tightly regulated by an N-terminal nuclear localization signal (NLS) and a C-terminal Nuclear Export Sequence (NES). Lastly, AID association with transcriptional machinery plays a role in targeting. Stalling by the factor, Spt5, RNA processing and degradation by the RNA exosome and binding of DNA by RPA all play a role in enabling AID access to the DNA in order to induce mutations. Lastly, genomic loci neighboring the Ig locus in the 3D space of the nucleus have an increased chance of mutation and translocation. Thus, it is likely that a subset of the nucleus is targeted by AID; however, the mechanisms involved here are unknown.

class-switch recombination while maintaining the ability to mutate (McBride et al., 2004). Replacement of the physiological NES with a heterologous NES cannot rescue CSR to wildtype levels, despite restoring export (Geisberger et al., 2009). Thus, it appears likely that either the C-terminus plays two roles—one that is export-specific and another that is CSR-specific, or that export is important for the mechanism of CSR, but is extremely sensitive to the efficiency of AID binding to exportins. In line with dual roles for the C-terminus, it has been shown that AID is actively imported into the nucleolus and export from this compartment relies on the C-terminus of AID. It is possible that movement in and out of the nucleolus is an important step in CSR, and thus the presence of this export sequence would be as well (Hu et al., 2012).

Though less well understood, recent findings have begun to uncover the mechanistic details of AID nuclear import. A nuclear localization signal (NLS) composed of several non-contiguous basic residues is located at the N-terminus of AID (Hu et al., 2012; Patenaude et al., 2009a). Proper nuclear import relies on the proper structural folding of AID, as the first 39 residues of AID are required (Hu et al., 2012) (Figure 1.8). Interestingly, the same signal is used to localize AID to the nucleolus, suggesting the importance of this compartment for CSR (Hu et al., 2012).

Last, it is likely that AID is retained in the cytosol by association with a cytosolic complex. Residues 158-198 of AID have been shown to be important for cytosolic retention. Replacement of this region with the analogous region in

Apobec 2 allows for passive diffusion of AID into the nucleus in the presence of the export inhibitor, LMB (Patenaude et al., 2009b). In addition, cytosolic AID has been purified from DT40 chicken B cells and shown to be part of a complex that sediments on a sucrose gradient with a sedimentation coefficient of 10-11S, suggesting that the complex is about 350 kDa (Häsler et al., 2011). Within this complex, AID has been shown to interact with the translation elongation factor, eEF1A; elimination of this interaction by mutating the residues involved results in an increase in the nuclear accumulation of AID in the presence of LMB (Häsler et al., 2011). Future experiments will be necessary to determine other key components of this cytosolic retention complex.

1.4.3.5 AID protein cofactors

The fact that AID is the only known B cell-specific factor necessary for SHM and CSR, coupled with the extraordinary specificity of SHM and CSR, has prompted intense investigation to understand how AID is targeted specifically to the Ig locus, sparing the rest of the genome. In addition, it still remains unknown how AID is locally positioned at the Ig locus, targeting the variable region during SHM and the switch regions during CSR. To this end, several labs have conducted screens to search for AID-interacting proteins that could be the elusive AID-targeting factor.

The ubiquitous family of 14-3-3 adaptor proteins were identified to bind the S region motif 5'-AGCT-3' with high affinity, thus suggesting them as a potential targeting factors (Xu et al., 2010). In fact, drug inhibition, knock-out or expression

of dominant negative forms of various 14-3-3 isoforms resulted in decreased accumulation of AID at S regions and a concomitant decrease in CSR (Xu et al., 2010). While these proteins may play a role in CSR recruitment, the 5'-AGCT-3' motif is not abundant in the variable region, suggesting that a different, and unknown, mechanism is needed to recruit AID to the Ig locus during SHM. It is possible, though, that differential targeting of AID during SHM versus CSR is mediated by its interaction with a different subset of cofactors. However, because 14-3-3 proteins are ubiquitously expressed, it seems unlikely that these proteins, alone, are sufficient to recruit AID specifically to S regions.

In addition to the recognition of DNA sequences, many chromatin modifying factors have been identified, which regulate the recruitment of AID and/or repair factors to the Ig locus. The histone chaperone, FACT complex, has been shown to play a role in both SHM and CSR, suggesting the importance of nucleosome dynamics in making the locus accessible for SHM and CSR protein factors (Aida et al., 2013; Stanlie et al., 2010). Interestingly, FACT deficiency during CSR does not affect transcription through the locus, but rather cleavage of the DNA (Stanlie et al., 2010). This suggests that chromatin dynamics are equally important for DSB formation and repair.

Unlike the FACT complex, many of the identified chromatin remodelers have specific effects on either SHM or CSR. KAP1, which recognizes H3K9me3, a histone modification associated with CSR, binds and recruits AID to the Sμ switch region. Loss of KAP1 results in a decrease in AID recruitment to S

regions, as well as decrease in CSR. Interestingly SHM is not affected in mice, which lack KAP1 (Jeevan-Raj et al., 2011). Similarly, the histone chaperone, Spt6, can interact directly with AID. Knockdown of Spt6 results in a significant decrease in CSR levels in the CH12 cell line (Okazaki et al., 2011). The mechanism by which Spt6 acts in this setting remains unknown. In contrast, the nuclear export factor, GANP, actively remodels chromatin specifically at the variable regions through its histone acetyl transferase activity. This modification allows access of AID to the region (Kuwahara et al., 2000; 2004; Maeda et al., 2010; Singh et al., 2013). Mice deficient in GANP exhibit normal CSR, but decreased SHM.

Many of the AID-interacting proteins discovered provide a link between AID activity and the well-known fact that transcription is highly correlated with SHM and CSR. First, AID has been shown to directly interact with RNA Polymerase II (RNAPII), thus providing a direct connection between transcription and mutation (Besmer et al., 2006; Nambu et al., 2003). Further, as already discussed, the ssDNA binding protein, RPA, is thought to play a role in stabilizing a ssDNA intermediate during transcription of SHM targets (Chaudhuri et al., 2004; Vuong et al., 2009). Last, the RNAPII associated stalling-factor, Spt5, interacts with AID and facilitates its interaction with RNAPII. Spt5 localizes with RNAPII and AID throughout the genome and sites at which Spt5 accumulate are predictive of AID-mediated mutation (Pavri et al., 2010). Based on these findings, it is likely that AID associates with RNAPII and the transcription machinery. As

RNAPII stalls during its elongation phase, AID disassociates, allowing for mutation. In line with this model, insertion of a transcription terminator containing a poly(A) signal downstream of the variable region promoter in DT40 cells increased AID-mediated mutation. Additional AID RNAPII stalling at the Poly(A) signal sequence likely accounts for the increase in mutation load (Kodgire et al., 2013). In an attempt to explain how AID activity is limited to a distance of 2 Kb from the promoter whereas mRNA transcription is not terminated until much later, it has been proposed that not all transcription complexes at the Ig locus are created equal. Non-AID containing RNAPII could be responsible for transcribing the complete mRNA, whereas AID/Spt5-containing transcription complexes may be more prone to stalling, releasing AID before transcription has terminated (Kodgire et al., 2013). This model would predict that several incomplete Ig transcripts exist in B cells stimulated for CSR and/or SHM. In fact, recent work from the Basu lab has shown that in the absence of a critical subunit of the RNA exosome, an accumulation of ncRNAs occurs at AID-targeted loci (Basu *et. al.*, unpublished, “Immunity and Tolerance” CSH Meeting, May/June 2013). Thus, it is possible that the existence of these truncated transcripts is masked in wildtype cells by the activity of the exosome. Future experiments are needed to piece apart the genesis and role of these ncRNAs, as well as the validity of this model.

The importance of RNA to SHM and CSR has become increasingly appreciated. Not only is transcription necessary, but also processing of the transcripts through splicing is required. As such, AID has been found to interact

directly with the splicing factors, CTNNBL1 and PTBP2 (Conticello et al., 2008; Nowak et al., 2011). Knockdown of PTBP2, specifically, results in a decrease in the accumulation of AID at the Ig locus (Nowak et al., 2011). Thus, it is possible that either the splicing machinery, or the processed RNA, itself, plays a role in AID stabilization at the locus. Either way, it is clear that RNA processing, including splicing and degradation by the Exosome, is intimately linked to AID activity.

As mentioned, the majority of AID in the cell is not in its active form at the Ig locus. Thus, it is likely that AID is a member of several different protein complexes throughout its lifetime. As discussed, cytosolic AID is complexed with both eEF1A and Hsp90 (Häsler et al., 2011; Orthwein et al., 2010). In addition, Regγ, a protein associated with ubiquitin-independent protein degradation, binds nuclear AID and also plays a role in regulating the abundance of AID protein (Uchimura et al., 2011). With time, more of the components of the cytosolic and nuclear AID complexes will be revealed, hopefully along with a better understanding of how the activity of AID is regulated. In addition, it has been shown once that PTMs on AID can affect protein-protein interactions. Specifically, phosphorylation of AID on S38 enhances binding to RPA. It will be interesting in the future to understand if, and how, other PTMs, such as ubiquitination, play a role in regulating AID complex formation. A list of identified AID cofactors is presented in Table 1.1.

Table 1.1 AID protein cofactors. A list of identified AID protein cofactors is given, along with a brief description of its function.

DNA BINDING	
14-3-3 Adaptors	Bind AGCT repeats in switch regions
CHROMATIN FACTORS	
FACT	Histone chaperone, involved in SHM and CSR
KAP1	Recognizes H3K9me3, involved in recruitment of AID to switch regions, CSR specific
Spt6	Histone chaperone, CSR specific
GANP	Histone acetyl transferase, makes variable regions accessible to AID, SHM specific
TRANSCRIPTION-ASSOCIATED	
RNAP II	RNA Polymerase
RPA	ssDNA binding protein, binds phosphorylated AID, involved in both SHM and CSR
Spt5	RNAP stalling factor, enhances binding of AID with RNAPII
RNA-ASSOCIATED	
Exosome	Processing and degradation of RNA, increases access of AID to template strand, involved in SHM and CSR
CTNNBL1	Spliceosome-associated factor
PTBP2	Splicing factor, stabalizes AID at Ig locus
CYTOSOLIC	
EEF1A	Maintains AID in cytosol
HSP90	Stabilizes AID in cytosol
OTHER	
PKA	Phosphorylates AID
REGγ	Degradation of nuclear AID

1.4.3.6 AID targeting and off-target mutation

It is clear that transcription and the transcription machinery plays a role in targeting AID. However, many genes in activated B cells are highly transcribed. Thus, it still remains unclear why AID activity accumulates at the Ig locus. ChIP-Seq studies of AID and Spt5 demonstrate that AID accumulates at many highly transcribed genes in the genome, suggesting that the specificity of mutation occurs downstream of initial AID targeting (Pavri et al., 2010; Yamane et al., 2011). Though it is clear that off-target AID activity occurs, as both mutation and DSB formation have been detected outside of the Ig locus, these mutations and translocations form at much lower rates than what is seen at the Ig locus. (Klein et al., 2011; Liu et al., 2008; Ramiro et al., 2004; Robbiani et al., 2009). Chromatin Capture studies coupled with deep-sequencing have revealed that genes which are frequently mutated by AID display proximity to the Ig locus in the 3D space of the nucleus (Rocha et al., 2012). Genes that are actively transcribed are often spatially localized, thus providing a connection with the earlier studies connecting AID to highly transcribed genes. However, again, not all transcribed genes are targets of AID and thus there must be unidentified factors that introduce greater specificity.

In addition to the trans-factors discussed, many labs have searched for cis-regulatory modules that recruit AID to the Ig locus. While the promoter is clearly important for the induction of transcription, the nature of the promoter and neighboring enhancer regions may play a role in the recruitment of AID. An

analysis of non-Ig genes that undergo AID-mediated mutation revealed the presence of three transcription factor binding sites: E-box motifs, binding sites for YY1 and C/EBP-β (Duke et al., 2013). These factors may work together to recruit AID, providing a mechanism for the specific recruitment to some transcribed genes, but not others. Experiments conducted in DT40 B cells have shown that regions 3' of the Ig locus participate in the recruitment of AID. Deletion of a 6 Kb region downstream of the 3' RR resulted in a loss of gene conversion (GCV) and mutation, despite high levels of transcription due to the insertion of an SV40 enhancer (Kothapalli et al., 2008). Further, insertion of a region 9.8 Kb downstream of the start site of the IgL locus in DT40 cells into a non-Ig locus was sufficient to activate hypermutation, thus suggesting that this region could recruit AID and/or other necessary factors (Blagodatski et al., 2009). Further experiments will be needed to work out the specific sequences involved in recruitment, whether it is AID or some other factor that is recruited here, and how these cis-regulatory modules work together with the transcriptional machinery and other trans-factors.

1.5 Statement of Problem

AID is a very potent enzyme with the ability to induce mutation in the genome of cells in which it resides. When harnessed properly, the capacity of this enzyme to mutate is able to produce a large and diverse repertoire of antibody specificities and functionalities to fortify the host's adaptive immune system. However, when used improperly, AID has the potential to wreak havoc

on the health and stability of the host's genome. This misuse of AID can, and has been shown to, result in the accumulation of point mutations and the formation of DNA breaks and aberrant chromosomal translocations at non-immunoglobulin loci in the genome---genomic abnormalities that leave the host vulnerable to the formation of cellular cancers. Because of this, the localization and activity of AID must be meticulously regulated to assure the health of the cell and maintain the balance of genomic instability within the immunoglobulin loci and genomic stability elsewhere.

Because of the great potential for harm due to the mis-regulation of AID, it is likely that several layers of regulatory mechanisms exist to ensure that AID-mediated deamination is targeted specifically. In fact, much of the research within the realm of AID biology is focused on understanding how AID, itself, as well as its activity, is monitored by the cell. Many of these modes of regulation have been discussed and likely involve the interaction of AID with other cellular proteins, which could either target AID to the immunoglobulin loci, regulate its activity, or participate in the recruitment of events downstream of AID-mediated deamination.

The interest in the involvement of AID protein cofactors in antibody diversification has led many laboratories to search for proteins, which can directly bind AID. However, this venture has proven to be particularly challenging because of the propensity for AID to become insoluble when ectopically expressed. Typically, ectopic expression of AID results in its mis-folding,

aggregation and presence in insoluble inclusion bodies. This makes it difficult to purify in order to study the protein *in vitro*, as well as study its interaction to other proteins using classical techniques such as the yeast-two hybrid assay (Y2H). Thus far, a variety of techniques, including immunoprecipitation followed by mass-spectrometry and RNAi knockdown screens, have been used to identify putative AID cofactors. These screens have suggested the synchronization of AID with other broadly important cellular pathways, including RNA splicing (Conticello et al., 2008), RNA processing and degradation by the RNA exosome (Basu et al., 2011), transcription (Besmer et al., 2006; Fukita et al., 1998; Lebecque and Gearhart, 1990; Nambu et al., 2003; Rada and Milstein, 2001; Storb et al., 2007; Tumas-Brundage and Manser, 1997; Winter et al., 1997), and RNA polymerase stalling (Pavri et al., 2010); however, a clear picture of the players involved in the AID reaction and the sequence of events remains, for the most part, unclear. Thus, the search for potential AID cofactors could still benefit from the development of better techniques for identifying interacting partners for insoluble proteins.

This thesis presents the development and validation of a novel screening approach to identify interacting partners for poorly soluble proteins, as well as the application of this screen to Activation Induced Cytidine Deaminase (AID). Using the screen, several novel putative AID cofactors have been identified along with some previously discovered interacting partners, verifying the robustness of the screening approach.

In addition, because of its interesting domain structure and the potential for interesting regulatory mechanisms on AID, one interacting partner identified in the screen, RING Finger Protein 126 (RNF126), was studied further. RNF126 was verified to be a *bona fide* AID binding partner and additionally determined to act as an E3 ubiquitin ligase, modifying AID with the addition of a single ubiquitin moiety. While poly-ubiquitination of nuclear AID has been implicated in its degradation, this discovery marks the first example of mono-ubiquitination of AID. In addition, this is the first example of the discovery of an AID ubiquitination event paired with the E3 ubiquitin ligase responsible. In addition to studying the function of RNF126 *in vitro*, a conditional knockout model was generated and used to assay the importance of RNF126 during various stages of antibody diversification. While it is clear that RNF126 is not essential for antibody diversification, these studies reveal the potential role of RNF126 in fine-tuning the activity of AID during CSR and SHM. In addition, these studies demonstrate the layered intricacies of both diversification mechanisms, suggesting the importance of analyzing putative AID cofactors in a variety of contexts, as it is very possible that the importance of a single protein factor might not be detectable in an otherwise wild-type context.

The findings presented here demonstrate the utility of a novel screening technique to search for interacting partners for insoluble proteins and, through its use, expands the list of putative AID cofactors. The discovery that the novel E3 Ubiquitin Ligase, RNF126, interacts with and modifies AID presents a novel

mode of regulation of the potent mutating enzyme and paves the way for future research to uncover the role of mono-ubiquitinated AID during SHM and/or CSR. Lastly, the generation of the RNF126 conditional knockout mouse model not only will prove invaluable for future studies of this enzyme during antibody diversification, but also in all other systems where RNF126 is expressed.

CHAPTER 2: DEVELOPMENT AND VALIDATION OF A NOVEL SCREEN TO IDENTIFY INTERACTING PARTNERS FOR INSOLUBLE PROTEINS

2.1 Motivation

The cell has evolved complex signaling networks to translate external and internal cues into the appropriate cellular response to foster the survival of the cell or of the multicellular organism to which it belongs. As our understanding of these signaling networks has increased, it has become clear that the involvement of many proteins, and thus of protein-protein interactions, is necessary to carry out the process faithfully. Traditionally, protein-protein interactions have been studied on a case-by-case basis using biochemical, biophysical or genetic techniques (Phizicky and Fields, 1995). However, with the onset of genome-wide sequencing came an interest in assigning function to the slew of new and unknown open-reading frames (ORF). Thus, it became increasingly appealing to develop high-throughput methods to discover novel protein interaction partners of functionally characterized genes in order to identify biological roles of protein-protein interaction networks and to provide useful clues to assign function to unknown ORFs.

Most of the genome-wide protein-protein-interaction data published to date has been produced with yeast two-hybrid assays or with *in vivo* pull-down approaches. Yeast-two-hybrid assays are often utilized for large scale protein-protein-interaction screens (Fields and Song, 1989; Ghavidel et al., 2005) and rely on the modular nature of eukaryotic transcription factors in which one domain

binds DNA in a sequence specific manner and the other domain acts to activate transcription. Thus, when one protein of interest is fused to the DNA binding domain and its interacting partner to the activation domain, proof of interaction can be determined by the transcription of a reporter gene. On the other hand, *in vivo* pull-down approaches toward identifying protein-protein interactions typically rely on the addition of an affinity tag to the protein of interest, expression in its native cell and purification of the protein along with its bound cofactors. Mass spectrometry is then applied to identify the proteins that precipitated along with the tagged protein of interest. Although these purification methods were first developed for small scale protein identification experiments, they have been successfully adapted for use in genome-wide proteomics studies (Alber et al., 2007a; 2007b; Gavin et al., 2006; Krogan et al., 2006). Both yeast-two-hybrid and affinity based techniques generate false positives and require independent confirmation of potential interaction partners.

More importantly, though, proteins that are poorly soluble or insoluble when ectopically expressed are least amenable to characterization using these tools. Poor solubility upon ectopic expression of a protein can result from either improper folding or the exposure of hydrophobic domains that would be masked when in complex with cofactors. This problem is particularly palpable for those who study the enzyme, AID. Because of the mutagenic potential of AID, it has been predicted that key regulatory mechanisms involve the interaction of AID with protein cofactors, which either regulate through binding or impart a post-

translational modification, another mechanism known to be invaluable for AID regulation. However, because AID is poorly soluble upon expression in bacterial cells, it has proven difficult to use traditional techniques to screen for AID-interacting partners.

2.2 Design of solubility-based interaction screen

To specifically address the problem of cofactor identification for poorly soluble proteins, such as AID, we have devised a new screening approach that actually takes advantage of the insolubility of the protein of interest to identify interacting partners. Often proteins that are insoluble when expressed alone will form soluble complexes when co-expressed with a native binding partner. In practice, this technique involves the co-expression and co-folding of two proteins, an unknown protein expressed from a cDNA library of choice, and a known, insoluble protein fused to an antibiotic resistance protein. As will be shown in the following proof-of-principle experiments, fusion of the gene for an insoluble protein with the gene coding for antibiotic resistance produces a fusion protein that is also insoluble. Thus, when this fusion gene is expressed, bacteria are *only* able to grow under drug selection when the fusion protein is solubilized, but not when it is in its insoluble form. Solubilization of the protein of interest, and of the fusion protein, rescues the ability of bacteria to grow under drug selection. For these experiments the drug resistance gene, Chloramphenicol Acetyl Transferase (CAT), was used, which confers resistance to the drug, chloramphenicol.

2.3 Validation of interaction screen

Prior to applying this technique to AID, it has been validated to establish that: (1) fusion of soluble and insoluble proteins to CAT yields soluble and insoluble fusion proteins, respectively, (2) drug resistance can be rescued by solubilization of a known insoluble protein with its native binding partner, and (3) positive interaction and subsequent solubilization can be detected even when the interacting partner constitutes the minority of a population of expressed proteins.

In order to establish that the solubility status of the fusion protein is directly correlated with that of the fused protein of interest, AID was used as a representative insoluble protein and the phage effector protein, Cell-Cycle Inhibitory Factor (CIF) (Hsu et al., 2008), a protein of similar size to AID, as a representative soluble protein. Bacterial expression of AID with a C-terminal fusion to CAT rendered cells unable to grow on chloramphenicol plates. In contrast, C-terminal fusions of CAT to CIF enabled cells to be resistant to even high doses of chloramphenicol, reaching up to 240 μg/mL (Figure 2.1).

In addition, it was necessary to establish that interaction of a known cofactor to the fused protein of interest could rescue solubility. In order to test this, we have used a known heterodimer in which one component is necessary for the solubilization of the other. It has been previously established that the tRNA editing enzyme adenosine-deaminase acting on tRNA (ADAT3) is insoluble unless co-expressed with the other component of the heterodimer, ADAT2 (Rubio et al., 2007). Utilizing an N-terminal fusion of ADAT3 to CAT, it can be

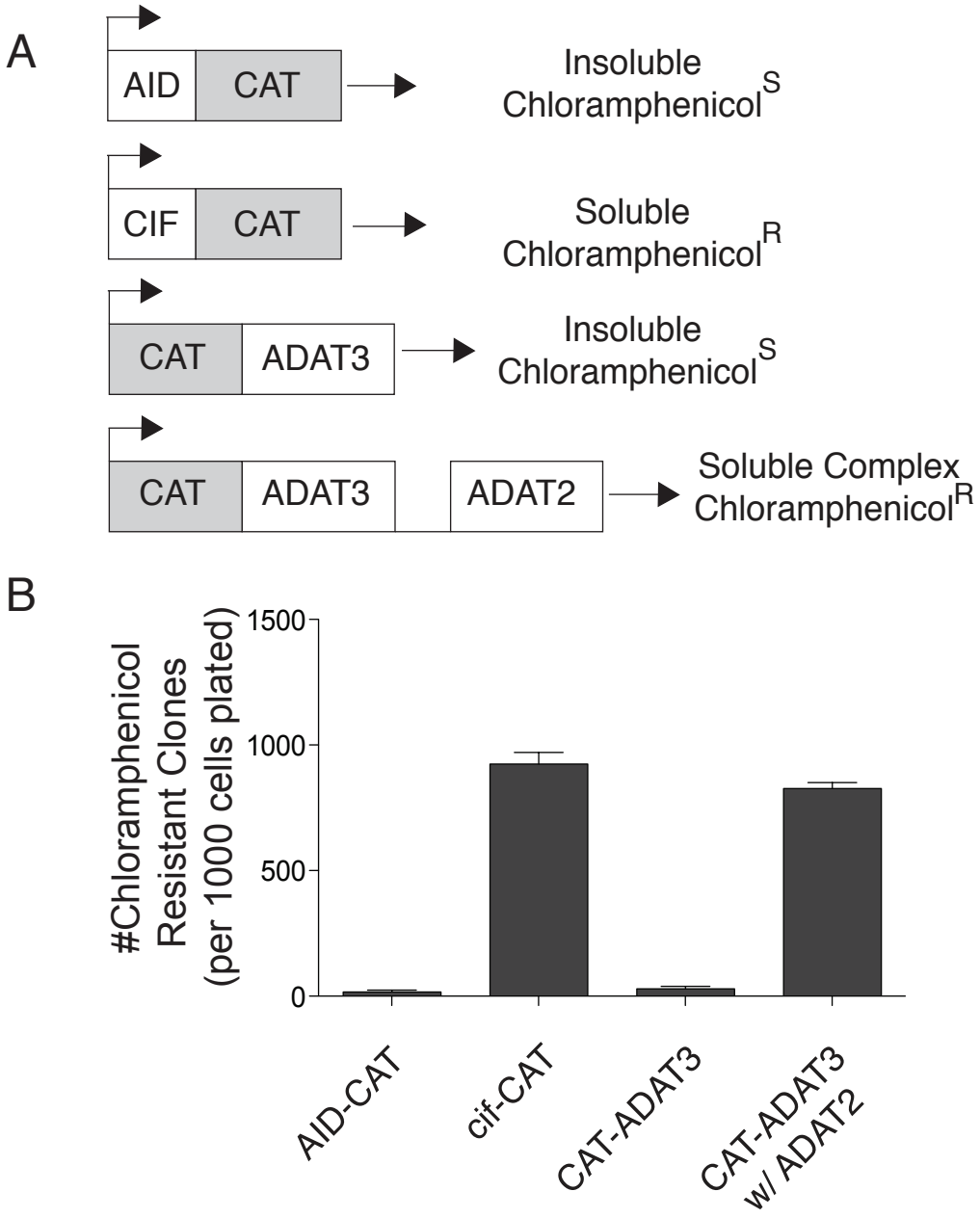


Figure 2.1. A genetic assay selects for the restoration of solubility of an insoluble protein. (A) A schematic depicts representative CAT fusion proteins used. AID and AID-CAT are insoluble and thus produce chloramphenicol sensitive cells; Cif and Cif-CAT are soluble allowing cells to be resistant to chloramphenicol; ADAT3 and ADAT3-CAT are insoluble unless coexpressed with its binding partner, ADAT2. (S=sensitive, R=resistant). (B) Plasmids carrying the indicated genes were transformed into BL21ai E. coli, plated on chloramphenicol containing plates under induction conditions. Resistant colonies were counted. Shown here are colony numbers obtained on LB plates containing 120 μg/ml chloramphenicol.

seen that, much like for AID, expression of ADAT3-CAT in bacteria results in drug sensitivity and the inability of colonies to form on chloramphenicol plates. Only upon co-expression of ADAT2 with ADAT3-CAT, can bacteria harboring the ADAT3-CAT fusion protein survive on chloramphenicol plates (Figure 2.1). Together, these results establish that fusions of CAT render cells resistant to chloramphenicol as long as the fused protein is soluble. Furthermore, at least for the case of the heterodimer ADAT2/3, reconstitution of a soluble complex by co-expression of a necessary binding partner rescues the ability of bacteria to be grown on chloramphenicol plates.

To assure that this screening technique would be able to identify novel AID interacting partners when expressed amongst a large library of sequences, we took advantage of proteins known to be able to bind AID. In particular, the C-terminal, RING domain containing, region of the E3 ligase mutant double minute (mdm2) has been shown to interact with AID by a yeast-two-hybrid screen (MacDuff et al., 2006). Co-expression of the C-terminal region of mdm2 with the AID-CAT fusion protein was able to rescue the solubility of the fusion protein and produce resistance to chloramphenicol when neither expression of AID-CAT alone nor mdm2 alone could do this (Figure 2.2). Thus, this technique was able to confirm the interaction, which was first shown using the traditional yeast-two-hybrid screen.

In order to demonstrate that an interaction could be efficiently isolated from a pool of known negatives, we generated a small library of irrelevant

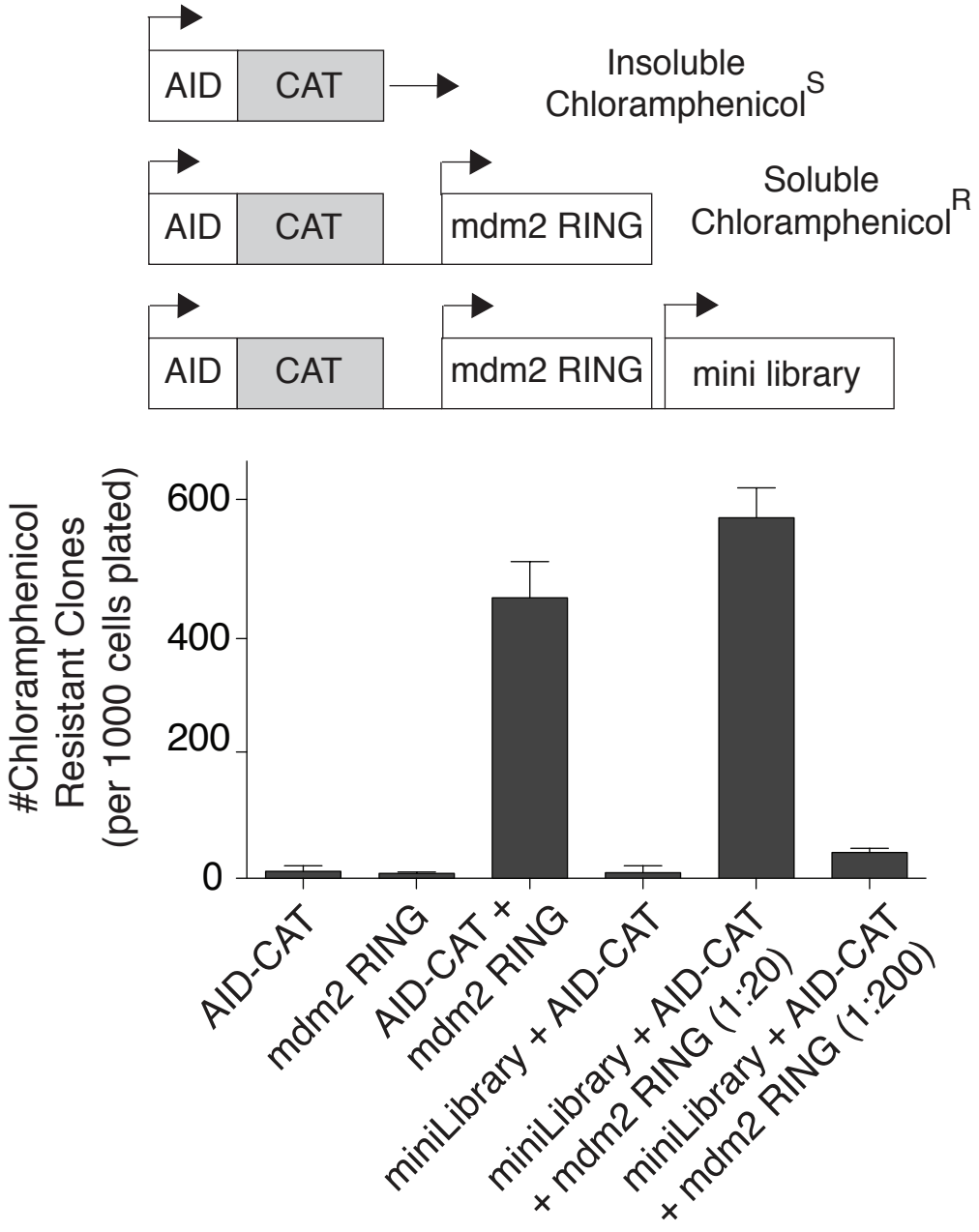


Figure 2.2 Detection of a positive interaction amongst non-interactors. (A) A schematic depicts representative CAT fusion proteins. (B) Chloramphenicol resistant colonies after co-transformation of the indicated expression plasmids were counted. Co-expression of AID-CAT and the Mdm2 RING domain allowed for efficient survival on chloramphenicol plates. These interactions were revealed even when the Mdm2 RING plasmid was diluted 1:20 and 1:200 in a mini-library of known non-interactors.

bacterial cDNAs, which consisted of 20 cDNAs encoding known bacterial virulence factors. Co-transfection of this mini-library with AID-CAT did not render bacteria able to survive on chloramphenicol plates, demonstrating that, as expected, none of these twenty proteins are able to bind AID. However, dilution of the mdm2 RING domain expressing plasmid within the mini-library in a 20:1 or 200:1 ratio followed by transformation and expression in the presence of AID-CAT, recovered chloramphenicol resistant clones (Figure 2.2). Importantly, all of these clones expressed mdm2 RING domain and AID-CAT, confirming that the solubilization of AID was due to the interaction with the RING domain of mdm2. Thus, this screening technique is sensitive enough to isolate positive AID interactions even when the interacting protein constitutes a minority of the total population of the library.

2.4 Concluding Remarks

Recent progress has been made in the development of novel techniques to assay protein-protein interactions on a high-throughput scale with the goal of filling in the gaps in the interactome left by the use of more classical techniques. Despite the many benefits of these new methods, though, none are particularly reliable for the identification of interactions with a protein that is poorly soluble or insoluble when ectopically expressed. The solubility-based screen developed here specifically targets this subset of proteins in an attempt to expand the repertoire of methods used to identify protein-protein interactions.

The set of experiments discussed thus far establish the utility of this technique when working with poorly behaved proteins. By generating fusion proteins

between a protein of interest and a drug-resistance protein, it has been established that there is a direct link between solubility and drug resistance. That is, fusion of a soluble protein produces a soluble fusion protein and drug resistance and fusion of an insoluble protein produces an insoluble fusion protein and drug sensitivity. In addition, solubilization induced through the interaction of an binding partner can rescue drug-resistance, providing a direct read-out for protein-protein interactions, thus rendering this technique useful for high-throughput screening for binding partners of a protein of interest.

CHAPTER 3: IDENTIFICATION OF RING FINGER PROTEIN 126 AS AN E3 UBIQUITIN LIGASE FOR AID

3.1 Application of Interaction Screen to AID

As a source of potential AID-interacting proteins a full length, normalized cDNA library was generated from the RAMOS human B cell lymphoma line. This cell line constitutively expresses AID and hypermutates, suggesting that necessary AID cofactors are also expressed (Sale and Neuberger, 1998). The library, which contains approximately 22,000 unique clones, was screened in triplicate in order to be reasonably confident of near-complete coverage. From these replicates, a list of 127 candidate interactors was generated, of which 36 were cloned more than once in each of the three experiments. The interaction of these 36 proteins with AID was further validated by co-expression experiments in *E. coli*. Each candidate was tagged with an N-terminal hexa-histidine tag and co-expressed in BL21 DE3 cells along side FLAG-tagged AID. Reciprocal affinity purification and western blots were conducted for all 36 candidates and 30 out of 36 co-purified with AID, suggesting a 16% false positive rate. An outline of the screen is presented in Figure 3.1

As validation for this novel screening technique, many of the previously determined putative AID cofactors were also identified in this screen. These include components of the RPA complex (Chaudhuri et al., 2004), the splicing factor, CTNNBL1 (Conticello et al., 2008), Karyopherins (Patenaude and Di Noia, 2010; Patenaude et al., 2009a) , and finally AID itself. The fact that AID-CAT binds co-expressed AID raises the question of why AID-CAT cannot simply

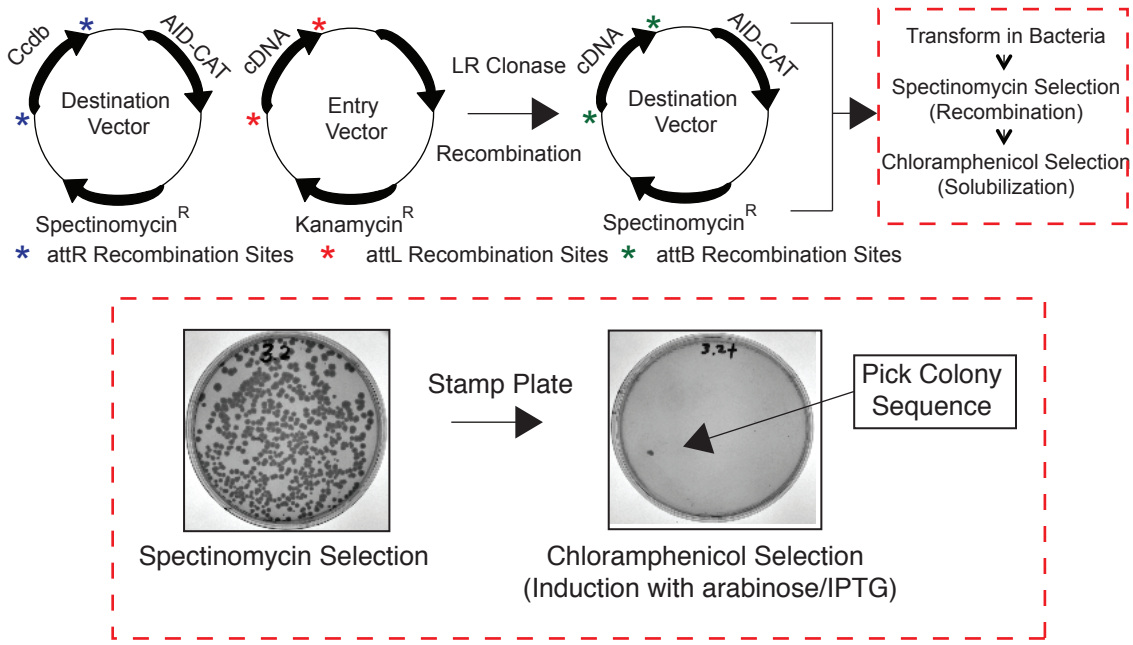


Figure 3.1 Detailed Schematic of Screening Process. The Gateway Cloning System (Invitrogen) was used to recombine an entry vector containing cDNA from the RAMOS cDNA library with a destination vector containing AID-CAT. Recombination occurs between “attR” and “attL” sites using the “LR Recombinase”. Notably, recombination replaces the toxic Ccdb gene with a cDNA from the library. In addition, because only the destination vector contains spectinomycin resistance, successful recombination can be selected for on spectinomycin plates. These plates are grown O/N at 30°C to minimize the potential toxicity due to trace expression of a toxic cofactor. The colonies from spectinomycin selection are transferred (using velvet pads) onto new plates containing spectinomycin, chloramphenicol to select for solubilization of AID-CAT, and arabinose and IPTG to induce T7 RNAP and AID-CAT/cofactor expression, respectively. Again, plates are grown O/N at 30°C. Colonies that grow on these plates represent potential AID-interacting candidates as they were able to solubilize AID-CAT. These colonies are picked, liquid cultures grown, and DNA prepared and sequenced to identify the unknown cDNA from the library.

dimerize and rescue solubility. This is likely related to the trimeric nature of active CAT, which would necessitate the placement of three AID molecules in the AID-CAT trimer at a significant distance from one another. Thus, multimerization of the AID moiety would preclude CAT trimerization and thus inhibit survival on chloramphenicol plates (Leslie et al., 1988).

The majority of the novel factors identified in the screen fall within the category of mRNA transcription, processing, splicing and export—processes that have already been shown to be necessary for CSR, though the mechanism remains unknown. In addition, a significant number of factors have been shown to be important for DNA repair, again a process necessary for CSR. The remaining identified cofactors are either unknown or from uncharacterized open reading frames (ORFs), or factors with known functions but whose involvement in AID-mediated processes is unclear. A list of all factors identified in triplicate, as well as their predicted functions, is presented in Table 3.1.

3.2 RING Finger Protein 126 (RNF126)

Because of the potential interesting characteristics of one of the unknown ORFs identified in the screen, Ring Finger Protein 126 (RNF126), further experiments were conducted to understand its influence on AID and AID-mediated events. RNF126 is a relatively small protein of approximately 310 amino acids and is well conserved throughout evolution. In fact, there exists a 62% similarity between human and zebrafish forms of the protein. Though virtually nothing is known about the role of RNF126, three domains exist that are

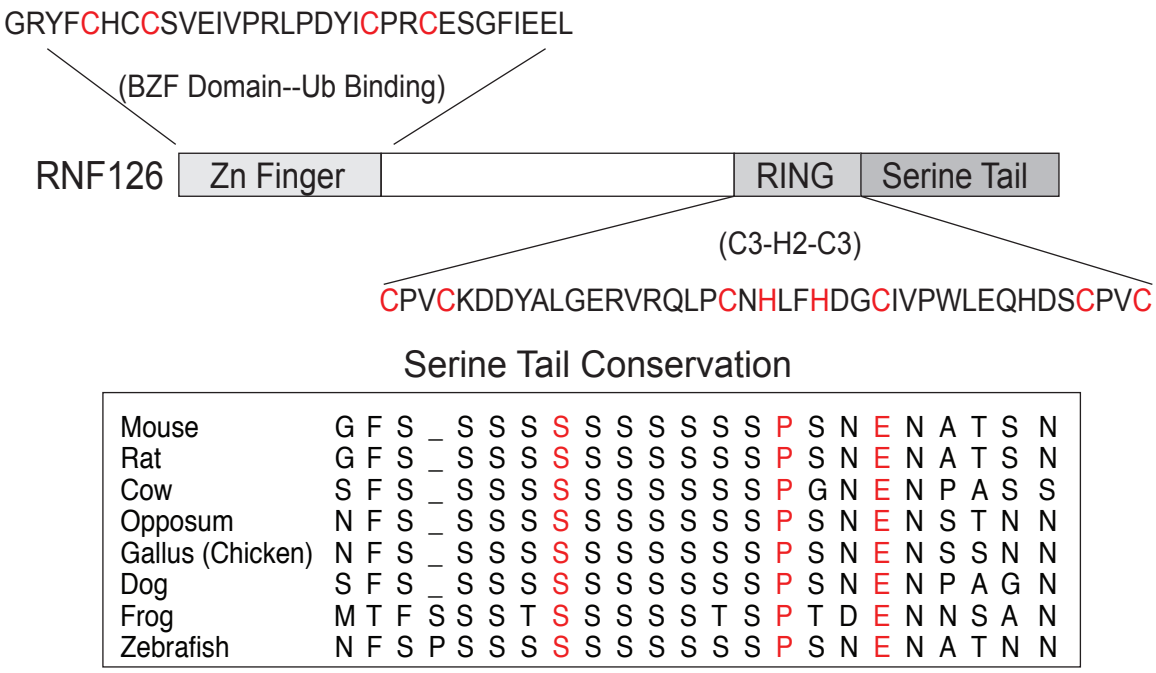
Table 3.1 List of identified putative AID-Cofactors. A list of all 36 candidates that were cloned in triplicate. The name and accession number of each gene are presented, as well as the number of times each gene appeared in the screen (# cloned). 30/36 hits were verified to interact by Co-IP (last column—Y=yes, N=no). In addition, the best description of the function of the gene found in the literature is presented. For many of the genes, the functions are not known and descriptions of other aspects of the biology of the gene are given. AID-interacting factors found by other screens are highlighted.

Factor	Accession #	Description	Function	#cloned	Co-IP?
RNA Processing					
hnRNP-M	NP_005959	Splicing? Coating transcript?	Unknown	17	Y
hnRNP-A0	NP_006796	Splicing? Coating transcript?	Unknown	35	Y
DDX21	NP_004719	Nucleolar RNA helicase	Unknown	10	Y
DDX19	NP_001014449	Nucleolar RNA helicase	Unknown	10	Y
LGTN	NP_008824	Sui1; RNA binding	Unknown	16	Y
CPSF73	NP_057291	Cleavage and polyadenylation specificity factor	PolyA Tail Addition	6	Y
XRN2	NP_036387	5'>3' exoribonuclease	mRNA term.	6	Y
EEF1A1	NP_001393	Translation initiation factor	Translation	9	Y
RBM39	NP_909122	U2AF homolog; splicing	splicing	14	Y
NIP30	NP_079222	Nucleolus-associated	Unknown	3	N
NOL11	NP_056277	Nucleolus-associated	Unknown	12	N
CTNNBL1	NP_110517	mRNA splicing	Import? Splicing?	6	Y
Nuclear Import/Export					
Nup93	NP_055484	Nucleoporin	Nuclear Pore Complex	12	Y
Importin Subunit β-1	NP_002256	Importin beta family	Nuclear import	11	Y
DNA Repair					
RAD51	NP_002866	Rad51 Homolog 1 Isoform 1	DNA repair	7	Y
RRM2B	NP_001025	Ribonucleoside Reductase	Deoxyribonucleotide Synthesis	8	Y
IER5	NP_057629	Novel repair factor?	Unknown	6	Y
FEN1	NP_004102	Flap endonuclease	DNA repair	6	Y
General Cellular Biology					
Actin	NP_001605	Gamma actin isoform	cytoskeleton	12	Y
RDH11	NP_057110	oxidoreductase	Nucl. Metabolism	6	N
PGAM1	NP_002620	Glycolysis/metabolism	Nucl. Metabolism	13	Y
IMPDH2	NP_000875	IMP dehydrogenase	Nucl. Metabolism	35	N
Protein Modification					
UbcH7	NP_055904	Ubiquitin ligase	Prot degradation	6	Y
PPP2R1A	NP_055040	Protein phosphatase	Reg subunit	6	Y
AID Biology					
AICDA	NP_065712	Cytidine deaminase	DNA mutation	8	Y
RPA1	NP_002936	Replication protein A	AID cofactor	15	Y
Chaperone Proteins					
TCP1-eta	NP_001009570	T-Complex Protein 1	chaperone	23	Y
BiP/HSPA5	NP_005338	78 kDa glucose-regulated protein precursor	Chaperone;HSP70 family member	31	Y
HERPUD2	NP_071768	homocysteine-responsive endoplasmic reticulum-resident ubiquitin-like domain member 2 protein	Upregulated during unfolded protein response	9	Y
Unknown Function					
C14orf94	NP_060285	Unknown	Unknown	7	Y
C22orf28	NP_055121	Unknown	Unknown	6	Y
CXorf9	NP_061863	Lymphoid sp.; SH2 domain	Unknown	18	N
ZNF44	NP_057348	Zn finger; DNA binding	Unknown	6	N
DAZAP2	NP_055579	Zn finger; DNA binding	Unknown	5	Y
ZMIZ2	NP_113637	MIZ-type Zn finger	Unknown	7	Y
RNF126	NP_919442	Ring Domain Containing Protein	E3 Ubiquitin Ligase	6	Y

very suggestive of function: an N-terminal Zinc Domain, a C-terminal RING Finger Domain and a very C-terminal stretch of serine residues (Figure 3.2). The C-terminal serine tail contains approximately 12 serines that are conserved from its emergence in zebrafish to human (Figure 3.2). Few other proteins contain such a domain and its conservation throughout evolution suggests that it plays an important role in the mechanism of RNF126. Just upstream of the C-terminal serine stretch lies a C3H2C3-type RING domain, which has been shown to be necessary for auto-ubiquitination in *in vitro* assays (Burger et al., 2006). Lastly, the N-terminal Zinc domain shows significant homology (69%) to a zinc finger domain on a related E3 ligase, BCA2, which has been identified as a BZF domain, a novel ubiquitin binding domain (Amemiya et al., 2008). The conservation of these domains prompts intense investigation to determine how each play a role in the physiology of RNF126.

3.3 Identification and validation of RING Finger Protein 126 as an AID-interactor

As with the other potential interacting candidates, the direct interaction of RNF126 with AID was validated by exogenous expression in both bacterial and mammalian systems. First, co-expression of a hexa-histidine tagged RNF126 and a Flag tagged AID followed by affinity purification shows that AID and RNF126 interact. His-RNF126 was purified on a Ni-NTA column and eluted with increasing amounts of imidazole. Flag-AID co-eluted with RNF126, but did not appear in elution fractions that did not contain His-RNF126 (Figure 3.3).



Serine Tail Conservation

Mouse	G	F	S	_	S	S	S	S	S	S	S	S	S	S	P	S	N	E	N	A	T	S	N
Rat	G	F	S	_	S	S	S	S	S	S	S	S	S	S	P	S	N	E	N	A	T	S	N
Cow	S	F	S	_	S	S	S	S	S	S	S	S	S	S	P	G	N	E	N	P	A	S	S
Opposum	N	F	S	_	S	S	S	S	S	S	S	S	S	S	P	S	N	E	N	S	T	N	N
Gallus (Chicken)	N	F	S	_	S	S	S	S	S	S	S	S	S	S	P	S	N	E	N	S	S	N	N
Dog	S	F	S	_	S	S	S	S	S	S	S	S	S	S	P	S	N	E	N	P	A	G	N
Frog	M	T	F	S	S	S	T	S	S	S	S	S	T	S	P	T	D	E	N	N	S	A	N
Zebrafish	N	F	S	P	S	S	S	S	S	S	S	S	S	S	P	S	N	E	N	A	T	N	N

Figure 3.2 RING Finger Protein 126. Cartoon of RNF126 domain structure. The position and sequences of the RING domain and Zinc finger domain are shown. The insert shows that the amino acid composition of the C terminal serine-rich domain is evolutionarily conserved.

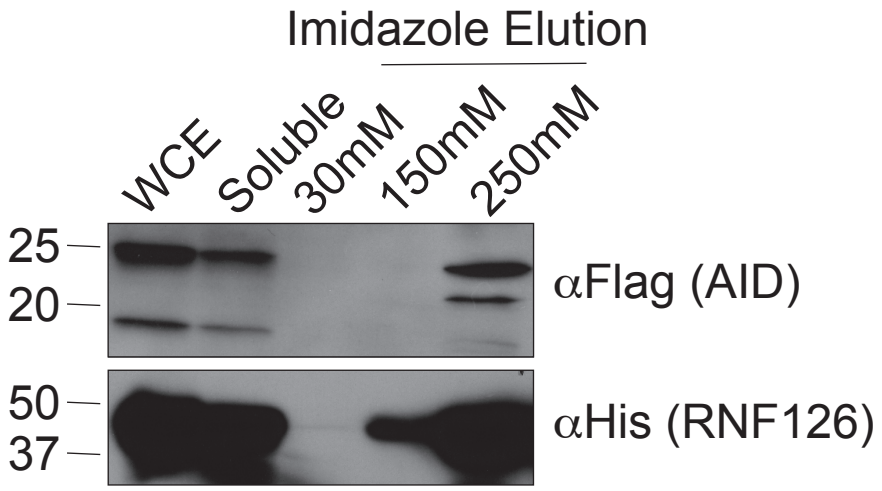


Figure 3.3 RNF126 interacts with AID in bacterial cells. AID co-purifies with RNF126 upon co-expression and purification in E.coli. Flag tagged AID was co-expressed with His-tagged RNF126. Purification of RNF126 on a Talon cobalt column and elution with imidazole reveals that AID co-elutes with RNF126.

In order to better characterize the interaction of RNF126 and AID, the regions on both AID and RNF126 that were involved in protein-protein binding were determined. Because AID is a small protein, truncation of the protein often leads to mis-folding and thus misleading results. To get around this, an approach developed by Conticello *et. al.* was used (Conticello et al., 2008). A set of chimera expression constructs, interspersing regions of the APOBEC family member, APOBEC2, within AID were used to determine the role individual domains of AID play in its interaction with RNF126. Co-expression of His-tagged AID and FLAG-tagged RNF126 in bacterial cells followed by co-affinity purification experiments revealed that, whereas chimeras A, B and D interacted well with RNF126, Chimera C did not (Figure 3.4). This region replaces residues 88-116 of AID with the comparable region of APOBEC2, and contains part of the zinc-coordinating motif (Figure 1.8), as well as the region of AID (aa 106-116) that has been found to confer specificity for AID hotspots. This region, when replaced with a comparable region from other APOBEC family members, alters the mutation preference of AID (Kohli et al., 2009).

To ascertain what regions of RNF126 are necessary for its interaction with AID, N-terminal truncations of Flag-tagged RNF126 were co-expressed with 6X-His tagged AID. AID was purified on a Ni-NTA column, followed by elution with increasing concentrations of imidazole. The most severe truncation, containing only the RING finger domain and the C-terminal serine stretch, was sufficient to interact with AID (Figure 3.5). This data fits well with the observed interaction of

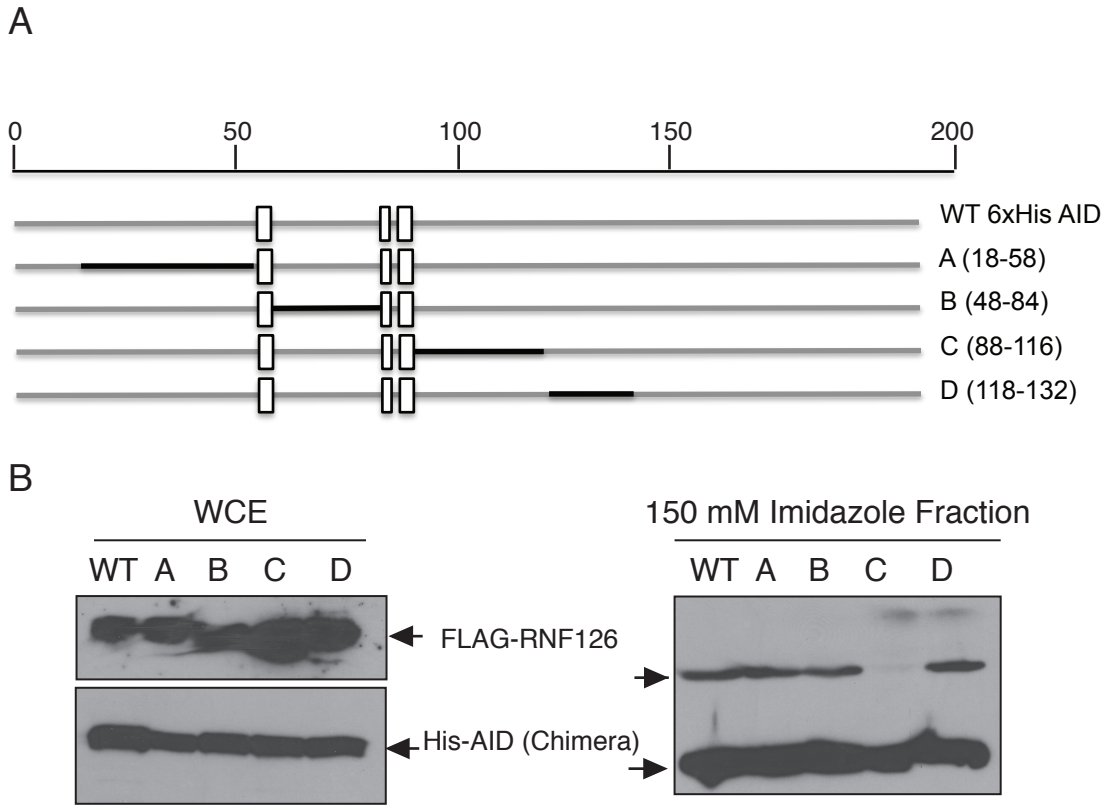


Figure 3.4 Delineation of interaction domain on AID. To define the regions of AID that interact with RNF126, a set of chimera expression constructs as in Conticello *et al.* were made, interspersing regions of the APOBEC family member, APOBEC2, within AID. Schematics are shown in the figure: the black line represents APOBEC2 sequence within the AID sequence, shown in grey; in addition, white boxes indicate the catalytic residues. Co-expression of Flag-tagged RNF126 and His-tagged APOBEC2/AID chimeras in bacterial cells followed by purification on a BD Talon Cobalt column demonstrates that whereas chimeras A, B and D co-purify with RNF126, Chimera C does not. Chimera C replaces the region of AID spanning residues 88-116 with the comparable region of APOBEC2. This suggests that this region of AID is the primary site of interaction between AID and RNF126. Left panel: anti-FLAG and anti-His blots to demonstrate equivalent expression in whole cell extract of RNF126 and AID (and its chimeras with Apobec2), respectively. Right panel: the membrane was first blotted with anti-FLAG to detect RNF126, followed by incubation with anti-HIS antibody to detect the AID/APOBEC2 chimera proteins. Co-purification of RNF126-FLAG and 6xHisAID/APOBEC2 was assessed in the 150mM imidazole fraction.

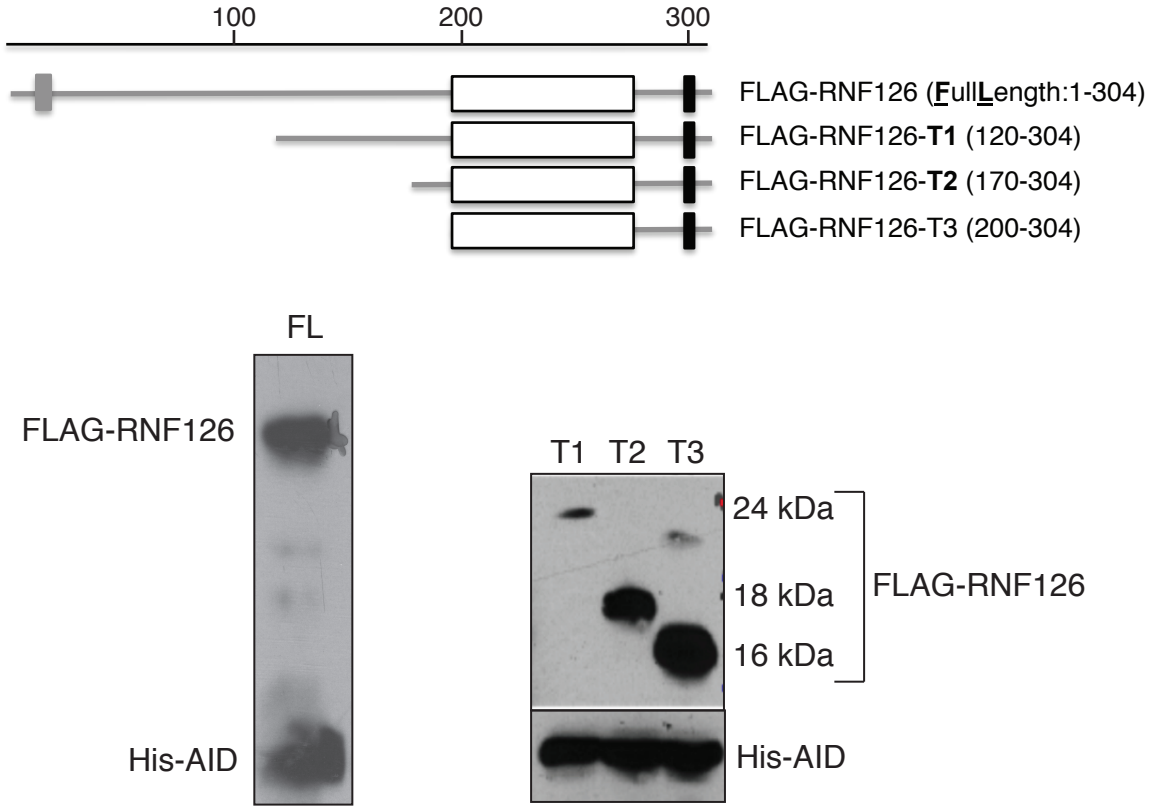


Figure 3.5 Delineation of interaction domain on RNF126. Truncations of RNF126 reveal that the RING domain interacts with AID.
Flag-tagged full-length RNF126 (FL) and various truncations of RNF126 (T1, T2, T3 - schematic shown) were co-expressed with His-tagged full-length AID (gray box-N terminal Zinc finger of RNF126; white box-the RING domain; black box-the serine tail domain) . Purification of His-AID on a BD Talon cobalt column (BD Biosciences) reveals that N-terminal truncations of RNF126 are still able to interact with AID. In addition, the interaction appears stronger as more of the N-terminus is deleted, which could result because of increased solubility of some of the truncation constructs. Importantly, the last truncation (200-end), which contains the RING domain and the serine tail, can interact strongly with AID. This suggests that AID can interact with isolated RING domains in general. Indeed, the RING domain of the E3 ligase, Mdm2, can interact with AID in a yeast-two hybrid screen as well as in our screen; however AID does not interact with full-length Mdm2.

the RING domain of the E3 ligase Mdm2 in a yeast-two hybrid screen. Thus, it is likely that it is also the RING domain of RNF126 that primarily interacts with AID (MacDuff et al., 2006).

To assure that these two proteins interact in the context of a mammalian cell, FLAG-tagged AID and Hemagluttin (HA)-tagged RNF126 were expressed in the mammalian 293T HEK expression system. Purification of either AID or RNF126 was completed by precipitation with anti-FLAG or anti HA antibodies. These studies, again, reveal that AID and RNF16 co-precipitate and thus interact (Figure 3.6). In addition to demonstrating an interaction, these studies also show that RNF126 and AID reside within the same cellular compartment.

Because AID is primarily expressed in B cells for the purpose of antibody diversification (Alt et al., 2013; Delker et al., 2009; Teng and Papavasiliou, 2007; Vuong and Chaudhuri, 2012; Xu et al., 2012), it was important to determine if RNF126 is expressed in a physiological setting alongside AID. Utilizing both quantitative RT-PCR and Western blotting, it is clear that RNF126 is expressed in primary B cells upon stimulation *in vitro* to undergo CSR (Figure 3.7 A, B). While, naïve, unstimulated B cells do express RNF126, there appears to be a slight increase at both the mRNA and protein level upon stimulation. In addition, the expression of RNF126 is not dependent on the presence of AID or the formation of AID-mediated breaks in these cells because a similar expression profile is shown in *aicda*$^{-/-}$ B cells (Figure 3.7 B).

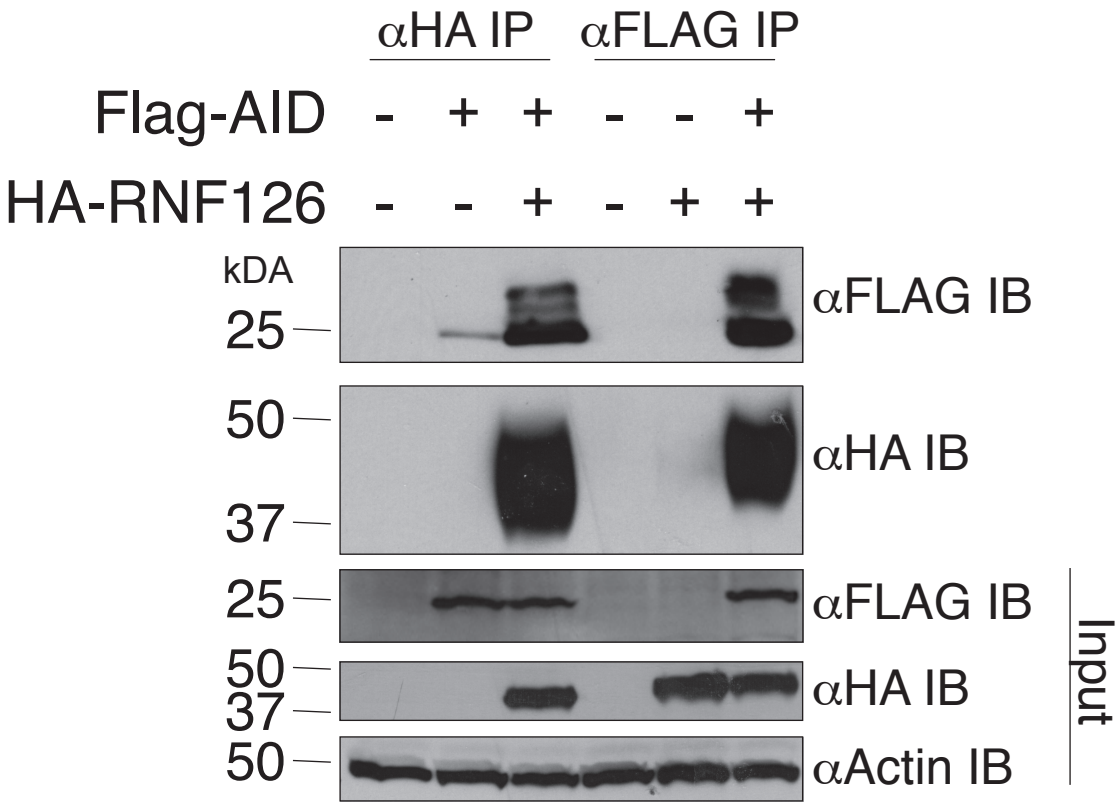


Figure 3.6 RNF126 interacts with AID in mammalian cells. RNF126 interacts with AID in a HEK 293T mammalian system. Flag-tagged AID and HA-tagged RNF126 were co-expressed. Affinity purification of either tag resulted in the purification of the binding partner only when both proteins were expressed.

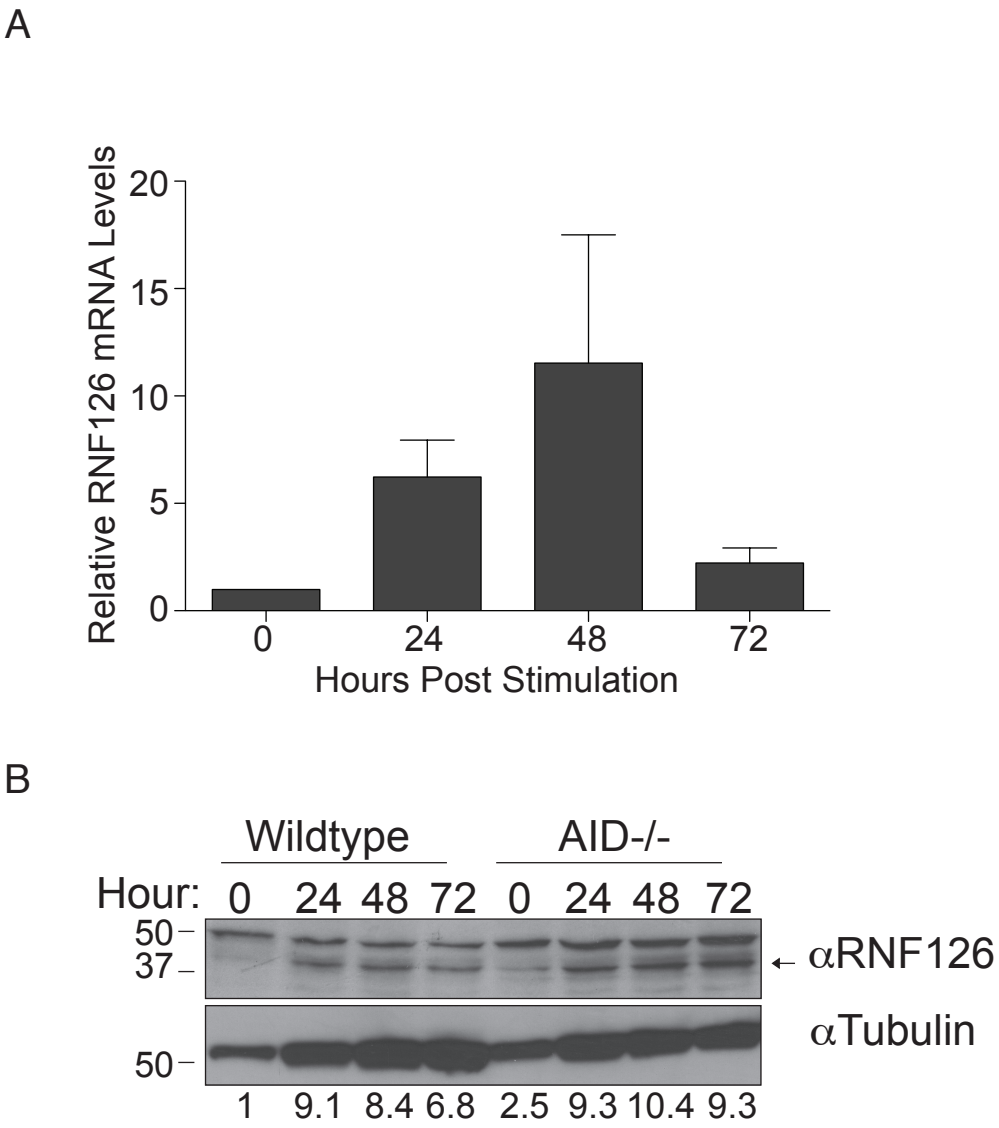


Figure 3.7 RNF126 is expressed in primary B cells. (**A**) Expression of RNF126 is induced in switching B cells after stimulation for class-switch recombination as assessed by qRT-PCR (normalized to CD19 levels). (**B**) RNF126 protein levels increase in murine B cells after stimulation to undergo class switch recombination in an AID-independent manner. Arrow denotes RNF126 band (Tubulin: loading control). The values beneath the blot denote a quantification of RNF126 protein levels, normalized to Tubulin.

3.4 RNF126 ubiquitinates AID

Because RNF126 contains a RING finger domain, a domain commonly found in E3 ubiquitin ligases, and RNF126 has already been shown to be able to auto-ubiquitinate *in vitro* (Burger et al., 2006), it was plausible that RNF126 could also modify AID in addition to simply interacting. In order to test this, a HEK 293T ubiquitination assay was developed. In this assay, AID is expressed with and without co-expression of RNF126 and/or ubiquitin. Upon addition of RNF126, but not without, AID is ubiquitinated (Figure 3.8). Interestingly, these experiments exhibited neither a single slow migrating band, representative of a mono-ubiquitination event, nor a higher molecular weight smear, representative of poly-ubiquitination events. The pattern of ubiquitination observed, a ladder of approximately three to four slower migrating bands, could be representative of either a small poly-ubiquitin chain addition on one residue of AID, or multiple-mono-ubiquitination events on different residues of AID. To determine which of these scenarios is occurring, two mutants of ubiquitin, a Lysine 48 mutation to Arginine (K48R) and a Lysine 63 mutation to Arginine (K63R) were co-expressed with AID and RNF126. Poly-ubiquitin chains generated through linkages at lysine 48 are typically found in proteins targeted for degradation by the proteasome; conversely, poly-ubiquitin chains generated through linkages at lysine 63 are commonly found on protein components of signaling modules (Pickart, 2001b). Although other poly-ubiquitin linkages have been seen, K48 and K63 are the most common. Expression of both K48R and K63R ubiquitin moieties in the 293T

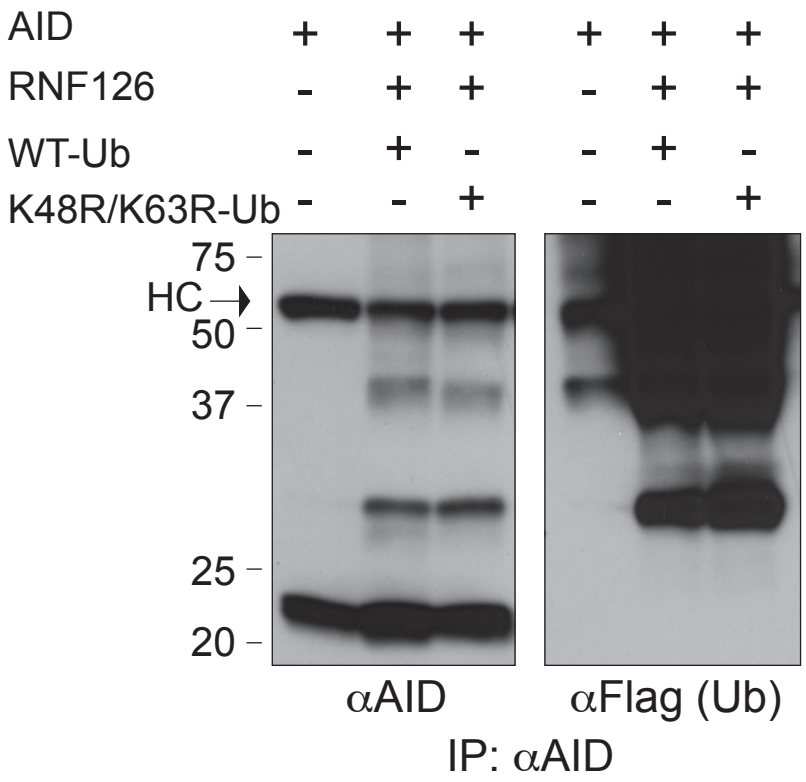


Figure 3.8 RNF126 ubiquitinates AID in HEK 293T Cells. AID is ubiquitinated by RNF126 in HEK 293T cells. Co-expression of hAID with HA-tagged RNF126 and Flag-tagged ubiquitin results in the formation of ubiquitylated AID, even in the presence of K48R/K63R mutant ubiquitin. RIPA extracts were prepared and AID immunoprecipitated with an anti-AID antibody. An anti-AID and an anti-FLAG (ubiquitin) blot are shown. The heavy chain band is marked as “HC.”

assay resulted in a similar three to four band laddering pattern, suggesting that this laddering pattern is not poly-ubiquitination (at least not of linkages at K48 or K63), but rather multiple mono-ubiquitination events (Figure 3.8). It has been shown that increased concentrations of E3 ligase can convert a mono-ubiquitination event into multiple mono-ubiquitination events (Li et al., 2003). Because this assay relies on the overexpression of RNF126 in the 293T system, it is possible that the laddering pattern observed is due to increased levels of RNF126 over what would be physiological levels of the protein.

To overcome this problem, an *in vitro* ubiquitination assay was developed and optimized using purified recombinant RNF126 and AID, as well as other necessary components of the ubiquitination cascade. Because UbcH5b was already identified as a compatible E2 ubiquitin conjugating enzyme for RNF126 (Burger et al., 2006), this E2 enzyme was chosen for these *in vitro* assays. Further experiments confirmed that only UbcH5b and Ubch5C, although to a lesser extent, are able to support *in vitro* ubiquitination of AID by RNF126; however, other E2s tested were not (Figure 3.9).

Just as in the 293T ubiquitination assays, ubiquitination of AID occurred in an RNF126-dependent fashion; furthermore, the addition of ubiquitin is dependent on all other components of the ubiquitination cascade, including Ube1 (E1), UbcH5b (E2), ubiquitin and energy (ATP) (Figure 3.10 A). While it was possible to detect a slight laddering of AID upon ubiquitination, the band representative of the addition of a single ubiquitin moiety to AID appears more

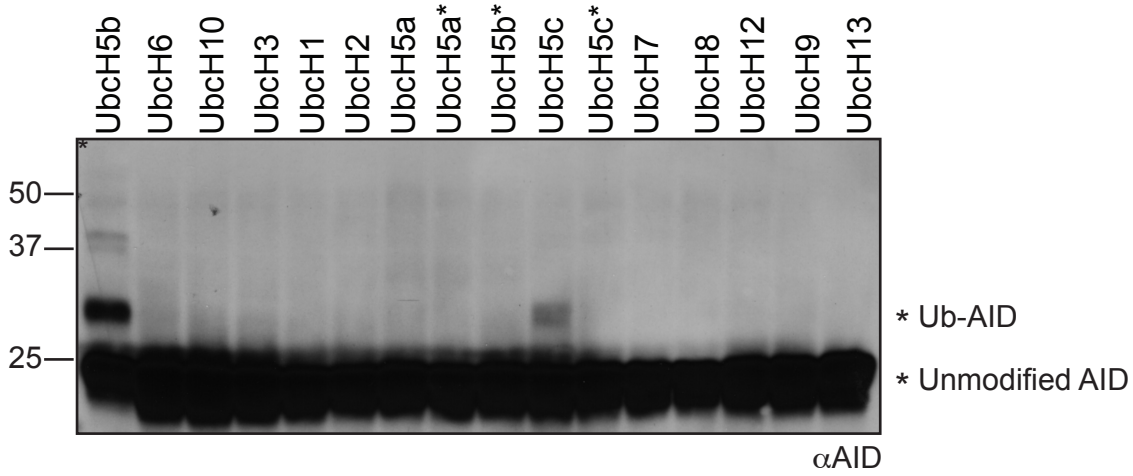


Figure 3.9 UbcH5b and UbcH5c support ubiquitination of AID. UbcH5b and, to a lesser extent, the highly related UbcH5c are able to support *in vitro* ubiquitination of AID by RNF126. A panel of human E2 enzymes is tested in the *in vitro* ubiquitination assay. An anti-AID immunoblot is shown; unmodified and Ub-AID are marked with asterices. The asterix used to mark UbcH5a,b,c denotes catalytically inactive forms of the enzyme.

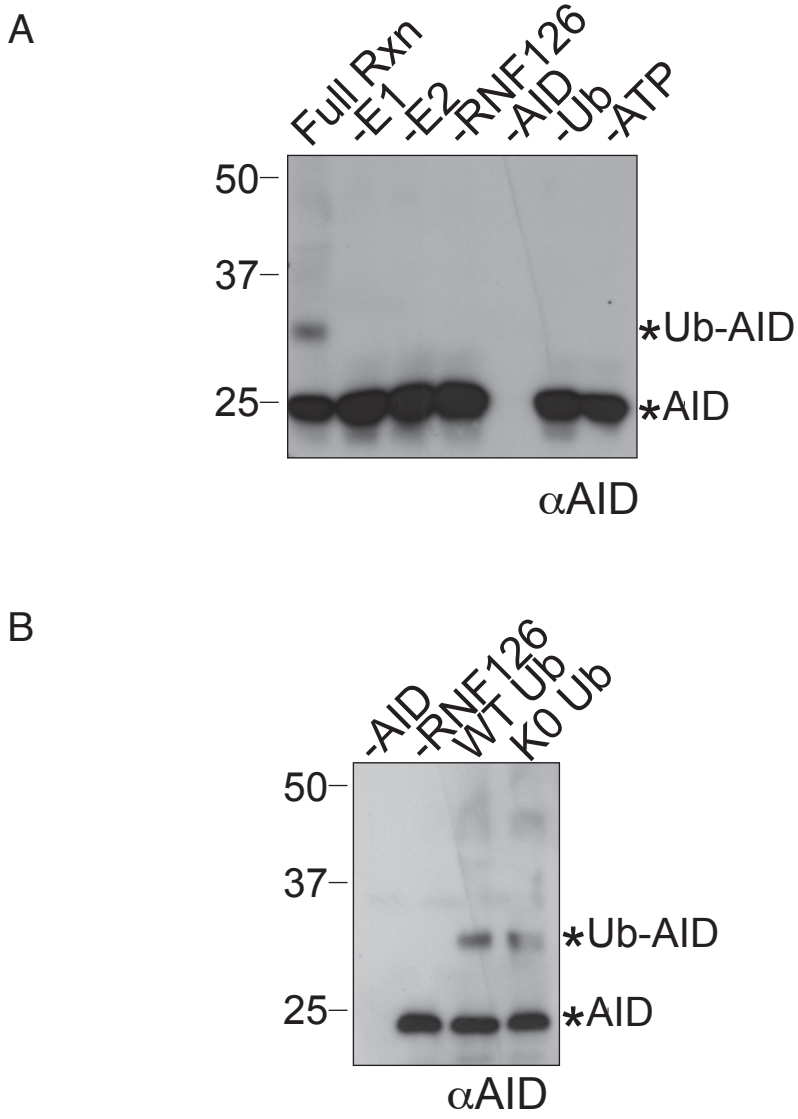


Figure 3.10 RNF126 mono-ubiquitinates AID *in vitro*. AID is ubiquitinated by RNF126 in vitro. Purified components of the ubiquitin reaction are incubated at 37°C. (**A**) All components are necessary for ubiquitylation to occur (lane 1).The absence of individual components (lane 2-7) prevents ubiquitylation. (**B**) RNF126 mono-ubiquitylates AID. Addition of a mutant ubiquitin in which all lysines are mutated to arginine (K0 Ub) shows that the laddering is representative of multiple mono-ubiquitylation events. anti-AID immunoblots are shown.

intense than even higher migrating bands, suggesting one main mono-ubiquitination event with the possibility of additional ubiquitination events on the same AID molecule. To confirm this, *in vitro* ubiquitination assays were performed using a mutant form of ubiquitin in which all seven lysines are mutated to arginine (K0). This mutant is unable to form poly-ubiquitin chains. The same slight laddering pattern of AID ubiquitination was observed using K0 ubiquitin as compared to wildtype, thus confirming that RNF126 mono-ubiquitinates AID (Figure 3.10 B).

Because the RING domain of E3 ligases is essential for substrate ubiquitination and because it has already been established that the RING domain of RNF126 interacts with AID, it was important to determine if mutation of the RING Domain resulted in a loss of ubiquitination. A mutant version of RNF126 (RNF126*), which contained mutations at two of the zinc-coordinating cysteines in the RING domain (C229A/C232A), was tested in the 293T ubiquitination assay. Unlike wildtype RNF126, RNF126* was unable to ubiquitinate AID, providing direct evidence for the necessity of the RING domain (Figure 3.11).

Though ubiquitination most commonly occurs on lysine residues, ubiquitination has also been observed on serine, threonine and cysteine residues (Cadwell and Coscoy, 2005; Wang et al., 2007). To determine the residue on AID that is ubiquitinated by RNF126, lysine to arginine mutants of AID were subjected to the 293T ubiquitination assay. Single lysine to arginine mutants, as well as grouped lysine to arginine mutants, were still able to be ubiquitinated (Figure 3.12

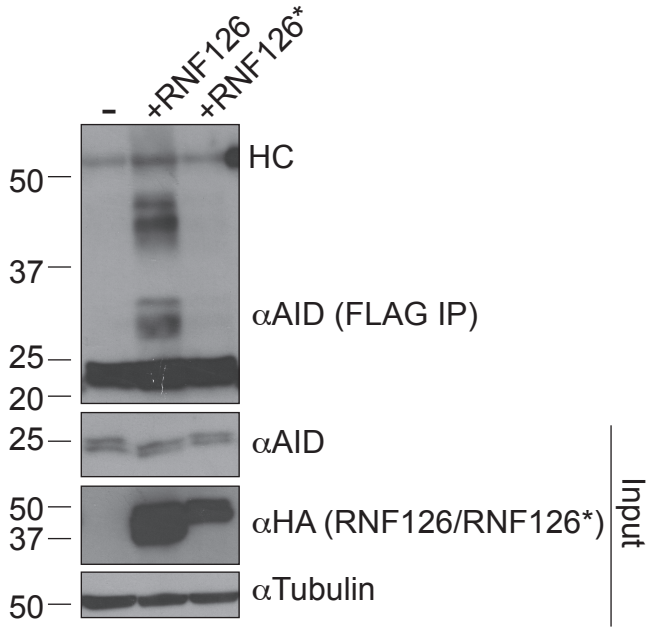


Figure 3.11 The RING domain of RNF126 is necessary for AID ubiquitination. Mutations in the RING domain of RNF126 abrogate the ability to modify AID. Mutant RNF126 (RNF126*), which contains cysteine to alanine mutations at the first two zinc-coordinating cysteines in the RING domain of RNF126 (C229A and C232A) is catalytically inactive. HA tagged wildtype RNF126 or RNF126* are co-expressed in 293T cells with Flag-AID. Wildtype RNF126 is able to induce ubiquitylation of AID (lane 2), but RNF126* is not (lanes 3). Expression of Flag-AID without exogenous ligase is shown as a control (lane 1). The top panel shows an αAID western blot of an αFLAG IP. The bottom three panels show levels of Flag-AID (αAID), HA-RNF126/RNF126* (αHA) and tubulin (αTubulin) in the input/whole cell extract. The heavy chain band from the antibody used in the IP is denoted as “HC.”

A). However, mutation of all lysines to arginine (AID K0) resulted in a loss of ubiquitination (Figure 3.12 A), suggesting that a lysine residue is ubiquitinated. However, reconstitution of *aicda*$^{-/-}$ B cells with AID K0 could not rescue CSR (Figure 3.12 B). Thus, it cannot be ruled out that structural changes to the protein are the cause of the loss of ubiquitination. In addition, mass spectrometric analysis of *in vitro* ubiquitinated AID was attempted; however, due to limited material, full coverage of the protein was not achieved and the ubiquitination site not identified.

3.5 RNF126 displays specificity toward AID

Thus far it has been established that (1) RNF126 can interact with AID, and (2) RNF126 can act as an E3 ligase to modify AID. To become more certain that these events were not non-specific, it was important to establish that RNF126 possesses specificity for AID as a substrate. To determine this, the ability of RNF126 to ubiquitinate AID was first compared to two other ligases expressed in B cells, RNF8 and BCA2.

RNF8 has been shown to play a distinct role in mediating the recruitment of protein factors necessary for double-strand break repair during class-switch recombination (Ramachandran et al., 2010). To assess the ability of RNF8 to modify AID, these factors were co-expressed in the HEK 293T ubiquitination assay. While co-expression of RNF126 with AID in 293T cells results in the ubiquitination of AID, co-expression of RNF8 and AID in this setting does not (Figure 3.13 A, left panel). This finding was confirmed using the *in vitro*

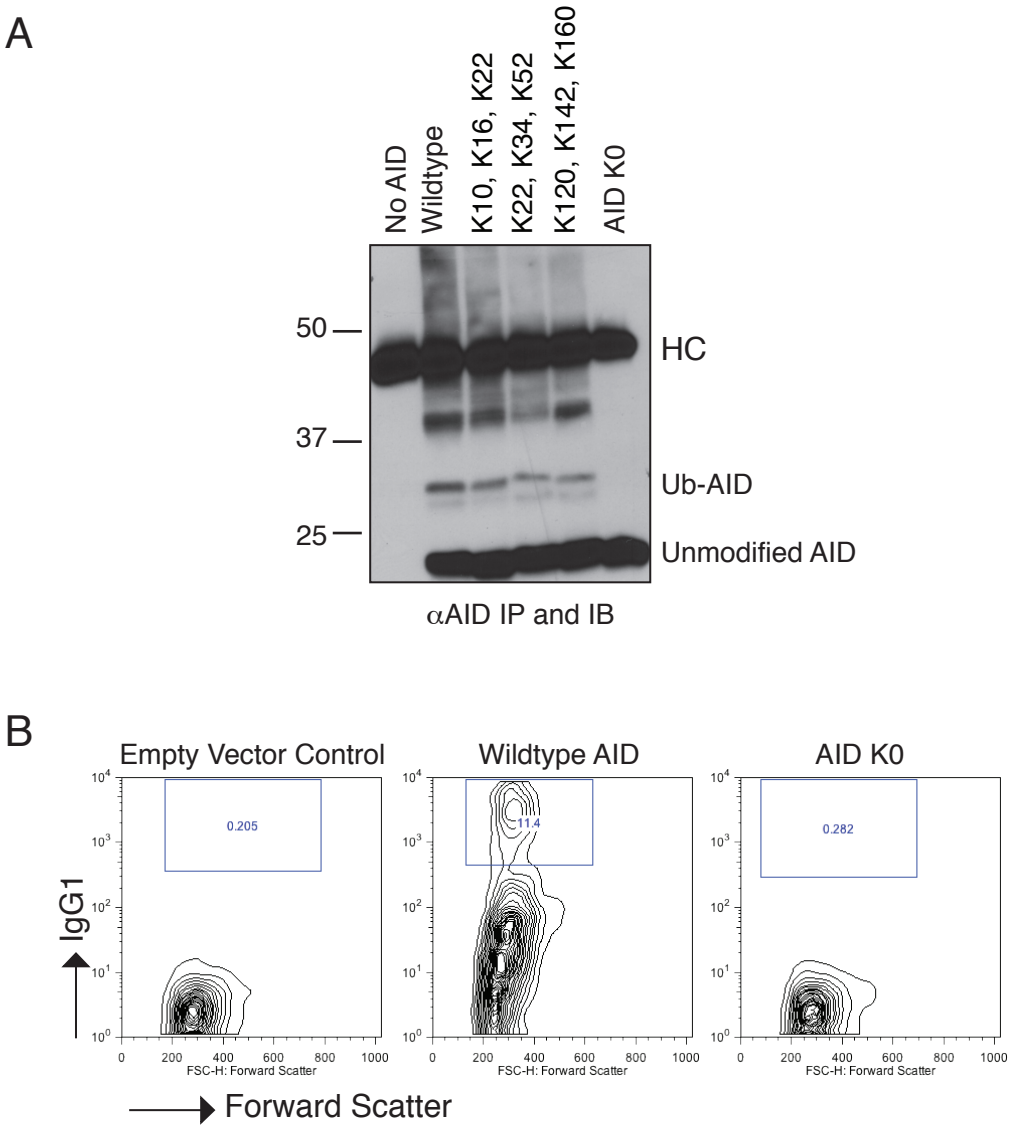


Figure 3.12 Lysineless AID is not ubiquitinated by RNF126. (**A**) Mutant forms of AID containing clusters of lysine-to-arginine mutations were subjected to the 293T ubiquitination assay. Only mutation of all lysines on AID to arginine (AID K0) resulted in a loss of RNF126-mediated ubiquitination. An anti-AID IP and immunoblot are shown. The unmodified AID band, as well as the ubiquitinated form are shown. Lane 1 contains a "No AID control." "HC" denotes the heavy chain of the antibody used for immunoprecipitation. (**B**) *aicda-/-* B cells were virally transduced with either empty vector, wildtype AID or AID K0 and assayed for CSR. While wildtype AID rescued CSR (middle panel), AID K0 did not (right panel). Representative FACS plots are shown. Y axis = surface IgG1, X axis = Forward Scatter.

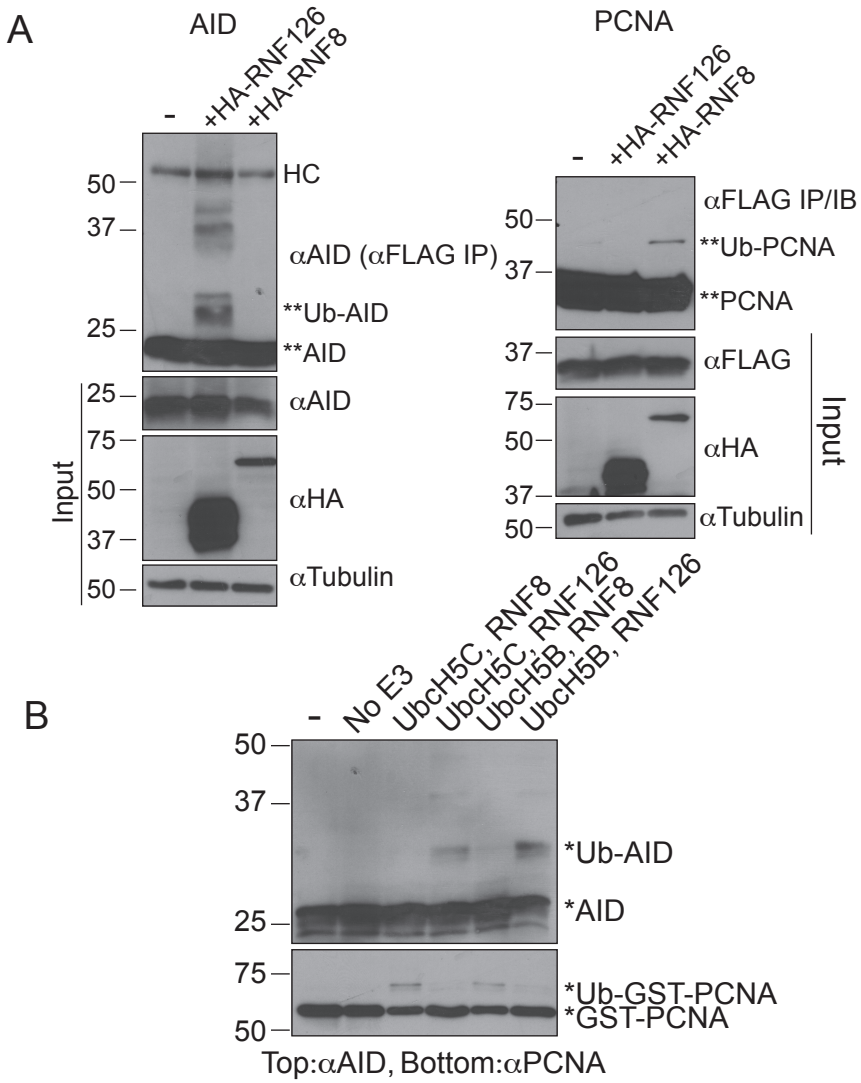


Figure 3.13 RNF126 selectively ubiquitinates AID when compared to RNF8. (**A**) 293T cells were transfected with Flag-AID, alone or with HA-RNF126 or HA-RNF8 (left panel), and Flag-PCNA, alone or in combination with the same ligases (right panel). Co-expression of RNF126, but not RNF8, results in AID ubiquitylation. Ubiquitylation of PCNA occurs upon co-expression of RNF8 (lane 3), but not RNF126 (lane 2). Both AID and PCNA were immunoprecipitated with αFlag and blotted with αAID and αFlag, respectively (top panels). Western blots of AID (αAID), PCNA (αFlag), RNF126/RNF8 (αHA) and tubulin (αTubulin) of input (whole cell extract) are shown in the bottom three panels. Asterices mark the unmodified and modified bands of AID and PCNA and "HC" denotes the heavy chain band. Lane 1 of each panel represents the expression of the substrate without exogenous ligase. (**B**) GST-RNF8 and GST-PCNA were used as an alternate ligase and substrate in the in vitro ubiquitylation assay. Reactions were carried out with either UbcH5b or UbcH5c as the E2 enzyme. In both cases, RNF8 selectively ubiquitylates PCNA and not AID (lanes 3,5) and RNF126 selectively ubiquitylates AID and not PCNA (lanes 4,6). Lane 1 excludes all components except the substrate and Lane 2 only excludes the E3 ligase. .

ubiquitination assay, in which it was clear that mono-ubiquitination of AID required RNF126 (Figure 3.13 B).

RNF126 shows incredibly high homology to another RING domain containing E3 ligase, BCA2. BCA2 shows 46% identity to RNF126 in overall amino acid content and 75% homology within the RING domain, lacking only the serine tail, which is unique to RNF126 (Burger et al., 2006) (Figure 3.14 A). Further, it is the RING domain of RNF126 that has been shown to bind AID (Figure 3.5). Thus, comparison of RNF126 with BCA2 in their ability to ubiquitinate AID provided a good test for the specificity of RNF126. In addition, BCA2 is expressed in B cells during CSR, providing physiological relevance (RNA-Seq data from (Fritz et al., 2013)). Despite the similarities between these two proteins, co-expression of BCA2 with AID in HEK 293T cells does not lead to AID ubiquitination (Figure 3.14 B). Again, this provides evidence for the specificity of the RNF126-AID interaction and modification.

In addition to these studies, which compare RNF126 to other E3 ligases, it was also important to test the ability of RNF126 to ubiquitinate alternate substrates. Though it is fairly common for ligases to ubiquitinate multiple substrates in different cellular contexts (Mailand et al., 2007; Zhang et al., 2008), it was still interesting to determine if RNF126 could ubiquitinate other proteins that are involved in CSR/SHM. One such substrate is Proliferating Cell Nuclear Antigen (PCNA), a homo-trimeric ring that encircles DNA and is necessary for DNA replication and repair in the context of AID mediated reactions (Langerak et al., 2009; Roa et al., 2008). Specifically, mono-ubiquitination of PCNA at lysine

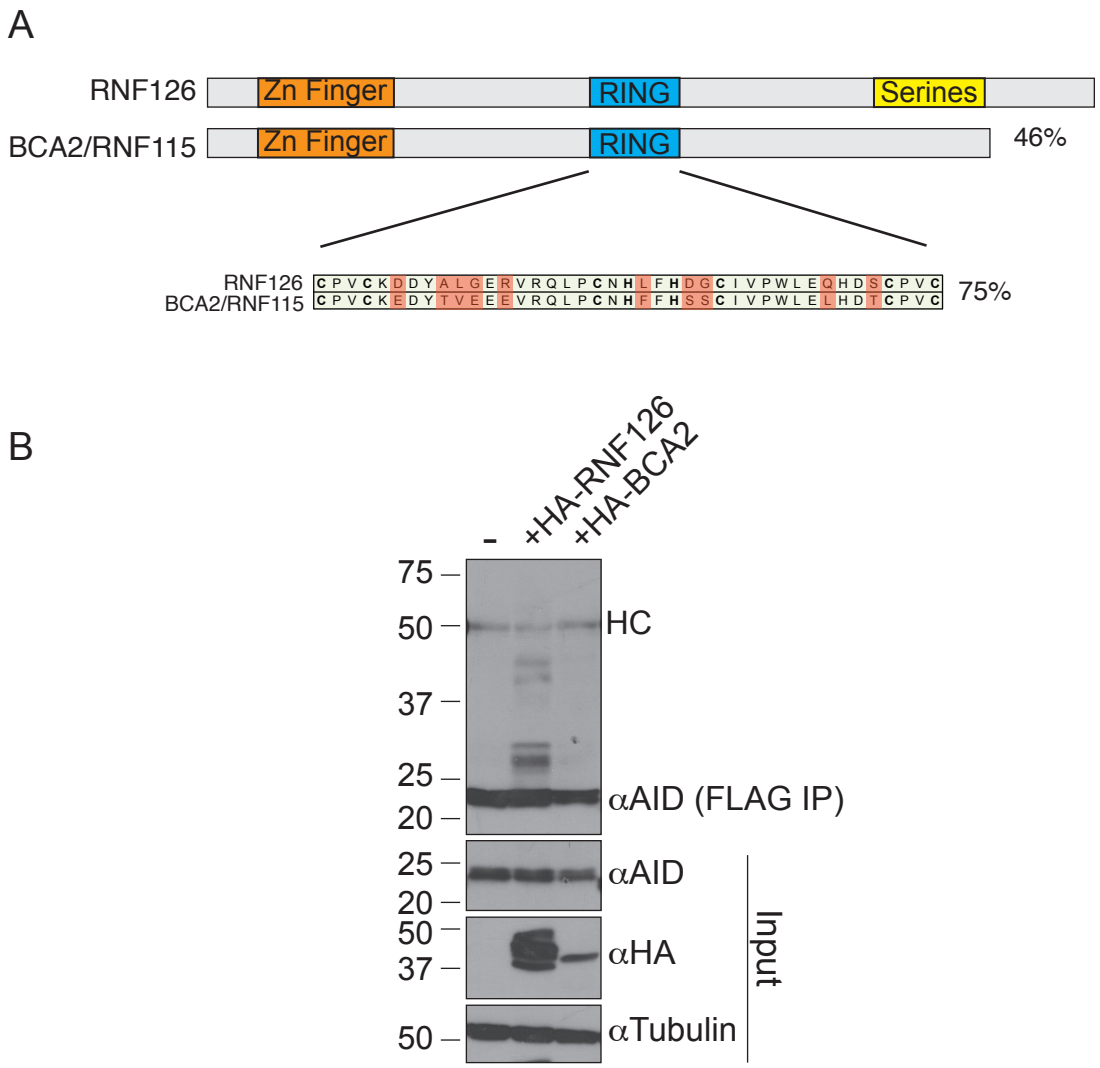


Figure 3.14 RNF126 selectively ubiquitinates AID when compared to BCA2. BCA2, a homologous E3 ligase, cannot ubiquitylate AID. AID is expressed in 293T cells alone (lane 1) or in the presence of either RNF126 (lane 2) or BCA2 (lane 3). Only co-expression with RNF126, but not BCA2, results in ubiquitylation. The top panel represents an αFlag IP and αAID immunoblot. Western blots of AID (αAID), RNF126/BCA2 (αHA) and tubulin (αTubulin) in the input/whole cell extract are shown in the bottom three panels. The heavy chain band is marked with “HC.”

164 (K164) is believed to recruit the error-prone DNA polymerase, Pol η, resulting in the accumulation of mutations at non-C:G base pairs within the variable region and switch regions of the immunoglobulin locus (Faili et al., 2004). Mono-ubiquitination of PCNA has been shown to be carried out by two different E2/E3 ubiquitin conjugating enzyme/ligase pairs: Rad6/Rad18 (Hibbert et al., 2011; Hoege et al., 2002) and UbcH5c/RNF8 (Zhang et al., 2008). However, additional enzymes are thought to allow for residual mono-ubiquitination even in the absence of these canonical modifying complexes. The ability of RNF8 and RNF126 containing complexes to ubiquitinate PCNA was determined in both the 293T ubiquitination assay, as well as the *in vitro* assay. In both assays it was clear that RNF8 had activity toward PCNA, but RNF126 was completely inactive (Figure 3.13 A, B).

Though every possible candidate substrate for RNF126, or conversely, every possible candidate ubiquitin ligase for AID, was not tested, these experiments establish that (1) RNF126 shows specificity for AID when presented with a substrate known to be present at the immunoglobulin locus during CSR and SHM and (2) that two other ligases, one of which has been demonstrated to be present at the immunoglobulin locus (RNF8) and one which is expressed in switching B cells and shows high homology to RNF126 (BCA2) cannot ubiquitinate AID.

CHAPTER 4: GENERATION AND CHARACTERIZATION OF AN RNF126 CONDITIONAL KNOCKOUT MOUSE MODEL SYSTEM

4.1 Motivation

A number of screening methods have been undertaken to show that AID interacts with a variety of cellular proteins, suggesting the involvement of several universal cellular processes in AID-mediated reactions. However, for many of these factors mechanistic details are still lacking. Often the AID cofactors identified are essential cellular proteins. This makes it difficult to generate knock-out models due to lethality as well as produces complicated phenotypes because of the existence of multiple cellular roles for a given factor. Thus, most findings have relied on shRNA knock-down studies and/or mutation of the empirically-determined binding interface. In addition, a variety of model systems have been utilized for these studies, including DT40 chicken B cells, the murine B cell lymphoma line, CH12s, and primary B cells *in vitro* and *in vivo.* Each of these contexts studies a slightly different AID-dependent mechanism and thus may utilize a different set of AID cofactors.

Having already established that the E3 Ubiquitin Ligase, RNF126, is able to interact with and mono-ubiquitinate AID (Delker et al., 2012), the role of RNF126 and RNF126-mediated ubiquitination in antibody diversification remained to be examined. Because of the difficulties associated with analyzing the importance of putative AID cofactors, the generation of an RNF126 conditional knockout mouse model was undertaken in order to provide the most thorough analysis of the role of RNF126 during antibody diversification. Through

the generation of a B-cell specific knockout, the role of RNF126 during SHM, CSR and affinity maturation could be assessed.

4.2 Generation of an RNF126 conditional knockout mouse model

Because *rnf126* is a relatively uncharacterized gene, it was unclear whether a full knockout would result in embryonic lethality. For this reason, generation of a conditional knockout mouse model, taking advantage of the Cre Recombinase system, proved to be the most prudent approach (Schmidt-Supprian and Rajewsky, 2007). The *rnf126* locus on chromosome 10 was targeted with a knock-out first construct, which contains a gene-trap cassette within intron 1, terminating translation after exon 1 when the full construct is present. Conditional knock-out potential can be restored by taking advantage of the FRT, or Flipase Recombinase sites, that flank the gene-trap cassette. In the absence of the gene-trap and drug resistant cassettes within intron 1, deletion of exons 2-8, and thus generation of a knock-out allele, is made possible by the presence of two LoxP sites which flank exons 2-8, the bulk of the coding region of the gene (Figure 4.1). A more thorough presentation of the strategy used to both generate and genotype the RNF126 conditional knockout mouse is presented in Figure 4.2

Targeted ES cells were derived from a C57BL/6N background, but with an agouti coat color, thus enabling detection of high chimerism by mosaic coat color while preserving the C57BL/6 background even in the F1 progeny. Breedings between several chimeric mice and C57BL/6 wildtype mice produced one offspring in which transmission of the targeted construct was detected both by

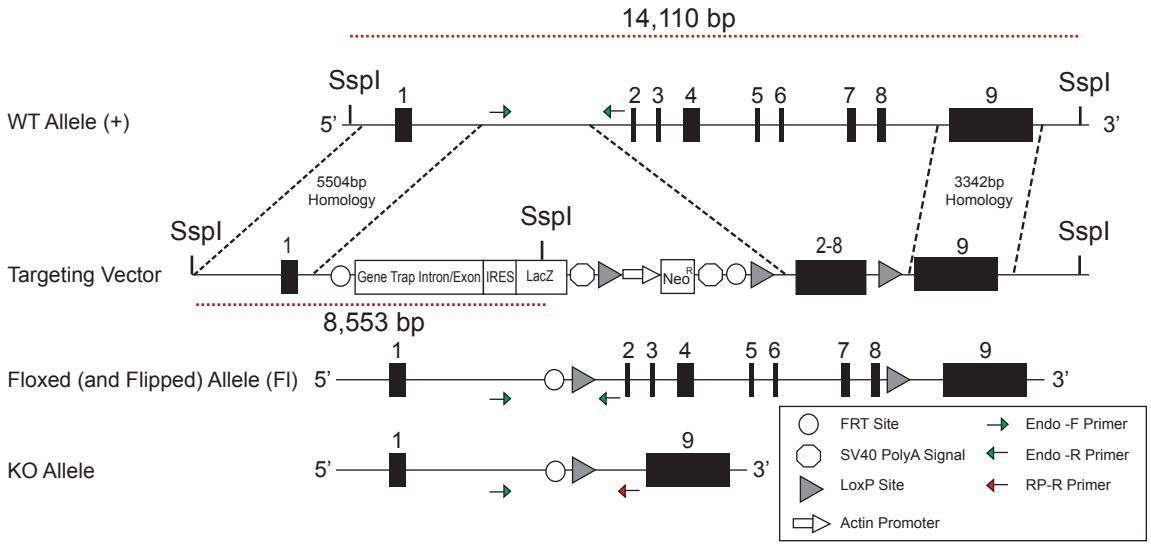
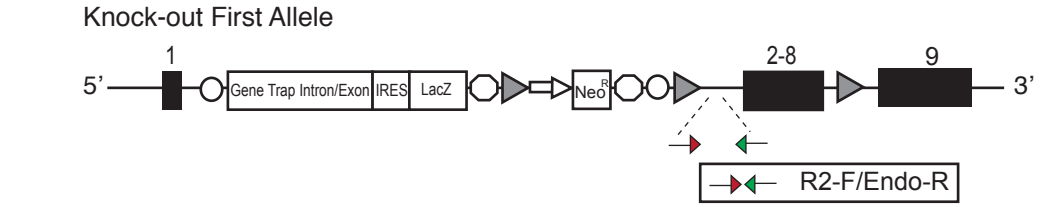
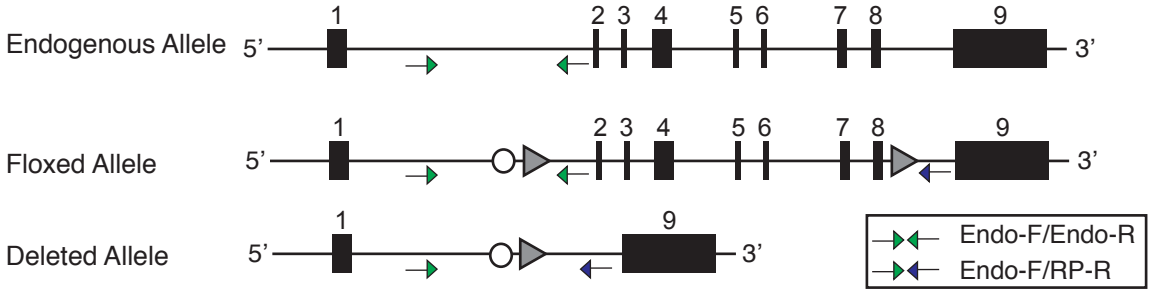


Figure 4.1 Schematic of targeting construct and resultant allele. The endogenous allele, targeting vector, targeted allele and knockout allele are shown. RNF126 contains 9 exons, of which exons 2-8 are flanked by LoxP sites. In addition, a gene-trap construct has been inserted within intron 1, which, when present, prevents translation through the locus thereby creating a knock-out allele. The gene-trap construct is flanked by FRT recombination sites and was deleted by breeding mice transgenically expressing the Flipase recombinase (see Floxed and Flipped allele). B-cell specific conditional knockouts were obtained by breeding mice containing the Floxed (and Flipped) allele with mice expressing the recombinase CRE from the mb-1 locus. SspI sites used for genomic DNA digestion and expected band sizes for the endogenous allele and targeted allele are shown. In addition, primers used for PCR genotyping the locus are shown as green and red arrows.

A. Full Knock-out Genotyping



B. Conditional Knock Out Mouse Genotyping



C. Genotyping PCR Products

Allele	R2-F/Endo-R	Endo-F/Endo-R	Endo-F/RP-R
Endogenous Allele	X	367 bp	X
Knock-Out First Allele	213 bp	X	X
Floxed Allele	213 bp	565 bp	X
Deleted Allele	X	X	329 bp

Figure 4.2 Detailed schematic of RNF126 locus and genotyping primers. (**A**) Genotyping of the targeted knock-out first allele. Prior to breeding to mice expressing Flipase, mice containing the knock-out first allele were genotyped with the R2-F/Endo-R primer set, yielding a 213bp product that is not produced from the endogenous allele. (**B**) Differentiation of the endogenous allele from the Floxed allele is obtained using the primer set Endo-F/Endo-R. Identification of deletion of exons 2-8 by CRE-induced recombination is determined using the primer set Endo-F/RP-R. (**C**) A table denoting genotyping primer sets used and the expected amplicon for each RNF126 allele.

coat color and southern blotting (Figure 4.3A). Deletion of the gene-trap cassette was accomplished by crossing mice containing at least one *rnf126*-targeted allele with Flipase expressing mice. Subsequently, the Flipase transgene was crossed out, leaving the final targeted allele (abbreviated "Fl"), which has knock-out potential when crossed with a Cre-expressing strain (Figure 4.3B). Lastly, attempts to breed mice containing one or two targeted alleles with intact gene trap constructs suggest that deficiency in RNF126 results in either insufficiencies in reproduction and/or embryonic lethality as mice containing targeted alleles were born at much lower frequencies than expected. Further, deceased pups found from breedings were enriched for mice containing homozygous targeted alleles (Figure 4.4).

4.3 Validation of the RNF126 conditional knockout mouse model.

In order to generate a B-cell specific *rnf126*$^{-/-}$ conditional knock-out, the established mb1-Cre strain of mice, which contains a humanized Cre gene knocked into the *mb1* locus, was used (Hobeika et al., 2006). The *mb1* gene encodes the Igα subunit of the B cell receptor and is expressed at the pro-B cell stage of B cell development. Thus, Cre expression is upregulated, and the floxed gene of interest deleted, in the earliest stage of B cell development.

To validate that the construct functions as it should, experiments were conducted at the DNA, RNA and protein level to confirm that deletion of *rnf126* was occurring in the B cells of mice with the genotype RNF126$^{Fl/Fl}$mb1$^{Cre/+}$. Importantly, PCR genotyping of genomic DNA derived from the tail reveals an intact locus, whereas genotyping of genomic DNA from splenic B cells from the

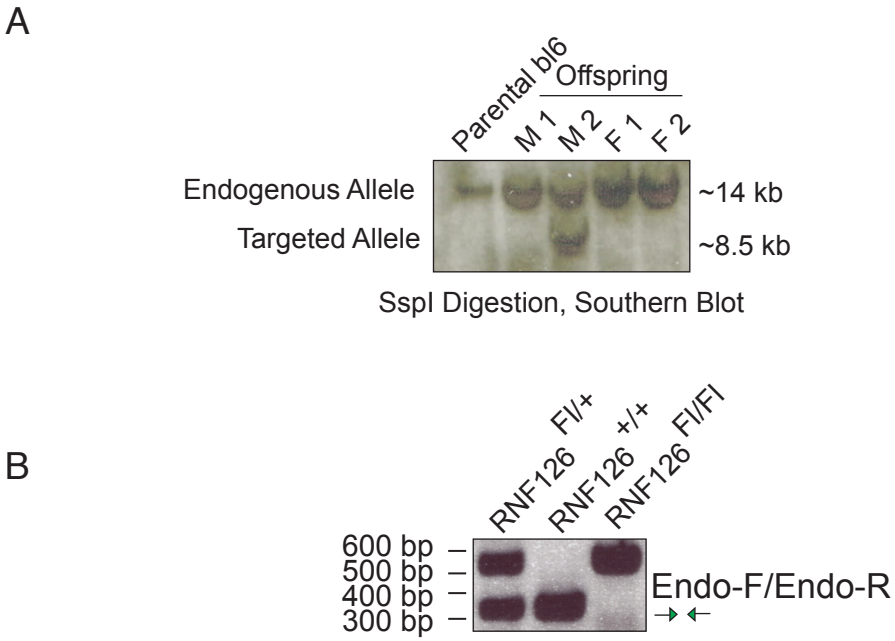


Figure 4.3 Genotyping RNF126 conditional knockout mice. (**A**) Validation of construct integration in founder mouse by Southern Blot. Genomic DNA from the offspring of an RNF126 chimeric male and c57bl/6 female was prepared. DNA from the parental female was also collected as a control. DNA was digested with the restriction enzyme, SspI, southern blotting performed. As expected, the bl6 female displayed one band of approximately 14,000 kb, representative of the endogenous allele. Only one offspring, Male 2, displayed banding expected from the digestion of one endogenous allele and one targeted allele (see schematic in Figure 4.1). (**B**) After the gene trap construct was removed from intron 1, mice were genotyped using primers Endo-F/R described in Figure 4.1. The endogenous allele (+) and the Floxed (and Flipped allele, Fl) are distinguished by a size difference due to the presence of an FRT and LoxP site present in the amplicon of the Fl allele. Genotyping products from RNF126(Fl/+), RNF126(+/+) and RNF126(Fl/Fl) mice are shown.

A

RNF126$^{-/-}$ X RNF126$^{+/-}$

Total Pups: 18	RNF126$^{-/-}$	RNF126$^{+/-}$
Expected	9	9
Actual	2	16

RNF126$^{+/-}$ X RNF126$^{+/-}$

Total Pups: 13	RNF126$^{-/-}$	RNF126$^{+/-}$	RNF126$^{+/+}$
Expected	3.25	6.5	3.25
Actual	0	9	4

B

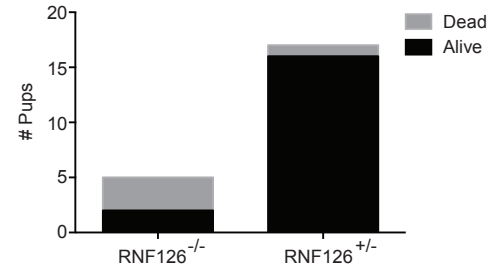


Figure 4.4 Full RNF126 knockout mice are produced at much lower rates than expected. Because the targeting vector contains a gene trap construct in intron 1, disrupting expression through the allele, mice that are homozygous for this allele are “functional” knock-outs (RNF126-/-). Very few functional knockout mice were produced from breedings. (**A**) Breedings between homozygous and heterozygous functional knockouts produced significantly fewer homozygous mice than expected. Tables depict total number of pups from each breeding type, and the expected and actual number of pups for each genotype. (**B**) When dead pups were found in cages they were genotyped. The majority of dead pups found were homozygous for the targeted allele and thus RNF126-/-.

same mouse reveals the expected deletion (Figure 4.5A). Thus, despite the presence of the *Cre* gene throughout the mouse, expression is limited to the B-cell subset. In some cases it appeared as if Cre expression was leaky and deletion occurred in a subset of cells derived from tail. When mice that contained leaky Cre expression were used in experiments, it was noted; however, no differences were observed between B-cells derived from these mice as compared to non-leaky Cre expressing mice.

In addition, it has been verified that deletion of exons 2-8 of the *rnf126* gene in B cells results in loss of gene expression, which can be shown both at the level of mRNA (Figure 4.5B) and protein (Figure 4.5C).

4.4 Analysis of RNF126 conditional knockout mice

4.4.1 B Cell Development is unaffected by the loss of RNF126

Because the recombinase Cre was expressed from the mb-1 locus, thus deleting the *rnf126* gene at the Pro-B cell stage (Hobeika et al., 2006), it remained a possibility that B cell development would be affected by the loss of RNF126. To assess this, flow cytometry was used to characterize the relative levels of Pro-, Pre- and Immature B cells in the bone marrow of $RNF126^{Fl/Fl}$ $mb1^{+/+}$ and $RNF126^{Fl/Fl}$ $mb1^{Cre/+}$ mice. $RNF126^{Fl/Fl}$ $mb1^{Cre/+}$ mice, when compared to both unrelated C57Bl/6 wildtype mice and $RNF126^{Fl/Fl}$ $mb1^{+/+}$ mice, had comparable levels of each B-cell subset and thus displayed no observable defect in B cell development (Figure 4.6 A), In addition, the proportion of mature IgM/IgD double positive B cells in the spleen of these mice was assayed and, again, no defect was observed in the $RNF126^{Fl/Fl}$ $mb1^{Cre/+}$ mice as compared to

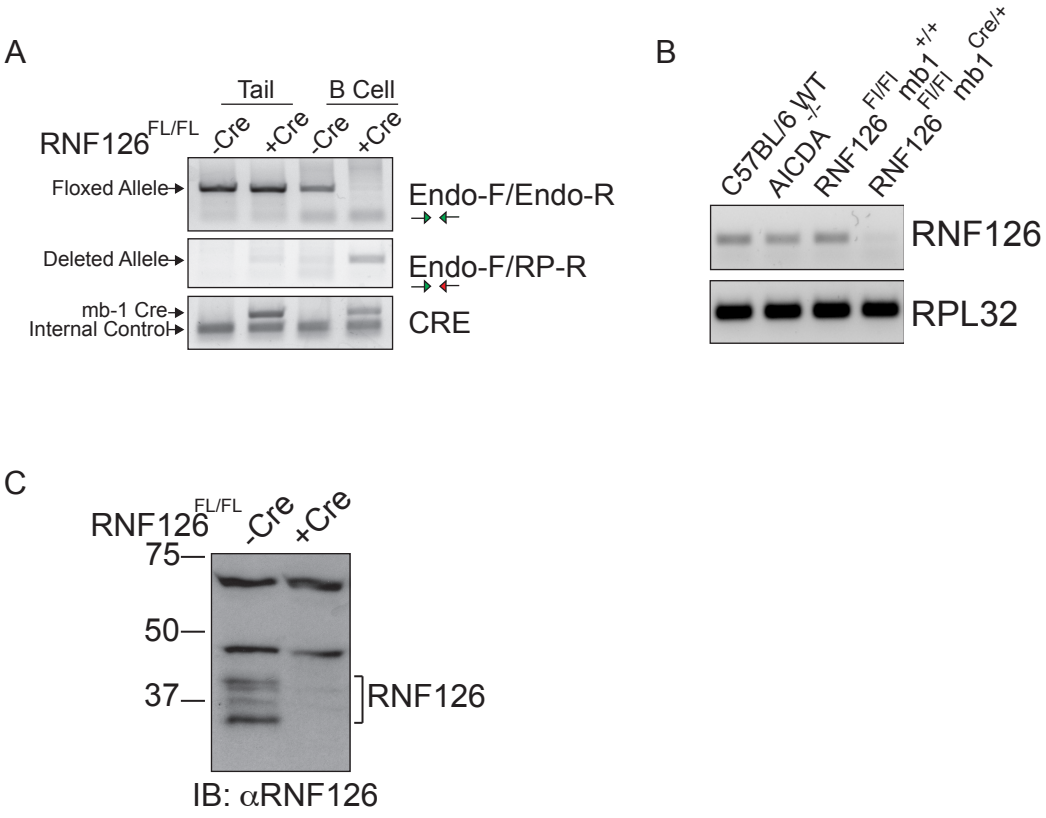


Figure 4.5 The use of mb-1 Cre efficiently deletes RNF126 in B cells. (**A**) PCR Genotyping using primers depicted in Figure 4.1 show deletion of RNF126 exons 2-8 in DNA prepared from splenic B cells, but not tail. Deletion is assayed with two primer sets: Top panel, shows the disappearance of a band due to the loss of the reverse primer binding site; middle panel, shows the appearance of a band due to the recombination of two distant LoxP sites surrounding exons 2 and 8. This deletion occurs only in the presence of the mb-1 CRE gene (bottom panel). (**B**) RNA was collected from splenic B cells stimulated for 72hr in vitro. B cells were derived from an unrelated c57bl/6 wildtype mouse, an unrelated *aicda-/-* mouse and related rnf126Fl/Fl mb1+/+ and rnf126Fl/Fl mb1Cre/+ mice. Primers that bind in Exon 1 and Exon 4 of the rnf126 gene were used to assay the presence of transcript. Primers within the rpl32 gene were used as a normalization control. (**C**)Western blot analysis of whole cell lysates derived from splenic B cells stimulated in culture to undergo CSR shows the absence of RNF126 protein in B cells in RNF126Fl/Fl mb-1CRE/+ mice, but not RNF126Fl/FL mb-1+/+ mice. RNF126 runs as multiple bands, as marked. An anti-RNF126 western blot is shown.

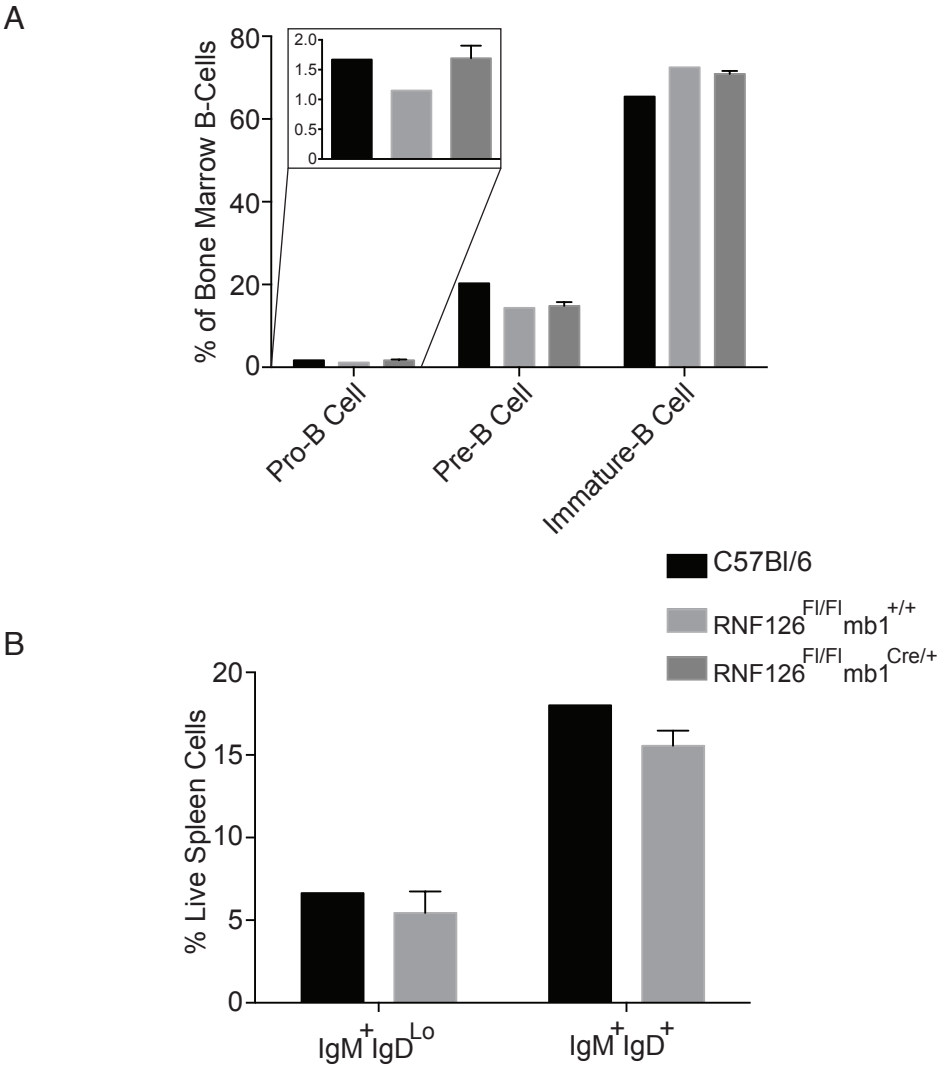


Figure 4.6 B cell development is not affected by loss of RNF126.
(**A**) Summary of FACS analysis of bone marrow cells derived from C57Bl/6 wildtype, RNF126Fl/Fl mb1+/+ and RNF126Fl/Fl mb1Cre/+ demonstrates comparable levels of pro-, pre- and immature B cells. Data from 1 C57Bl/6 widltype mouse, 1 RNF126Fl/Fl mb1+/+ mouse and 2 RNF126Fl/Fl mb1Cre/+ is shown. Error bars denote standard deviation. (**B**) Summary of FACS analysis of spleen cells derived from the same mice as in A. Again, there are comparable levels of IgM+IgD(Lo) and IgM+IgD+ B cells in all genotypes tested.

wildtype mice (Figure 4.6 B). Thus, normal B cell development is maintained in the absence of RNF126.

4.4.2 Genetic deletion of RNF126 results in a slight reduction of CSR *in vitro*

As AID is the key component of antibody diversification and loss of AID results in complete abrogation of antibody diversification in mature B cells, it was of great interest to determine if deletion of RNF126 also affected CSR. Thus, experiments were conducted in order to determine if primary B cells derived from $RNF126^{Fl/Fl}$ $mb1^{Cre/+}$ mice displayed compromised levels of CSR. Splenic B cells derived from $RNF126^{Fl/Fl}$ $mb1^{Cre/+}$ and $RNF126^{Fl/Fl}$ $mb1^{+/+}$ mice were stimulated *in vitro* with IL-4/αCD-40 to induce a switch from the IgM isotype to that of IgG1. Levels of CSR were measured by flow-cytometry and reported as the percentage of B cells, which express surface-bound IgG1. While RNF126-deficient B cells are able to undergo CSR, they do so at a slightly reduced efficiency, both when CSR levels are measured at 72hr and 96hr post-stimulation (Figure 4.7). In response to these stimuli, RNF126-deficient B cells displayed a slight delay in proliferation, which may account for the subtle defects observed in CSR, as it is known that CSR is dependent on progression through the cell cycle (Figure 4.8).

4.4.3 Loss of RNF126 results in a subtle defect in Affinity Maturation in response to NP-CGG

In addition to affecting CSR, mis-regulation of AID can also result in aberrant SHM and affinity maturation. In fact, it is entirely possible that there exists a different subset of AID cofactors for each of these processes, thus

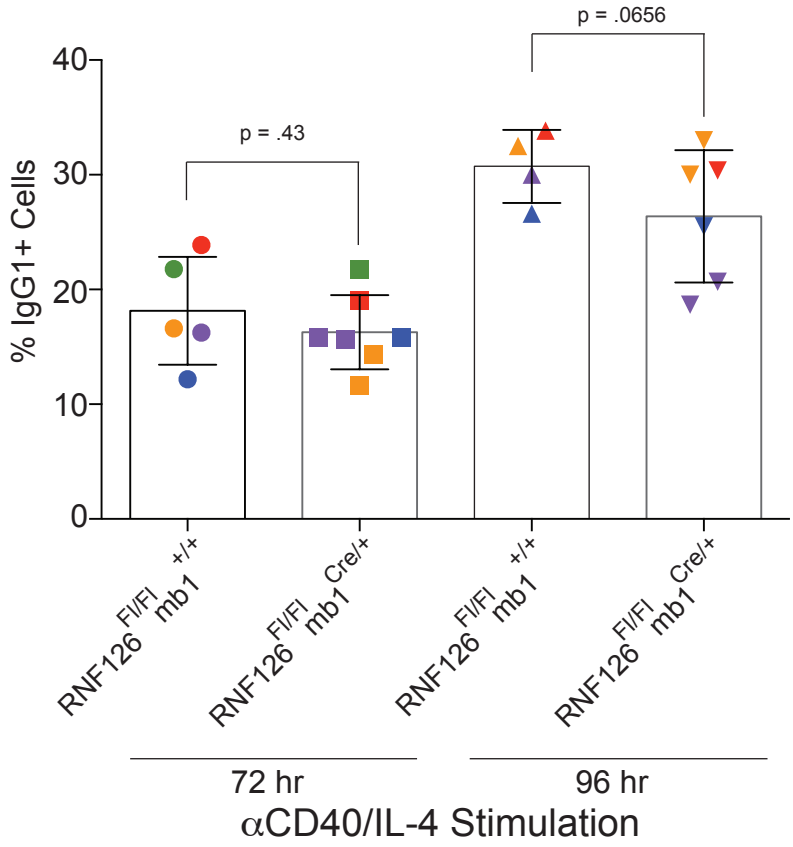


Figure 4.7 Genetic loss of RNF126 does not significantly impair CSR. Naïve splenic B cells derived from RNF126Fl/Fl mb1+/+ and RNF126Fl/Fl mb1Cre/+ mice were stimulated in vitro with aCD-40 and IL-4 to induce a CSR event to IgG1. FACS analysis completed at 72hr and 96hr post stimulation was used to determine the percentage of cells which express surface IgG1. Colors indicate experimentally paired mice. P-values, determined by the unpaired student's t-test, are shown.

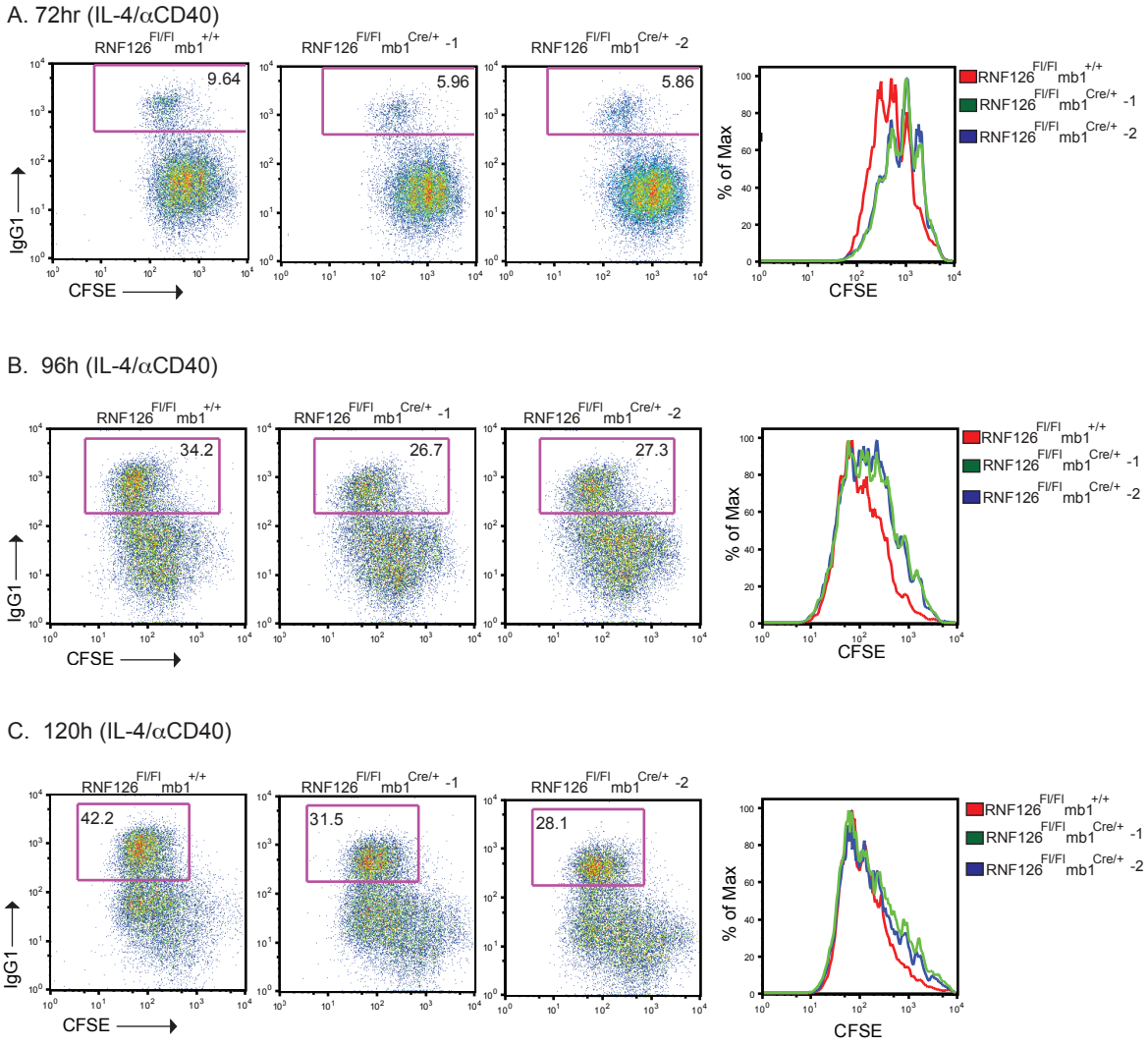


Figure 4.8 RNF126 knockout B cells exhibit a slight delay in cell proliferation. RNF126Fl/Fl B cells were purified from 1 RNF126Fl/Fl mb1+/+ mouse and RNF126Fl/Fl mb1cre/+ B cells were purified from 2 RNF126FL/Fl mb1Cre/+ mice (labeled 1 and 2). Cells were loaded with CFSE as described in Materials and Methods and stimulated in vitro with IL-4 and anti-CD40. FACS plots show switch rates to IgG1 relative to cell division (CFSE) and histograms overlay CFSE profiles of each genotype. Analysis was completed at (A) 72hr, (B) 96hr and (C) 120hr post-stimulation.

contributing to the difference in outcomes. At present, and for unknown reasons, SHM in B cells cannot be assayed *in vitro*. Thus, to determine if loss of RNF126 results in a defect in SHM and Affinity Maturation, an *in vivo* immunization model was used. The established and well-studied nitro-phenol (NP) immunization model was chosen.

To assay affinity maturation, or the increase in antibody affinity to a given antigen as a result of somatic hypermutation and B cell-selection in the germinal center, animals were immunized with NP_{16}-CGG and serum was collected to assess titers of NP-specific IgG1 antibodies. Affinity to NP conjugated to BSA in a ratio of 30 (NP_{30}-BSA) was used as a measure of total NP-specific IgG1 antibodies and in a ratio of 3 (NP_3-BSA) as a measure of high affinity NP-specific antibodies. Affinity maturation was then calculated as the ratio of high affinity to total NP-specific antibodies (NP_3/NP_{30}). As expected, there was an increase in titers of total and high affinity NP-specific antibodies as time progressed post-immunization; this increase occurred in both the $RNF126^{Fl/Fl}$ $mb1^{+/+}$ and $RNF126^{Fl/Fl}$ $mb1^{Cre/+}$ mice, however a significant defect was observed in the production of high affinity anti-NP antibodies in $RNF126^{Fl/Fl}$ $mb1^{Cre/+}$ mice (Figure 4.9, (representative of Cohort 2--refer to Figure 4.11)). In total, 28 mice have been immunized and peripheral blood collected at day 21 post-immunization. NP3/NP30 ratios have been measured in these mice and show that loss of RNF126 results in a subtle, but consistent, defect in affinity maturation (Figure 4.10).

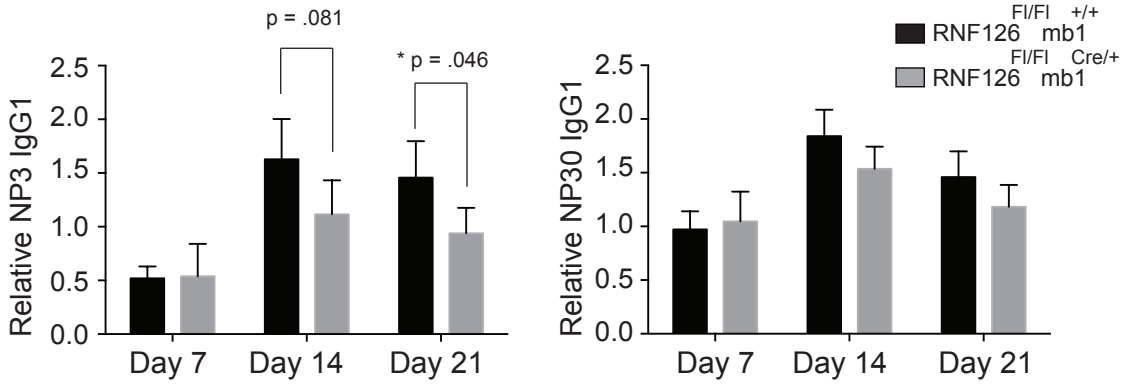


Figure 4.9 Loss of RNF126 hinders production of high affinity antibodies. Titers of NP3-binding IgG1 and NP30-binding IgG1 were determined by ELISA for cohort 2, which contained 4 RNF126FL/FL mb1+/+ and 4 RNF126Fl/Fl mb1Cre/+ mice at D7, D14 and D21 post-immunization. NP-specific IgG1 levels were normalized to total serum IgG1. Error bars denote standard deviation. P-values, determined by the unpaired student's t-test, are shown.

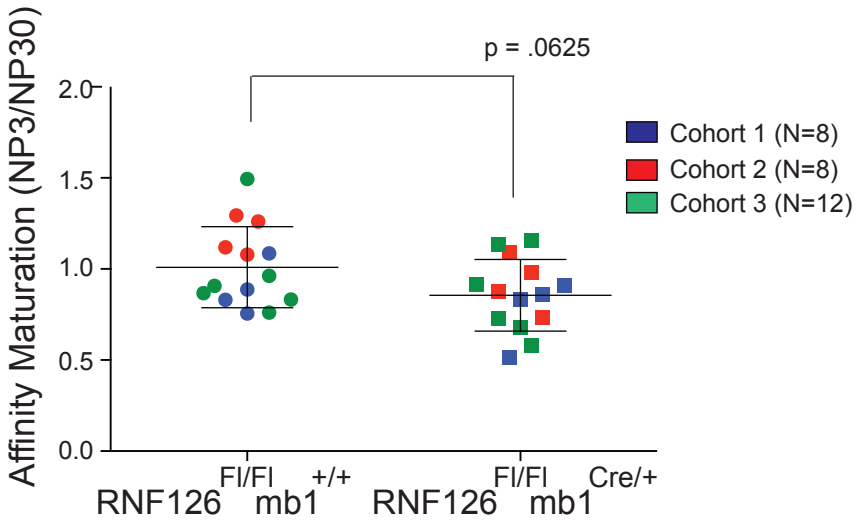


Figure 4.10 RNF126-deficient B cells display a defect in affinity maturation. NP3/NP30 ratios at D21 post-immunization, or a measure of affinity maturation, are shown for each mouse within each genotype (X-axis) and cohort (color coordinated). A decrease in affinity maturation occurs in mice with RNF126-deficient B cells, with a p-value of .0625 (unpaired student's t-test). The bars denote the average NP3/NP30 ratio and the standard deviation.

In addition, Figure 4.11 presents the relative titers of (1) high affinity NP-specific antibodies and (2) total NP-specific antibodies measured at D21 post-immunization in the three cohorts of mice immunized. As can be seen, the overall immune response in Cohort 3 does not appear to be as robust as Cohort 1 and 2. For completeness, though, this data is still included in the affinity maturation analysis. It is possible, though, that the inclusion of Cohort 3 has skewed the significance of difference in affinity maturation between wildtype and RNF126 conditional knock-out mice and should be taken into consideration. Affinity maturation measurements shown in Figure 4.10 are color-coded based on cohort number and, as can be seen, a much greater defect is observed in the first two cohorts.

Thus, while RNF126 does not appear to be essential for either CSR and affinity maturation, genetic loss of RNF126 does result in subtle defects in both processes. This suggests that RNF126 plays a role in fine-tuning the reactions.

4.4.4 Loss of RNF126 results in altered mutation patterns during Somatic Hypermutation

Another benefit to using the NP-immunization model is that NP induces a characteristic immune response, allowing one to sequence the variable region of the immunoglobulin gene and assess mutation rates and patterns. Due to high germline affinity for NP, B cells which have produced a V(D)J rearrangement containing the V186.2 V region and the JH2 J region are selected for post-immunization. Thus, primers specific for these two regions can be used to sequence for mutations within the variable region.

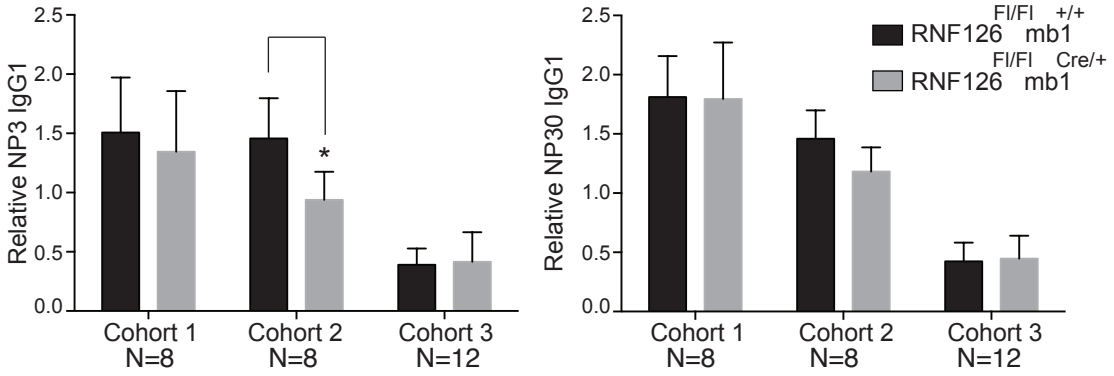


Figure 4.11 Complete presentation of antibody titers in three cohorts of mice. A total of 28 mice, within 3 cohorts, were immunized with NP-CGG as described in Materials and Methods. Cohort 1 and 2 each contained 4 RNF126FL/FL mb1+/+ and 4 RNF126Fl/Fl mb1Cre/+, and Cohort 3 contained 6 RNF126FL/FL mb1+/+ and 6 RNF126Fl/Fl mb1Cre/+. Average titers of NP3-binding IgG1 (left panel, high affinity) and of NP30-binding IgG1 (right panel, total) at day 21 post-immunization in control (black) and RNF126-deleted mice (gray) for cohorts 1-3 are shown. The asterisk denotes that the decrease in relative NP3-binding IgG1 levels in the RNF126-deleted mice of cohort 2 compared to wildtype is statistically significant ($p=.046$, determined by unpaired student's t-test). Error bars denote the standard deviation.

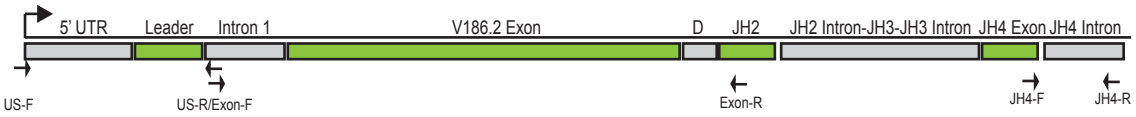
To investigate whether RNF126 is necessary for somatic hypermutation *in vivo*, two RNF126$^{Fl/Fl}$ mb1$^{+/+}$ and three RNF126$^{Fl/Fl}$ mb1$^{Cre/+}$ mice were immunized intraperitineally with nitro-phenol conjugated to chicken gamma globulin (NP_{16}-CGG). Mice were sacrificed 14 days post-immunization, germinal center B cells FACs-sorted using the cell-surface markers FAS, GL7 and CD19 and genomic DNA prepared. Somatic mutations were assessed in three regions of the immunoglobulin heavy chain gene: (1) the rearranged V186.2 variable region, the dominant form of antibody selected for in response to NP, (2) the un-rearranged JH4 intronic region downstream of the rearranged variable region, and (3) the region upstream of the V186.2 variable exon and downstream of the promoter region (schematic presented in Figure 4.12). Mutations found within each of these regions can be informative. First, because the V186.2 exonic region encodes for the majority of the antigen-binding domain of the antibody, mutations found here are under selection pressure to improve affinity. In contrast, mutations located downstream of the rearranged variable region in the JH4 intron, which are not informative of the antibody produced, represent those mutations that are not under selection pressure. Thus, analysis of these mutations can be more informative of AID activity and repair pathways utilized because selection has not occurred to erase the mark of those mutations. Lastly, the third region, termed the V186.2 Upstream Region, was sequenced to assess the existence of any alterations in the spatial location of mutations relative to the promoter, where AID is thought to be loaded along with RNAP II (Peters and Storb, 1996) and induce mutation upon RNAP II stalling (Pavri et al., 2010). Differences in the pattern of

mutations found in this region could be informative of the presence of defects in AID complex formation at the Ig locus. Analysis of SHM in RNF126$^{Fl/Fl}$ mb1$^{+/+}$ and RNF126$^{Fl/Fl}$ mb1$^{Cre/+}$ mice is broken down into the regions described. A summary of the data obtained from the three regions is shown in Figure 4.12.

A. V186.2 Exon

Because this region codes for the portions of the antibody that come into direct contact with antigen, mutations here have the most impact on affinity. Thus, an analysis of mutations in this region should correlate with an analysis of affinity maturation. A defect in the generation of high affinity antibodies, and thus affinity maturation, was detected in RNF126$^{Fl/Fl}$ mb1$^{Cre/+}$ mice (Figure 4.9 and 4.10). Intuitively, then, it would be expected that the loss in affinity maturation is due to a decrease in mutation within this region. Unexpectedly, this was not what was found. As a note, for all analyses presented unique mutation frequencies were calculated. This is a more conservative approach because it not only excludes PCR duplicates (where the entire sequence is the same), but also excludes duplicate mutations at the same base in the same mouse. These duplicates could represent *bona fide* AID-mutations, but more likely represent a mutation that was carried through to an additional round of mutagenesis. Thus, this one mutation appears in multiple sequences derived from clonally divided B cells. Unique mutation frequencies in the V186.2 exonic region were not significantly different in RNF126$^{Fl/Fl}$ mb1$^{Cre/+}$ mice as compared to RNF126$^{Fl/Fl}$ mb1$^{+/+}$ mice (Figure 4.12). Thus, loss of RNF126 does not inhibit the introduction of mutations by AID.

A



B

	V186.2 Exon		JH4 Intronic Region		V186.2 Upstream Region	
	$RNF126^{Fl/Fl}mb1^{+/+}$ N=2	$RNF126^{Fl/Fl}mb1^{Cre/+}$ N=3	$RNF126^{Fl/Fl}mb1^{+/+}$ N=2	$RNF126^{Fl/Fl}mb1^{Cre/+}$ N=3	$RNF126^{Fl/Fl}mb1^{+/+}$ N=2	$RNF126^{Fl/Fl}mb1^{Cre/+}$ N=3
Sequences Analyzed	64	96	99	190	116	163
Mutated Sequences (%)	64 (100%)	95 (99%)	53 (54%)	82 (43%)	27 (23%)	44 (27%)
Unique Mutations (Total)	145 (313)	220 (585)	77 (122)	113 (150)	40 (72)	64 (131)
Mutation Frequency (Unique, 10^{-2}) (p-value)	6.9	7.0 (.96)	3.6	3.5 (.9)	2.3	2.4 (.8)
Mutation Frequency (Non-Unique, 10^{-2}) (p-value)	1.4	1.7 (.002)	.35	.22 (.0002)	.21	.27 (.09)

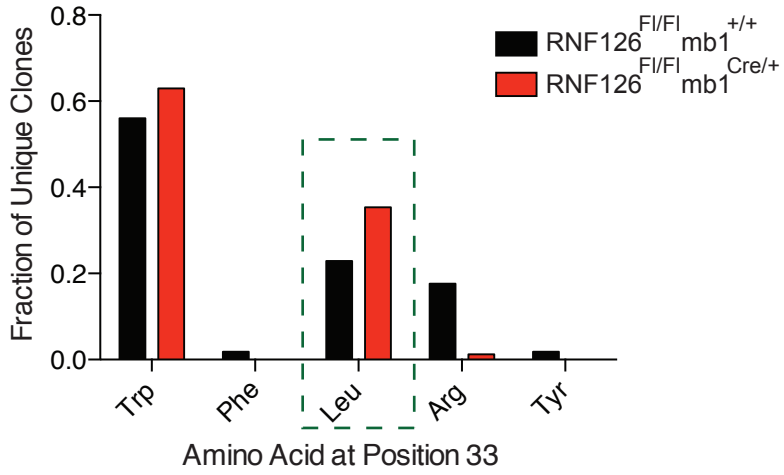
Figure 4.12 Overview of somatic hypermutation analysis in immunized mice. (**A**) Schematic of regions used for mutation analysis. Regions were cloned from genomic DNA prepared from germinal center B cells from 14-day immunized mice (rnf126Fl/Fl mb1+/+ and rnf126Fl/Fl mb1Cre/+). Primers used to amplify each region are shown: US-F/R were used to amplify the 295bp V186.2 upstream region, which spans from the promoter to Intron 1 of the recombined antibody gene, Exon-F/R were used to amplify the 351bp V186.2 exonic region, which makes up the bulk of the antigen binding domain of the translated antibody and JH4-F/R were used to amplify the 357bp JH4 intronic region, beginning at the exon/intron junction. (**B**) Detailed representation of mutation analysis. A table is used to describe the following information for each genotype and region sequenced: (1) the total number of sequences analyzed, (2) the number of mutated sequences as an absolute number and a percent of total sequences, (3) the number of unique and total mutations, and (4) the unique mutation frequency and (5) the non-unique mutation frequency, each with its corresponding p-value, as calculated by the chi-squared test. The number of mice of each genotype used is stated (N=2 for rnf126Fl/Fl mb1+/+ and N=3 for rnf126Fl/Flmb1Cre/+).

This data, though, leaves open the question of why affinity maturation is defective in RNF126 conditional knockout mice. This inconsistency can be explained in a few different ways. First, it is known that particular mutations (Tryptophan 33 to Leucine (W33L)) contribute greatly to NP-affinity. Thus, a decrease in this mutation could explain defective affinity maturation. In addition, it has been observed before that the introduction of additional mutations in the region either by overactive AID or AID that persists post-germinal center can also be detrimental to affinity (Teng et al., 2008). Analysis of V186.2 sequences from $RNF126^{Fl/Fl}$ $mb1^{Cre/+}$ and $RNF126^{Fl/Fl}$ $mb1^{+/+}$ mice revealed that, while there were no differences in the prevalence of the W33L mutation, there was a shift toward a larger number of mutations present per sequence/B-cell in the RNF126 conditional knockout mice (Figure 4.13). In fact, an analysis of the non-unique mutation frequency reveals that RNF126 conditional knockout mice exhibit a greater number of mutations than their wildtype counterparts (Figure 4.12). These additional mutations could account for the loss of affinity for immunogen. Further experiments will be necessary to determine the mechanism of how RNF126 limits AID mutation in the context of the V186.2 exon.

B. JH4 Intron

As mentioned, the JH4 intron lies downstream of the rearranged variable region. Thus, it does not encode for antibody and mutations here can be used to assess the mutagenic load in the absence of germinal-center selection. Just as was seen in the V186.2 exonic region, unique mutation frequencies in the JH4 intron from $RNF126^{Fl/Fl}$ $mb1^{Cre/+}$ and $RNF126^{Fl/Fl}$ $mb1^{+/+}$ mice were not

A



B

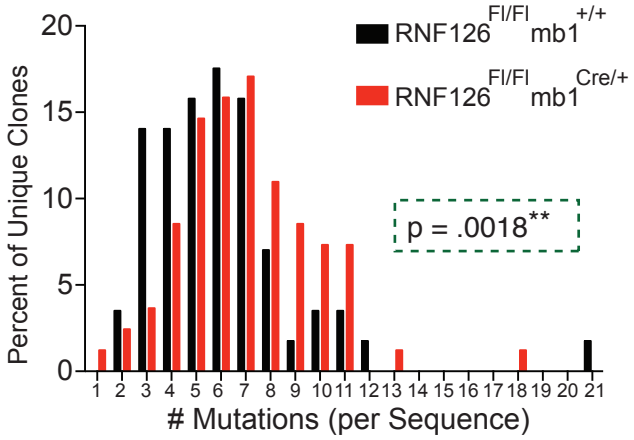


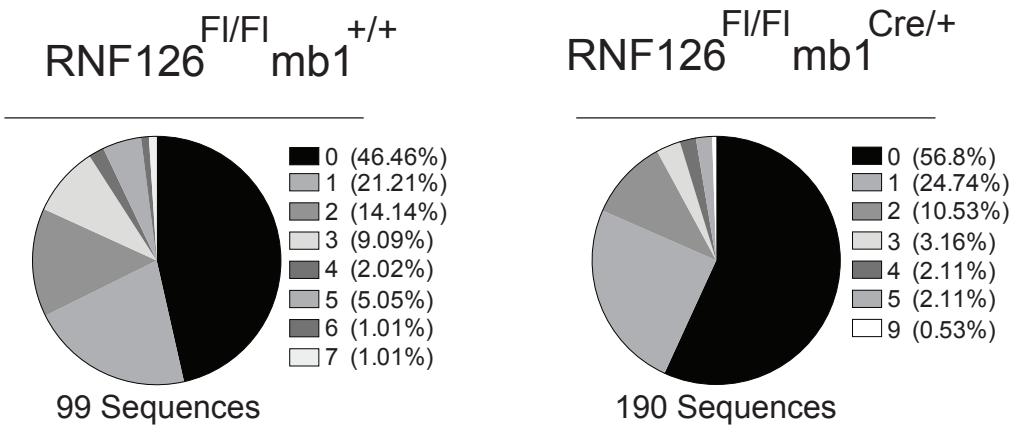
Figure 4.13 Loss of RNF126 results in altered mutation patterns in the V186.2 exon. (**A**) An analysis of the types of mutations at residue 33 (W33) reveals that there is no significant decrease in the known high-affinity mutation W33 to L33 in RNF126 conditional knockout mice. The fraction of sequences mutated to other amino acids is also shown. (**B**) An analysis of the number of mutations per clone reveals that sequences derived from RNF126-deficient B cells contain more mutations than those derived from wildtype B cells. This difference is significant (p = .0018), as determined by the Mann-Whitney test. In both analyses, unique sequences were used so as to avoid PCR duplicates. Wildtype data is presented in black and RNF126 conditional knockout data presented in red.

significantly different (Figure 4.12). Again, it is not likely that RNF126 affects the mutagenic activity of AID. In line with this, there does not appear to be a significant difference in the number of mutations per sequence in the JH4 region (Figure 4.14).

Despite the lack of difference in total mutation frequency, it is particularly evident in this region that the pattern of mutation in RNF126$^{Fl/Fl}$ mb1$^{Cre/+}$ mice is different from that of RNF126$^{Fl/Fl}$ mb1$^{+/+}$ mice. It has been well-established that mutations at C:G basepairs are not strand biased—that is, cytosine bases on the template (T) strand are just as likely as cytosine bases on the exposed non-template (NT) strand to be targeted by AID and mutated. In contrast, mutations at A:T basepairs do show a strand bias. Adenine bases on the NT strand are mutated more frequently than the T strand. This has been explained by the greater accessibility of the error-prone repair pathways, including Pol η, to the NT strand.

As expected, analysis of JH4 sequences from RNF126$^{Fl/Fl}$ mb1$^{+/+}$ mice exhibited strand bias in A:T mutations, but not C:G mutations (Figure 4.15). Strikingly, however, an analysis of strand bias in JH4 sequences derived from RNF126 conditional knockout mice revealed a gain in bias toward the T strand in mutations at C:G bases and a loss of bias at A:T mutations (Figure 4.15). These results suggest that, in the absence of RNF126, there is enhanced access for either AID or error-prone repair proteins to the T strand. *In vitro* transcription and deamination studies would be useful to determine if the presence of RNF126 inhibits AID access to the T strand, or if these results are better explained by the

A



B

RNF126 Fl/Fl mb1 +/+ (%)

Fr/To	G	C	A	T	Sum
G	0	5.9	10.5	4.7	21.1
C	12.8	0	4.3	8.6	25.7
A	18.0	6.0	0	13.2	37.3
T	5.0	7.0	4.0	0	16.0

(G:C = 0.8: 1 ; A:T 2.3 : 1)

RNF126 Fl/Fl mb1 Cre/+ (%)

Fr/To	G	C	A	T	Sum
G	0	8.5	13.7	6.8	29.0
C	3.1	0	3.1	7.8	14.1
A	16.7	7.0	0	9.7	33.5
T	6.6	9.5	7.3	0	23.4

(G:C = 2 : 1 ; A:T 1.4 : 1)

Figure 4.14 Summary of JH4 mutations. **A**. Pie charts show the fraction of total sequences that contain between 0 and 9 mutations for rnf126Fl/Fl mb1+/+ and rnf126Fl/Fl mb1Cre/+ mice. **B**. A table depicts the percentage of mutations from each base in the sequence (vertical) to each possible base (horizontal). Values are corrected for base composition of the sequence (27% G, 15% C, 26% A, 32% T). Wildtype data is presented on the left and RNF126 conditional knockout data is presented on the right. Highlighted squares and arrows depict a decrease (red) and increase (green) in mutation rates in the RNF126 conditional knockout, respectively.

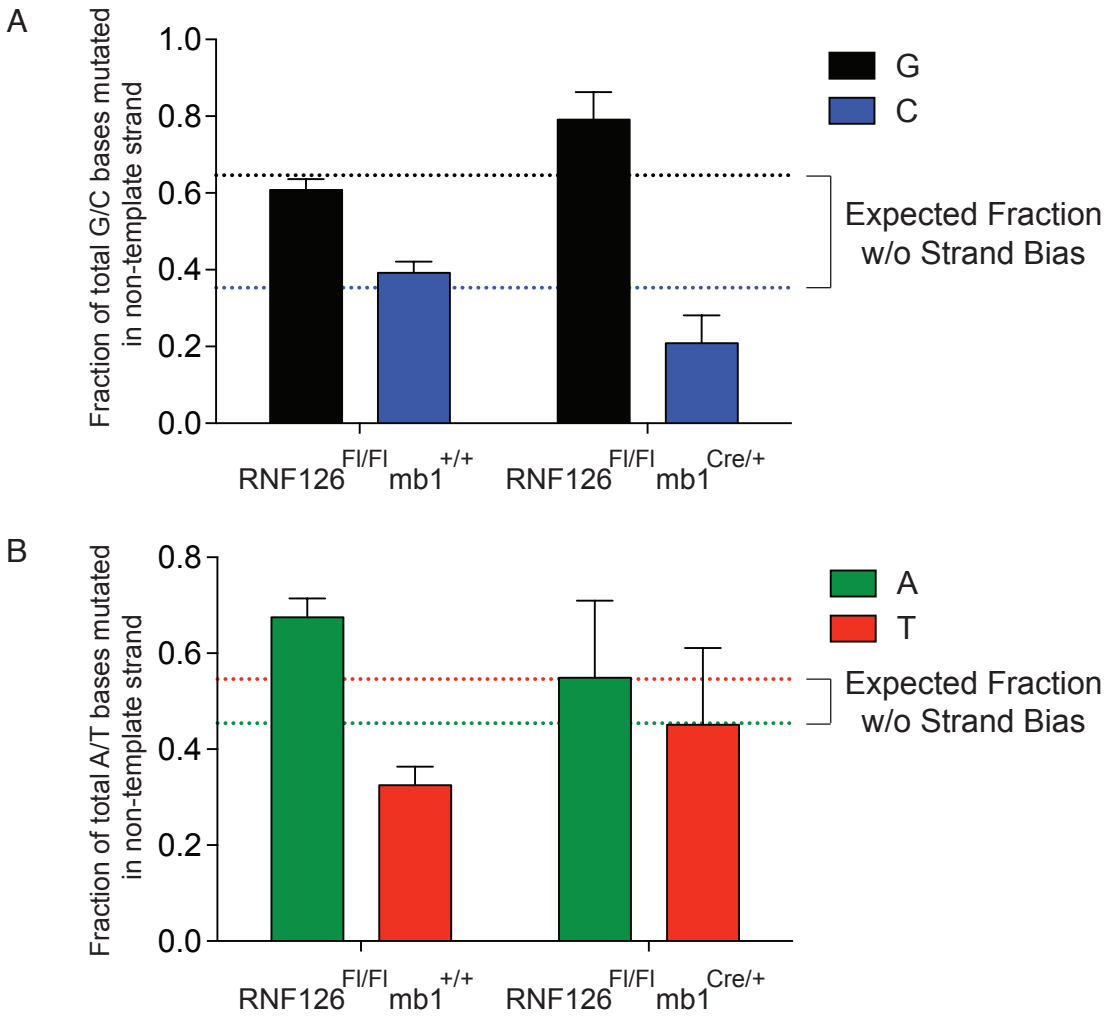


Figure 4.15 Loss of RNF126 results in an increase in template strand mutations in the JH4 intronic region. (**A**) The fraction of total G/C mutations that occur at NT-strand G bases (or C bases on T strand, black) and NT-strand C bases (blue) is presented for wildtype and RNF126 conditional knockout mice. Dotted lines represent expected number of mutations at G and C respsectively, assuming that there is no strand bias. As previously shown, wildtype mice exhibit no strand bias in G/C mutations (p value = .45, chi-squared test); however, RNF126 condtional knockout mice display a bias toward mutation at the template strand (G>C, p value = .0026, chi-squared test). (**B**) A similar analysis is presented for mutations at A:T basepairs. It has been observed that there is a strand bias toward greater mutations at A bases on the NT strand as compared to template strand. Accordingly, wildtype mice in this experiment exhibit the expected strand bias, with more mutations occurring at A bases rather than T bases on the NT strand (p value = < .0001, chi-squared test). A/T mutations in RNF126 conditional knockout mice show a loss of this strand bias, with an increase in A mutations on the T strand (p value = .0538, chi-squared test). Chi-squared tests are performed on actual mutation load versus expected mutation load assuming no strand bias.

differential recruitment of error-prone repair at the T strand. It should be noted that this alteration in strand-bias is not seen in the other two regions analyzed; however, this could be explained by the fact that this pattern is lost from the V186.2 exon due to multiple rounds of selection and cannot be detected in the V186.2 US region because of the scarcity of mutations.

C. V186.2 Upstream (US) Region

From 116 US sequences from RNF126$^{Fl/Fl}$ mb1$^{+/+}$ and 163 US sequences from RNF126$^{Fl/Fl}$ mb1$^{Cre/+}$ mice, there appeared to be no clear difference in either the total mutation frequency or the spatial distribution of mutations throughout the region (Figure 4.12 and 4.16). Thus, it does not appear that loss of RNF126 affects the loading of AID on the transcription complex. In addition, the majority of sequences from mice of both genotypes contained no mutations, reinforcing the observation that AID-mediated mutation occurs approximately 100 nucleotides downstream of the promoter (Figure 4.17).

Further, there appears to be no difference in the spatial distribution of mutations between RNF126$^{Fl/Fl}$ mb1$^{+/+}$ and RNF126$^{Fl/Fl}$ mb1$^{Cre/+}$ as assessed within the V186.2 exon and JH4 intronic region (Figure 4.16), again illustrating that RNF126 does not contribute to the spatial localization of AID-mediated mutation during SHM.

4.5 Knockdown of RNF126 results in a significant decrease in CSR

Because the process of generating a knockout mouse model is arduous, gene knock-down studies, which hijack the cellular RNAi machinery to inhibit translation of a target gene, are often the first step in studying the role of a

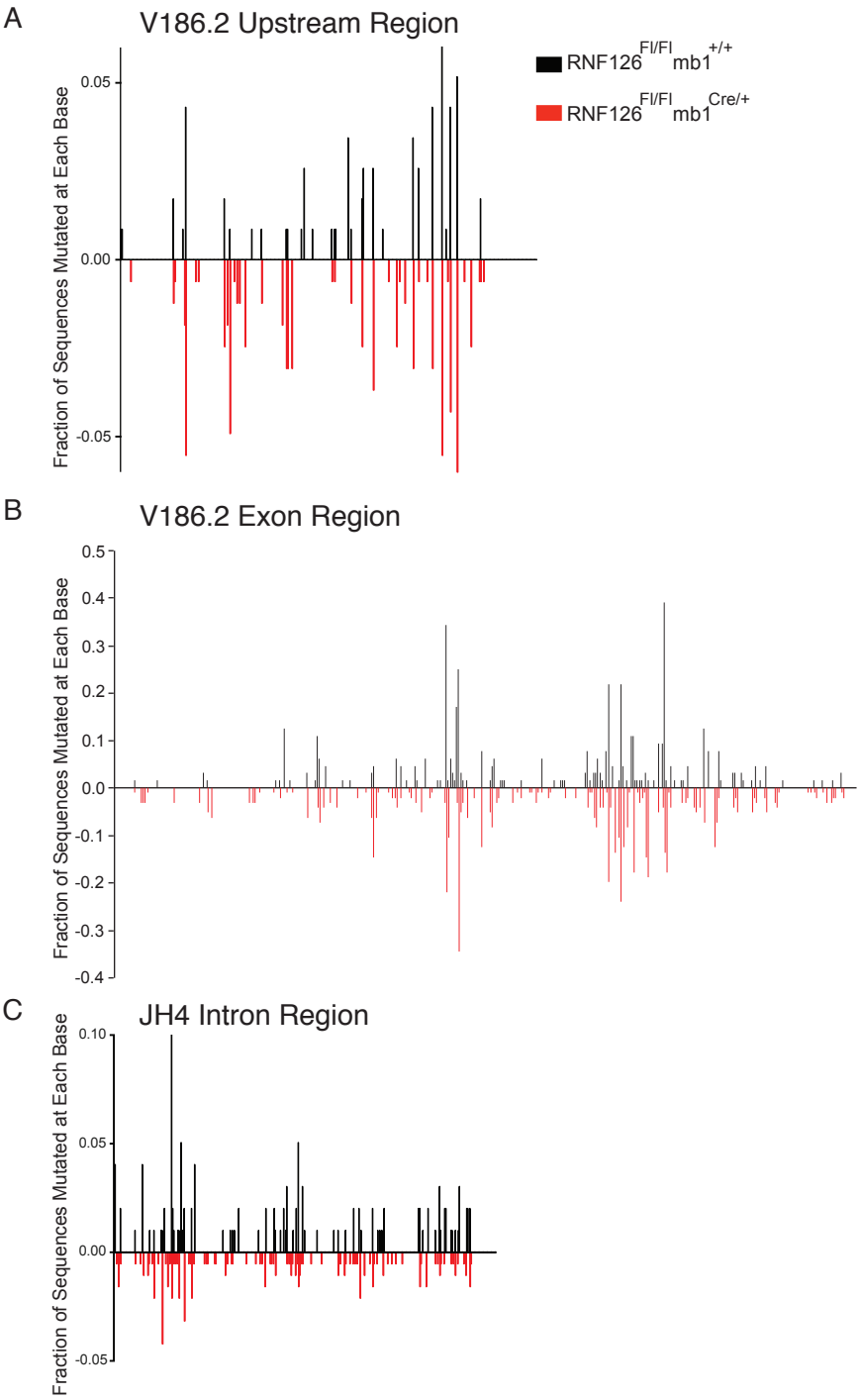


Figure 4.16 The spatial distribution of mutations is unaffected in RNF126 conditional knockout mice. Using SHMTool (Maccarthy et. al., 2009), plots were generated depicting the frequency of mutation at each base sequenced within the three regions analyzed. Each base is represented along the X axis, from 5' (left) to 3' (right). The Y axis depicts the fraction of sequences mutated at each base, including all mutations (non-unique). RNF126FL/FL mb1+/+ data is presented in black, above the axis and RNF126Fl/Fl mb1Cre/+ is presented in red, below the axis.

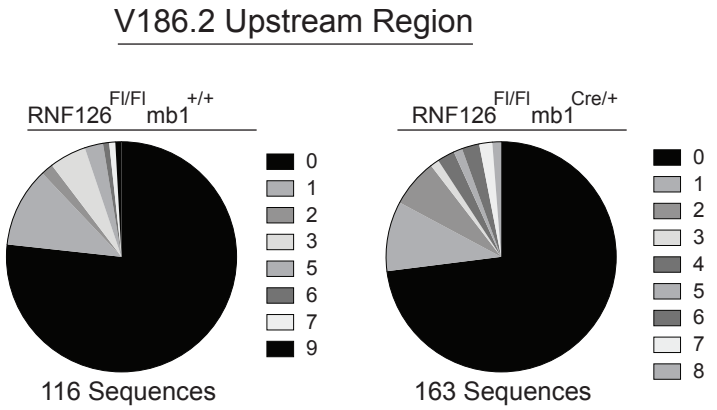


Figure 4.17 The majority of sequences in the V186.2 upstream region are unmutated. Pie charts show the fraction of total sequences that contain between 0 and 9 mutations for rnf126Fl/Fl mb1+/+ and rnf126Fl/Fl mb1Cre/+ mice.

particular gene. The inefficiencies involved in infecting primary B cells, coupled with the short time-frame of CSR experiment (3-5 days) make RNA knock-down studies difficult to conduct in primary B cells. For this reason, the CH12 cell line, which can be induced to switch from IgM to IgA at high rates, can be used. As expected, shRNA knock-down of AID in CH12 cells results in an almost complete loss of CSR (Figure 4.18). Similarly, shRNA knock-down of RNF126 using three different hairpins that target either the 3'UTR or the coding region of the transcript show a significant decrease in the level of CSR (Figure 4.18). This loss in CSR is accompanied by a loss in RNF126 at the protein level (Figure 4.18).

Though there is a slight defect in CSR upon genetic deletion of RNF126 from B cells, it is nowhere near the defect seen in the knock-down studies in CH12 cells. There are several possible explanations for this. First, shRNA knock-down of target genes is notoriously non-specific, especially in mammalian cells. It is common that hairpins targeted against one gene result in a change in gene expression profiles of many genes (Kaelin, 2012). For this reason, multiple hairpins against RNF126 were tested. Second, CH12 cells are an immortal cell line that contains many chromosomal abnormalities. Thus, it is possible, but unlikely, that the mechanism of CSR is different in these cells as compared to primary B cells. And third, the process of gene knock-down in CH12s, which is representative of the mature B cell state is different than a genetic knock-out in B cells at an earlier stage of development. As mentioned, mb-1 Cre is expressed at the Pro-B cell stage and thus RNF126 is knocked-out at the earliest stage of B

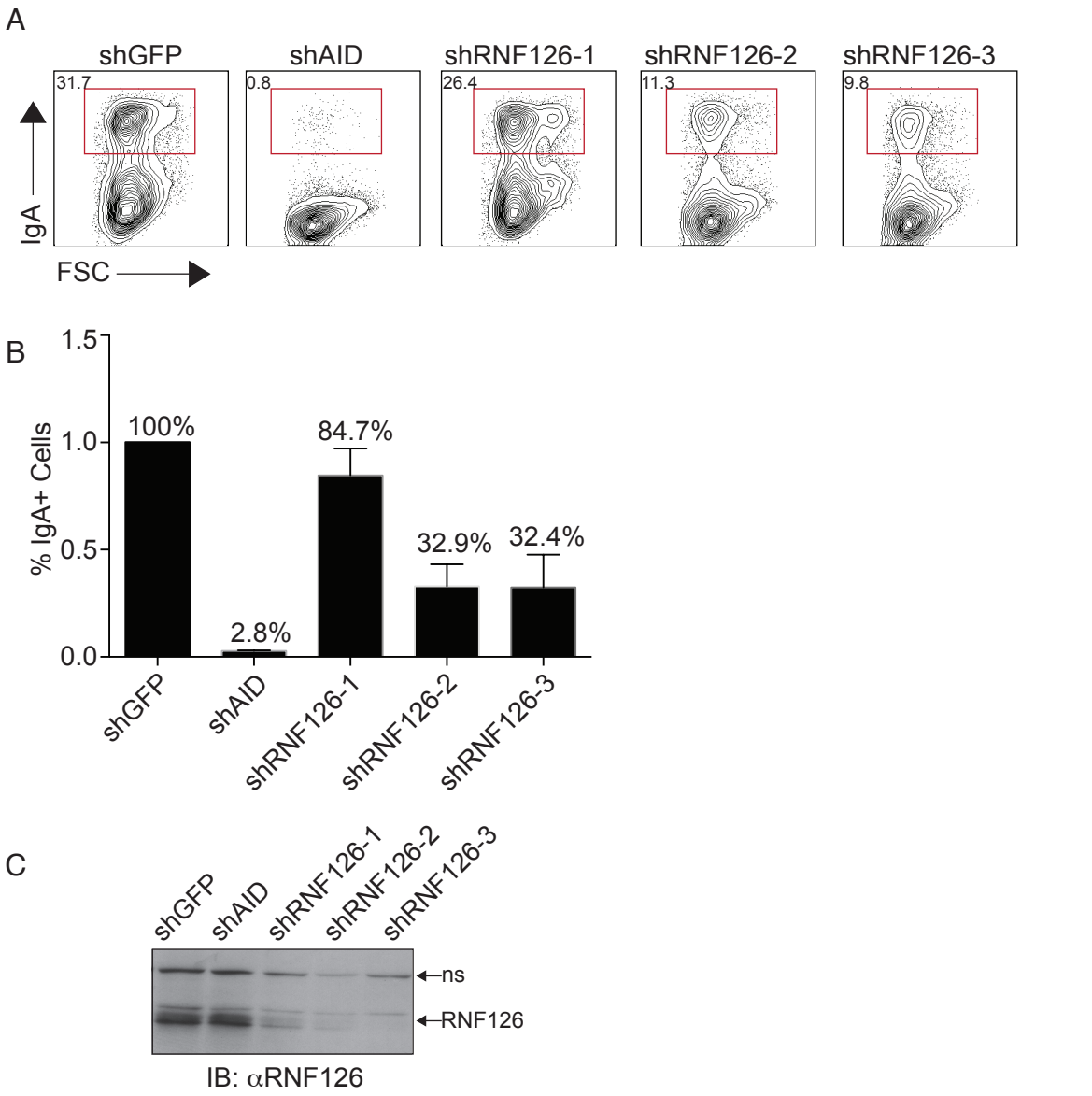


Figure 4.18 shRNA knockdown of RNF126 in CH12 cells results in decreased class-switch recombination. (**A**) Representative FACS plots from one experiment are shown. Knockdown of AID, as compared to the negative control (shGFP) results in an almost complete loss of CSR, as measured by IgA+ cells. In addition, knockdown of RNF126 using three different hairpins results in a decrease in CSR-levels. (**B**) A summary of three independent experiments is shown. CSR levels in the shGFP hairpin expressing line are set to 100. The percent loss of IgA+ cells is depicted above each bar. Error bars are representative of the standard deviation. (**C**) Anti-RNF126 western blot demonstrates that the hairpins against RNF126 (shRNF126-1,2,3) successfully result in a loss of RNF126 protein. Arrows denote (1) a nonspecific band (ns), which can be used as a normalization control, and (2) RNF126.

cell development; however, CSR, SHM and Affinity Maturation are not assayed until the B cells have matured and left the bone marrow. The time difference between knockout and experimentation allows for the possibility of compensation by a homologous protein. In addition to this possibility, expression of a similar protein in mature B cells may lead to redundancy, or the ability of multiple proteins to play one role in the cell. Under physiological settings one protein is often preferred, however the other can compensate upon loss of the first (Barbaric et al., 2007). The difference in results obtained through knock-down and knock-out could support this idea.

CHAPTER 5: DISCUSSION

5.1 Advantages and disadvantages of the solubility-based interaction screen

Protein-protein interaction data provides valuable insight into molecular networks within living cells. Popular methods to obtain protein interaction data in a high-throughput fashion include affinity purification followed by mass spectrometric analysis of proteins co-purified with the tagged protein of interest, and yeast-two-hybrid approaches. These approaches can be best thought of as complementary to one another, as each has distinct drawbacks that the other lacks. For instance, affinity based pull-down methods do not provide information about binary protein-protein-interactions; rather, they describe the assembly of multiple proteins that are stably associated with the tagged bait protein. In addition, these techniques also suffer from an inherent inability to detect transient protein-protein interactions. Overall, affinity based methods tend to be biased toward the identification of large, stable complexes. On the other hand, yeast-two-hybrid screens investigate interactions in the yeast nucleus between overexpressed fusion proteins. It is estimated that over 50% of such interactions are false positives because they cannot be confirmed by other methods such as co-immunoprecipitation (Sprinzak et al., 2003). This approach also generally fails to identify weak or transient interactions that do not sufficiently transactivate the reporter. Thus, yeast-two-hybrid screens seem biased toward the identification of strong binary interactions. Recent progress has been made to attempt to fill the gaps in the interactome left by these more classical techniques. New protein-

protein interaction screens include, but are not limited to, split molecule complementation, which leads to fluorescence upon interaction (e.g. split GFP approaches (Cabantous and Waldo, 2006), chemical and photo-crosslinking followed by specific immunoprecipitation to lock more transient interactions in place (reviewed in (Lowder et al., 2011)), and the fusion of a bait protein with molecules that can modify and "mark" neighboring proteins (e.g. biotin ligases such as BirA (Roux et al., 2012)). Despite the many benefits of these methods, though, none are particularly reliable for the identification of interactions with a protein that is poorly soluble or insoluble when ectopically expressed. The solubility-based screen discussed in this thesis further expands our ability to probe the interactome by specifically targeting this subset of poorly behaved proteins.

There are a few disadvantages of this technique. First, similar to other interaction screens, this assay is unlikely to pick up transient interactions. Because solubilization and drug resistance are the read-outs of this particular screen, a stable interaction between the drug-resistance gene containing fusion protein and the unknown cDNA-derived protein is necessary. In addition, this screen assays the ability of proteins to bind in a non-physiological cell-type. While this may reveal interactions that may not have been picked up through affinity pull-downs, putative interactions must be verified in a more physiological setting. For example, eukaryotic proteins, which may be physically separated due to cellular localization would still be able to bind in a bacterial cell. Lastly, the exposure of hydrophobic domains on proteins often makes them "sticky" and

more likely to bind non-specifically to exposed hydrophobic domains on other proteins. Thus, it is possible that if the co-expressed unknown partner is also poorly behaved, binding and solubilization may observed simply because of the presence of two exposed hydrophobic domains rather than due to a physiological interaction. Despite these disadvantages, though, this screen provides a platform to better understand the interactome of insoluble proteins; putative cofactors identified in the screen can easily be tested for validity.

5.2 Putative AID cofactors identified by interaction screen

The specific motivation behind the development of this assay was to identify interaction partners for AID, a potent DNA mutator highly expressed in B lymphocytes (Muramatsu et al., 1999; 2000) and poorly soluble when expressed in all systems tested. In mature B cells, AID deaminates cytidines at the immunoglobulin locus to initiate a cascade of error prone repair that results in either point mutations (during SHM) or genomic recombination (during CSR), depending on the location within the Ig locus (Longerich et al., 2006). Although AID preferentially targets the immunoglobulin locus, loci elsewhere in the genome have been shown to be targeted at a lower frequency by the detection of point mutations or genomic translocations (Liu et al., 2008; Oliveira et al., 2012; Robbiani et al., 2008; 2009). For this reason, it is likely that AID is tightly regulated. In fact, evidence for regulation at the transcriptional, post-transcriptional and post-translational level has been established (Delker et al., 2009). In addition, the sub-cellular localization of AID is tightly controlled (Brar et al., 2004; Ito et al., 2004; McBride et al., 2004).

Many attempts have been made to identify AID interactors using either affinity-based purification methods or yeast-two-hybrid approaches; as a result, a handful of proteins have been implicated in the AID reaction (Chaudhuri et al., 2004; Conticello et al., 2008; MacDuff et al., 2006). Using the solubility based interaction screen, a number of previously reported AID co-factors have been successfully identified, including components of the ssDNA binding complex, RPA, which are thought to play a role in the recruitment of DNA break repair machinery to the Ig locus (Chaudhuri et al., 2004; Vuong et al., 2009) and CTNNBL1, a splicing factor recently identified by a yeast-two-hybrid approach (Conticello et al., 2008). The importance of splicing of the immunoglobulin gene CSR has been proposed, but the mechanism remains unclear (Hein et al., 1998). In the past, mdm2, a RING domain containing protein that targets p53 for degradation, has been shown to interact with AID (MacDuff et al., 2006); however, full-length mdm2 was not identified in this novel screen, even though the C-terminal (RING domain containing) truncation of mdm2 can clearly solubilize AID as previously demonstrated (Figure 2.2 and (MacDuff et al., 2006)). This is not surprising because the library used here does not contain mdm2 truncations and it is only the very C-terminus RING domain of the protein that was shown to interact in previous yeast-two-hybrid studies (MacDuff et al., 2006).

The fact that this new screening approach has identified known co-factors underscores its validity. More importantly, however, the screen has identified factors not previously known to complex with AID. These include proteins that

are likely to play important roles in DNA repair, such as Rad51 and FEN1. In addition, the RRM2 complex was identified, which is responsible for the synthesis of deoxyribonucleotides during DNA repair (Niida et al., 2010). We have also identified factors with known roles in RNA processing, such as the polyadenylation factor and endonuclease, CPFS73. RNA processing at the immunoglobulin locus has increasingly been recognized as a necessary precursor to the completion of CSR and, although the mechanism is still not understood, there likely exists a tight coordination between transcription at the Ig locus, AID-mediated mutation, and double-strand break repair.

We have also identified a number of ubiquitous factors that are potentially very important for the subcellular localization of AID by either mediating cytoplasmic-nuclear trafficking (such as karyopherin (Patenaude and Di Noia, 2010) and Nup93), or by playing a role in retaining AID in the cytoplasm and excluding it from the nucleus (e.g. TCP1-eta and Bip in analogy to other chaperones (Orthwein et al., 2010). Finally, a number of factors that are of completely unknown function have been identified and it will be important to determine their relevance to CSR and SHM in the future.

5.3 Potential roles for RNF126 and mono-ubiquitinated AID

RING Finger Protein 126, a very interesting AID interacting protein, was revealed using this screen. In addition to verifying that RNF126 is a *bona fide* AID interactor in bacterial and mammalian cells (Figures 3.3, 3.4, 3.5, 3.6), it was determined that it acts as an E3 ubiquitin ligase in a complex that mono-ubiquitinates AID (Figures 3.8, 3.10). In addition, other E3 ligases expressed in B

cells with specific functions in B cell biology do not appear to have such activity, arguing for selectivity (Figures 3.13, 3.14). Though AID ubiquitination has been noted previously by Reynaud *et al.* (Aoufouchi et al., 2008), the type of ubiquitination observed here (mono-) is quite different than the poly-ubiquitination observed previously. Thus, it is likely that, like phosphorylation, ubiquitination of AID is a more general PTM used to regulate AID in a variety of different ways.

A number of roles for RNF126 in the context of AID and antibody diversification can be hypothesized based on sequence similarities of the domains of this E3 ligase with homologous domains of known function present in better characterized proteins. For example, in addition to the RING domain, which interacts with and modifies AID (Figures 3.5, 3.11), RNF126 contains several intriguing domains that could be suggestive of function. For instance, its N-terminal zinc finger domain is homologous to a recently identified ubiquitin-binding domain (Bacopulos et al., 2012) potentially suggestive of regulation. The co-existence of a ubiquitin ligase and ubiquitin binding domain within the same protein is reminiscent of the proposed regulation of translesion DNA repair. Ubiquitination of pol η is thought to result in a closed conformation through the interaction of the ubiquitin moiety with the ubiquitin binding domain at the N-terminus of the polymerase. As a result, this prevents the interaction of pol η with PCNA. Thus, ubiquitination can either positively or negatively regulate the assembly of the repair complex, depending on which substrates are targeted by ubiquitin (Bienko et al., 2010). Whether this phenomenon occurs in the biology of RNF126 still needs to be determined.

Lastly, RNF126 contains a conserved stretch of serines at the very C-terminus (Figure 3.2). Though the function of this domain is unknown at present, it is homologous to similar stretches of serines found in proteins that have been shown to act as transcriptional activators at RNAPII-dependent promoters (Bates and DeLuca, 1998; Kretzschmar et al., 1994; Miau et al., 1997). These include viral proteins, which bind to general transcription factors in order to hijack the transcription machinery of the host cell (Bates and DeLuca, 1998), but also include cellular proteins, such as the transcriptional coactivator, PC4 (Ge and Roeder, 1994; Ge et al., 1994) and other proteins with roles in transcription (Miau et al., 1997). Given the multitude of genetic and proteomic data linking AID to transcription initiation, elucidation of the role of this particular domain of RNF126 in the context of a potential dual interaction with the Ig promoter and AID will be interesting to pursue.

This thesis presents an analysis of only the RING domain of RNF126 and its activity in ubiquitination. The role that these other domains play in the biology of RNF126, itself, as well as in AID biology remains a very interesting avenue of research to pursue in the future. The generation of the RNF126 conditional knockout mouse makes it easier to investigate the role of individual domains on RNF126. Bone marrow reconstitution of RNF126 conditional knockout mice with mutant forms of the protein (for example, with mutant Zn Finger domains, or truncations of the serine tail) can be used to determine the role single domains play in SHM and affinity maturation.

The role of RNF126-mediated ubiquitination and of mono-ubiquitinated AID in antibody diversification can be ascertained from an analysis of the changes in each step of antibody diversification upon loss of RNF126. Of course, the role of RNF126, itself, may go beyond ubiquitination of AID and this must be taken into consideration. Despite this, hypotheses derived from experimental evidence regarding the role of RNF126 in antibody diversification are presented below.

5.3.1 In Somatic Hypermutation

The relatively uncharacterized protein, RNF126, has been identified as an E3 ubiquitin ligase with the ability to mono-ubiquitinate AID (Delker et al., 2012). Ubiquitination, which was originally thought to only play a role in protein degradation, has been shown to play a variety of roles in many cellular processes (Pickart, 2001b), thus suggesting that RNF126 may play an interesting regulatory role for AID and AID-mediated steps of antibody diversification. It has already been established that mono-ubiquitination is important for both somatic hypermutation and class-switch recombination. Mono-ubiquitination of the polymerase processivity factor, PCNA, on lysine 164 (K164) has been shown to recruit translesion synthesis polymerases necessary for error prone repair. Mice that express ubiquitin-null mutants of PCNA (K164R) display reduced class-switch recombination *ex vivo* and a reduction in mutations at A:T base pairs during *in vivo* somatic hypermutation (Roa et al., 2008). In addition, the mono-ubiquitination of the histones H2A and H2AX by RNF8 is necessary for the recruitment of RNF168 and subsequent RNF168-mediated poly-ubiquitination.

Both events are important for the recruitment and stabilization of the repair factor, 53BP1, and faithful DSB repair (Doil et al., 2009; Jackson and Durocher, 2013; Mailand et al., 2007; Mattiroli et al., 2012; Ramachandran et al., 2010).

Thus, given these examples, it is plausible that mono-ubiquitination of AID plays a similar role in recruiting necessary repair proteins to the Ig locus. In line with this hypothesis, loss of RNF126 results in an alteration of the pattern of mutations observed during SHM. Most notably, in the JH4 region, a greater mutation load was observed on the template strand in the absence of RNF126. This suggests that either AID, itself, or error-prone repair factors gained greater access to the template strand upon loss of RNF126. However, based on studies conducted in mice lacking both UNG and MSH2, it appears that AID has equal access to both template and non-template strands (Xue et al., 2006). Thus, it is more likely that the difference observed is due to access of downstream repair proteins. The only other instance in which the template strand was observed to accumulate more mutations was in patients deficient for NBS1, a member of the MRN complex (Du et al., 2008). In addition, overexpression of NBS1 in the murine B cell line, RAMOS, increases the strand bias of mutations toward the non-template strand (Yabuki et al., 2005). Thus, through an unknown mechanism, NBS1 prevents error-prone repair on the template strand. It is possible then that RNF126, and possibly ubiquitinated AID, functions in the same pathway. It is conceivable that, in contrast to ubiquitinated PCNA, which recruits error-prone repair, mono-ubiquitinated AID results in the recruitment of error-free repair, thus contributing to the local battle between error-free and error-prone

repair. It is also conceivable that RNF126 plays a role at the G1-to-S cell cycle check point. Persistence of U:G mismatches into S-phase, where repair is thought to be dominated more by error-prone repair (Li et al., 2012), could result in the observed increase in template strand mutations.

Though the same alteration in strand bias was not observed in the V186.2 exon region, the altered mutation rate observed here also supports this hypothesis. In the absence of RNF126, an increase in the number of mutations per B cell clone was observed. As mentioned, this could represent overactive AID or the presence of AID outside of the germinal center, but could also be representative of an increase in error-prone repair or the persistence of U:G mismatches into S-phase. If RNF126 promotes error-free repair, then in its absence, a greater number of mutations would accumulate during the process of replication and the generation of new B-cell clones.

It would be very interesting to test these hypotheses directly. Schatz *et al*. (Unniraman and Schatz, 2007) have generated transgenic mice, which contain a SHM substrate containing a tract of A/T bases with either one central C (NT strand): G (T strand) basepair or one central G (NT strand): C (T strand) basepair. These mice enabled them to determine that, while AID could efficiently deaminate the C base on both strands, A/T mutations only accumulated when the deaminated C lay on the NT strand. Thus, strand bias is determined mainly by the access of error prone repair proteins (Unniraman and Schatz, 2007). Utilizing this system, it would be predicted that in the absence of RNF126, A/T

mutations would accumulate equally well with the deaminated C on the T strand, as compared to the NT strand.

In addition, if RNF126 is truly involved in the recruitment of error-free repair, then it likely plays a role outside of the Ig locus to correct for off-target AID mutation. An analysis of off-target mutation load and translocation frequency in the RNF126 conditional knockout mouse model would informative on this front.

5.3.2 In Class Switch Recombination

It is much more challenging to discern the role of RNF126 during CSR because the phenotype in the RNF126 conditional knockout with regard to CSR is much more subtle. Though there does appear to be a subtle defect in CSR in the knockout B cells, it is not statistically significant. This is further complicated by the fact that there is a much greater defect upon knockdown of RNF126 in a B cell line (discussed more in section 5.4). While it is possible that RNF126 performs the same function in CSR as it does in SHM, it is also important to remember that this is not necessarily the case. Precedent for this exists for the binding factor, RPA, which likely stabilizes ssDNA during SHM, but was co-opted to play a DNA repair role during CSR. Further, the mono-ubiquitinated form of AID generated by RNF126 may play two separate roles in SHM and CSR. An analysis of mutations generated in the switch regions during CSR in wildtype and RNF126 knockout B cells would help determine if the mutation pattern is altered as was seen during SHM. Further, *in vitro* deamination assays on transcribed SHM targets compared to transcribed CSR targets (Switch regions) in the presence and absence of RNF126 would help to elucidate the presence of an

inherent difference in the role of RNF126 in regulating AID activity on distinct substrates.

5.4 The possibility of redundancy and/or compensation in the RNF126 conditional knockout mouse model

Often the *in vivo* function of unknown genes/proteins is approached genetically through the generation of knockout mouse models. This approach, however, can be complicated by the presence of genetic robustness, or the ability of an organism to sustain either environmental or genetic changes and maintain normal function. Genetic robustness can occur through two main pathways—(1) genetic buffering, or the use of alternative pathways to accomplish the same goal, and (2) functional complementation, where two genes have redundant functions (Barbaric et al., 2007; Gu, 2003). Studies in *S. cerevisiae* and *C. elegans* have shown that genes that exist as single copies with no sequence similarity to other genes have a greater likelihood of producing a phenotype when knocked out or knocked-down with RNAi (Gu, 2003; Gu et al., 2003; Kamath et al., 2003). Furthermore, several examples of compensation exist in knockout mouse models. For example, redundant, but not completely overlapping, functions exist for the genes MyoD and Myf5, which are both involved in skeletal muscle development. Mice lacking both MyoD and Myf5 do not develop skeletal muscle and die shortly after birth; however, mice lacking either of the individual genes have phenotypes ranging from normal to slightly perturbed (Barbaric et al., 2007; Braun et al., 1992; Rudnicki et al., 1992; 1993). In fact, upon MyoD inactivation, Myf5 is upregulated, presumably to account for

the loss of MyoD (Rudnicki et al., 1992). Other examples include knockouts of individual genes within highly related gene families—for example, knockouts of single caspase genes result in a variety of phenotypes ranging from prenatal lethality (caspase 8) to no detectable phenotype (caspase 12) (Barbaric et al., 2007). Compensation and redundancy have even been predicted to play a role in masking additional functions for AID. Various studies have suggested that AID may play a role in DNA demethylation by deaminating a methylated cytosine residue to generate a T:G mismatch, which is repaired to an unmethylated cytosine (Fritz and Papavasiliou, 2010); however, the *aicda*$^{-/-}$ mouse model does not have any observable defects outside of its canonical role in antibody diversification. To reconcile this difference, it has been shown that, while *aicda*$^{-/-}$ cells are functional in a reprogramming assay, cells in which AID has been knocked-down are not. This suggests that compensatory mechanisms that exist in the context of a knock-out may be overcome by the shorter time-frame of knock-down experiments, thus drawing a distinction between genetic and “acute” losses of genes/proteins (Bhutani et al., 2012). It should be noted that more recent studies addressing the role of AID in reprogramming have found that AID-null cells less proficiently maintain and stabilize a stem-cell like state 4 weeks post reprogramming and that many genes at this time point are found to be hypermethylated as compared to wildtype induced pluripotent stem cells (iPSC). Thus, this suggests that AID does have an alternate role in DNA demethylation (Kumar et al., 2013). However, these findings do not invalidate the earlier

studies, which showed that a knock-down of AID results in a loss of reprogramming.

Because many knock-out mice without obvious phenotypes are not published, it is hard to estimate how many null mutants result in no observable phenotype; however, it has been estimated that approximately 10-15% of knock-out mice made do not show a phenotype, likely due to the presence of redundancy (Barbaric et al., 2007). Thus, it is very likely that the subtle defects in antibody diversification observed in RNF126 conditional knock-out mice are due to imperfect compensation by other proteins present in the cell. Similar to the scenario in which acute loss of AID reveals potential functions not seen with the knock-out mouse model, knock-down of RNF126 in a murine B cell line results in a much more significant decrease in CSR levels as compared to that seen in the conditional knock-out model. It remains to be seen whether these discordant results are due to a difference in cell type, primary cells versus CH12 cell line, or whether compensation that has occurred in the mouse system is not present in the cell line. Because Cre was expressed from the mb-1 locus, RNF126 was knocked out at the pro-B cell stage; however, antibody diversification was not assayed until the mature B cell stage, allowing ample time for compensation to occur, which would preferentially occur if RNF126 normally plays a role during the development process. Experiments are underway to address this issue. Use of mice, which ubiquitously express the Cre recombinase fused to a mutant estrogen receptor (Cre-ERT), can be used to induce deletion of a gene of interest upon treatment of cells with the estrogen-mimic, tamoxifen (Feil et al., 1997).

Thus, treatment of splenic B cells *in vitro* directly prior to CSR stimulation will eliminate the possibility of compensation during B cell development. Unfortunately, it is much more difficult to get at the question of redundancy in mature B cells. The generation of multiple gene knockouts in one mouse, or a combination of gene knockout with gene knockdown may reveal a stronger phenotype if redundancy is, in fact, occurring.

In the 293T ubiquitination assay developed, BCA2/RNF115, an E3 ligase that is highly homologous to RNF126, cannot ubiquitinate AID (Delker et al., 2012). However, it still remains a possibility that, in the absence of RNF126, BCA2/RNF115 could acquire the ability to modify AID. In addition, hundreds of E3 ligases are expressed in activated B cells, allowing for the definite possibility of compensation. At this point it is still uncertain what E3 ligase is actively ubiquitinating AID *in vivo*, though the ability of RNF126 to do so *in vitro* suggests its function *in vivo*. Transcriptome-wide comparisons of gene expression in wildtype versus RNF126 knockout B cells could provide evidence for the up-regulation of alternate E3 ligase enzymes, which would be likely candidates for compensatory mechanisms; however, it is also possible that redundancy can occur with an enzyme that is already highly expressed in activated B cells.

Future studies of the role of RNF126 in antibody diversification would likely benefit from the generation of knock-in mice containing mutant forms of the RNF126 gene rather than knockout mice. The complete loss of the protein likely contributes to the onset of compensation and/or redundancy. Thus, the presence of a mutant form of the protein may inhibit this. Knock-in mice containing RING

domain mutants could be used to study the role of RNF126-mediated ubiquitination and mice containing mutations of the Zn Finger or truncations of the serine tail could be used to determine the roles of these other intriguing domains.

5.5 A model for the role of RNF126 during Antibody Diversification

The change in the measured levels of CSR and SHM observed in the RNF126 conditional knockout as compared to wildtype is subtle, at best; however, the pattern of mutations observed at the Ig locus during SHM is significantly altered. Thus, it is likely that the system of Antibody Diversification has evolved several mechanisms that function in parallel to account for the possible loss of one component. Given the alteration in the pattern of mutations in the JH4 region, there are several hypotheses that can be generated to explain the role of RNF126 during SHM. First, it is possible that RNF126 participates in a complex that prevents access of either AID or downstream repair factors to the template strand of DNA. Thus, in the absence of RNF126, AID and/or mutagenic repair factors, such as pol η, would gain greater access to the template strand, thus accounting for the observed shift in strand bias observed. Second, as mentioned previously, it is possible that RNF126 actually promotes error-free repair on the template strand, thus targeting error-prone repair on the non-template strand. Thus, in the absence of RNF126, error-prone machinery can extend to the template strand, resulting in the observed mutation pattern. Though the benefit of limiting mutagenic repair to one strand of DNA is not directly clear, it is possible that this type of mechanism could act to limit the

frequency of DSB formation in the variable region during SHM. As such, this could be a way to differentiate the repair processes that promote the formation of point mutations during SHM and the formation of DSBs during CSR. In addition to these two hypotheses, which both suggest that RNF126 plays a role in preventing mutation on the template strand, it is also simply possible that RNF126 is involved in the recruitment of known, canonical repair factors, such as Pol η. Then, in the absence of RNF126, non-canonical repair enzymes would be recruited to the Ig locus in an attempt to compensate for the loss of RNF126, thus producing an altered pattern of mutation. This third model opens the exciting possibility that, analogous to ubiquitinated PCNA, RNF126-mediated ubiquitinated AID is necessary for the proper recruitment of repair factors to the Ig locus. These three models are depicted in Figure 5.1.

In order to test the validity of these three models, further experimentation will be necessary. It will be useful to cross the RNF126 conditional knockout mouse to mice containing knockouts of various translesion polymerases, such as pol η, pol κ, pol ι. If, in the context of a double-knockout, the altered pattern of mutation is no longer observed, then it can be reasonably assumed that the targeted polymerase works with RNF126 in generating the pattern of mutation. Further, chromatin immunoprecipitation (ChIP) experiments for known repair factors can be conducted in the presence and absence of RNF126 to determine if the factors recruited to either the variable or switch regions is altered in the absence of RNF126.

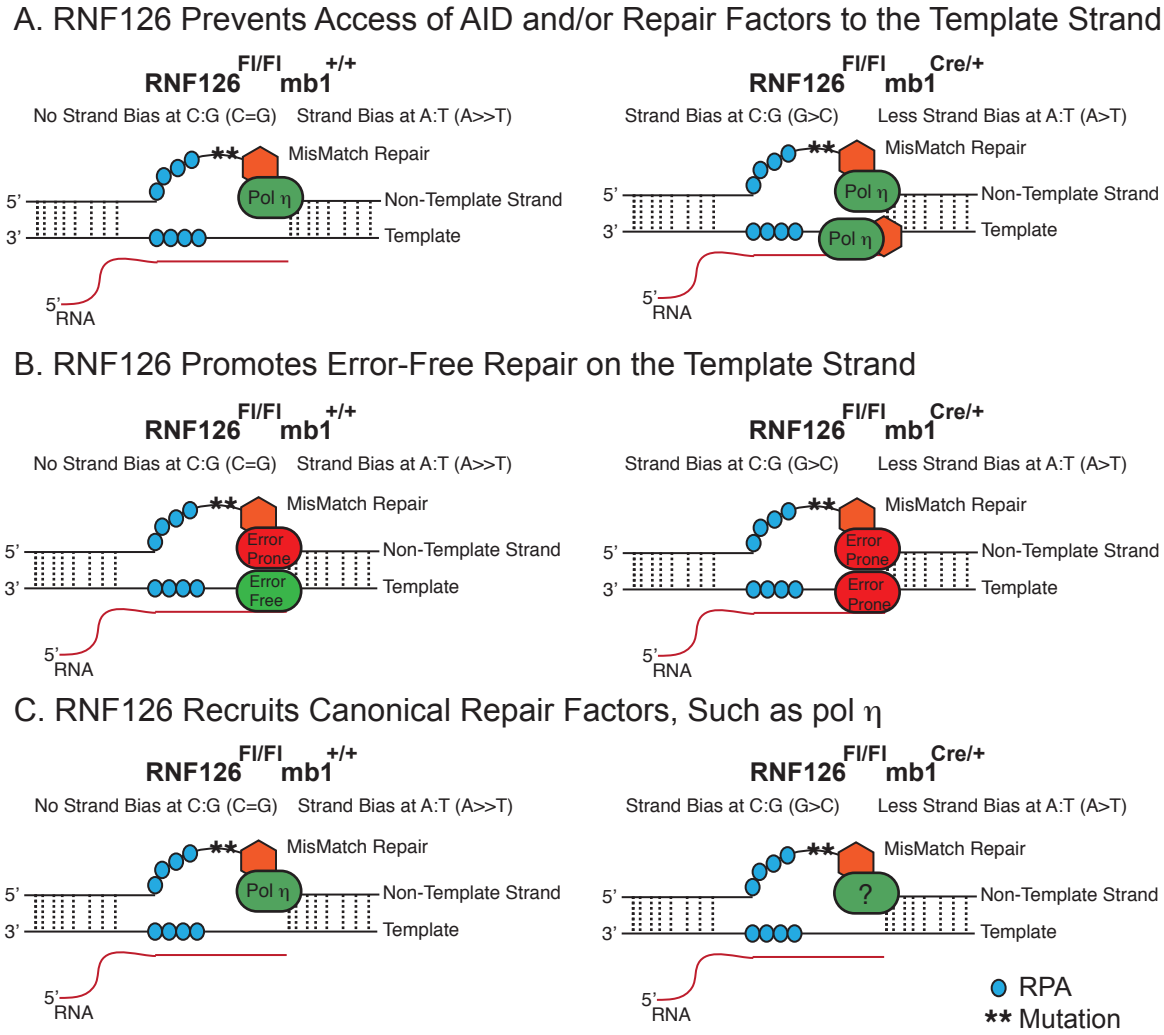


Figure 5.1 Potential Models to Explain the Role of RNF126 in Antibody Diversification. Images depicted represent a scenario for wildtype mice (left), in which there is no strand bias at G:C basepairs (G=C) and strand bias at A:T basepairs (A>>T) and RNF126 conditional knockout mice (right), in which there is strand bias at G:C basepairs (G>C) and less strand bias at A:T basepairs (A>T**). A**. RNF126 prevents access of AID and/or repair factors to the template strand. Thus, in the absence of RNF126, error-prone repair occurs on both the non-template and template strands. **B**. RNF126 promotes error-free repair on the template strand, thus targeting error-prone repair to the non-template strand. In the absence of RNF126, it is possible that error-prone repair could extend to the template strand. **C**. RNF126 recruites canonical repair factors, such as pol η. Thus, in its absence, non-canonical polymerases are recruited to compensate for the loss, resulting in the altered pattern of mutation.

In addition, though, it will be important to determine if the observed phenotype in the RNF126 conditional knockout mouse is due to a loss in RNF126-mediated ubiquitination of AID. Along with experiments to determine the involvement of known translesion polymerases, *in vitro* studies can be conducted to determine if ubiquitinated AID can directly interact with each of the translesion polymerases, thus providing a link between RNF126 and repair. Furthermore, when the residue on AID that is targeted by ubiquitination is determined, mice harboring ubiquitin-null mutants of AID can be generated. Analogous SHM and CSR assays can then be conducted in the context of a ubiquitin-null mutant of AID to determine if the same phenotype exists as was seen in RNF126 conditional knockouts. This would provide reasonable evidence that the phenotype observed upon loss of RNF126 is due to a loss in RNF126-mediated ubiquitination of AID.

Thus far, these hypotheses focus specifically on the role of RNF126 during SHM. A more thorough analysis of the pattern of mutations in the switch regions during CSR in RNF126 conditional knockout mice will be necessary to determine if these hypotheses can extend to CSR, as well.

5.6 Functions for RNF126 outside of the immune system

Additional substrates for RNF126 in other cellular contexts have been identified recently, providing excitement for the utility of the RNF126 conditional knockout mouse model as a generalized tool to study this E3 ligase throughout cell types in the mouse. First, RNF126 has been implicated in cancer cell proliferation by targeting the cell cycle regulator, p21, for ubiquitin-mediated

degradation (Zhi et al., 2013); second, RNF126 and the homologous E3 ligase, BCA2/Rabring7/RNF115, have been shown to be involved in the internalization and endosomal sorting of the epidermal growth factor receptor (EGFR) (Smith et al., 2013). In addition, expression of RNF126 has been documented in a very specific portion of the hippocampal region of the mouse brain, suggesting a potential role for this E3 ligase in memory formation or brain function (Lein et al., 2007). By generating crosses of the RNF126-floxed mouse with cell-type-specific CRE expressing mice, the importance of RNF126 can be investigated in each of these systems.

5.7 Looking forward

The work presented here creates a foundation for future research both in terms of the novel tools developed, as well as in discoveries made. First, the solubility-based protein interaction screen developed can be applied to any insoluble proteins of interest, thus making it a useful tool to study proteins in all biological pathways. Second, the RNF126 conditional knockout can be bred to cell-specific CRE expressing mice in order to study the role of RNF126 in many different cellular contexts. In addition to these tools, though, this work has also identified a novel mode of regulation of the enzyme, AID. The regulation of AID likely involves a complex interaction of a variety of cellular factors, including both protein and RNA. In addition, it has become clear that there must exist a balance between error-free and error-prone repair downstream of AID-induced mismatches and how this decision is made remains unknown. The finding here that RNF126 mono-ubiquitinates AID expands the list of AID regulatory

mechanisms and provides a foundation for future studies to elucidate the exact role of this modification during the various stages of antibody diversification. Ubiquitination is a commonly used mechanism to regulate cellular processes so it is not unexpected that it is utilized during AID biology; however, as has been discussed, it will be interesting to determine if ubiquitination of Ig-locus associated proteins, for example AID versus PCNA, provides a mechanism for the decision made between different types of repair pathways. The findings presented here provide the foundation and tools necessary to begin to better understand the role of ubiquitinated AID in antibody diversification.

CHAPTER 6: EXPERIMENTAL METHODS

6.1 Pertaining to all *in vitro* work: CHAPTERS 2 and 3

6.1.1 Mice and Cells

Wildtype and *aicda*$^{-/-}$ C57BL/6 mice between the ages of 8-12 weeks were used for *in* vitro class switch recombination assays. For bacterial expression experiments BL21DE3 *E. coli* cells were used. For mammalian expression HEK 293T cells were used.

6.1.2 Antibodies Used

The following antibodies were used for this study: anti-Flag (Sigma M2 clone, beads and HRP conjugate), anti-HA beads (Sigma EZ View Red HA beads), anti-HA (Roche, clone 3F10), anti-HA HRP (Miltenyi, cat# 130-091-972), anti-His HRP (Santa Cruz, cat# 8036), anti-AID (Cell Signaling, EK25G9 clone), anti-RNF126 (Sigma, cat# HPA043050), anti-H3 (abcam, cat# 1791), anti-PCNA (Santa Cruz, Clone PC10, cat# 56), anti-Tubulin (Sigma, clone DM1A).

6.1.3 Primers Used

AID-Forward: 5′-atggacagcctcttgatgaaccg-3′

AID Linker-Reverse: 5′-

gcatccatGAGGGGAAGATGTCCCTGCACATTaagtcccaaagtacgaaatgcgtctcg-3′

Linker CAT-Forward: 5′-

ggacttAATGTGCAGGGACATCTTCCCCTCatggatgcaaaacaaacgcggcag-3

CAT-Reverse: 5′-ttatttggacgttctacgctgcgtataaatcgcatccatc-3′

RNF126 Q1-F: 5’-GCAGCCCGGACGGTACT-3’

RNF126 Q1-R: 5’-AGCTCCTCAATGAAGCCAGACT-3’

RNF126 Q2-F: 5'-CCCCCACCGACCAGAAC-3'

RNF126 Q2-R: 5'-ATCGTCGAAGATGCCAAAGG-3'

CD19 QPCR-F: 5'-AGAAGGAAAAGGAAGCGAATGA-3'

CD19 QPCR-R: 5'-GGAGAGCACATTCCCGTACTG-3'

6.1.4 Plasmids Used

A. Screen. The AID-CAT fusion gene was assembled by PCR. The human AID cDNA and the chloramphenicol transferase (CAT) cDNA were amplified by standard PCR using the primer pairs AID-Forward/AID-Linker-Reverse and Linker-CAT-Forward/CAT-Reverse, respectively. This was followed by an overlap-extension PCR using the AID-Forward and CAT-Reverse primer pair. AID-CAT was then cloned as an NdeI/XhoI fragment into the second expression cassette of a pCDF-duet vector (Novagen). The first expression cassette contained either a fragment of Mdm2 (a gift from Dr. Reuben Harris, University of Minnesota) or candidate interactors.

The fragment of Mdm2 was also cloned into a pCOLA-duet vector (Novagen) as a SalI/NotI fragment. Similarly, a number of unrelated virulence factors (cagA, sopB, sopE, sifA, invB, invC, cdtA. cdtB, yopK, yopJ, yopH and others) were cloned to generate the mock mini-library of factors that are highly unlikely to interact with AID.

Finally, the CAT-ADAT3 construct was generated similarly to that of AID-CAT except that it was an N-terminal fusion. The C terminal fusion (ADAT3-CAT) sterically hindered the interaction with ADAT2, ADAT3's only known co-factor.

The cDNA library was derived by Invitrogen from the RAMOS Burkitt's lymphoma cell line, which hypermutates constitutively. To generate the destination vector for the cDNA library, cDNA encoding the AID-CAT fusion protein and the attR1/attR2 recombination cassette containing the ccdB gene were sequentially cloned into the second and first cloning site of pCDF-Duet (Novagen) using the restriction enzyme combinations NdeI/XhoI and NcoI/NotI, respectively. The chloramphenicol resistance cassette located within the attR cassette was removed by cutting with SacII/BamHI followed by religation. One Shot® ccdB Survival™-T1R Chemically Competent Cells (Invitrogen) were used to propagate the destination vector.

B. Bacterial Expression. We used the pCDF-Duet vector (Novagen) to clone C-terminally Flag tagged AID into the second multiple cloning site using NdeI/XhoI. The first multiple cloning site was used to clone various putative interacting factors (including RNF126). The same vector was also used to clone RNF126 on its own, for purification purposes, or fragments of RNF126 together with AID for domain delineation and interaction purposes.

C. Mammalian Expression. The mammalian expression vector pcDNA3 (Invitrogen) was used for transient expression of hAID, hRNF126 and ubiquitin (and mutants thereof) in HEK 293T cells. In these constructs AID is either untagged or FLAG-tagged, as indicated for each experiment. RNF126 is HA-tagged. Ubiquitin is either untagged or FLAG-tagged, as indicated for each experiment. Flag-PCNA, HA-RNF8 and HA-BCA2 were expressed from the pCMV-Sport6 vector (Invitrogen).

6.1.5 Screening Protocol

Destination vectors based on pCDF-duet and containing the attR1-ccdB-attR2 construct in the first expression cassette and AID-CAT in the second, were directionally recombined with the cDNA library (22,000 cDNAs, split into 42 pools containing 800-1000 unique colonies per pool) using the Gateway system (Invitrogen). Each recombination reaction was then transformed into BL21DE3ai bacteria (Invitrogen) and plated on spectinomycin plates (50μg/ml) and grown at 30°C to determine transfection efficiency. Spectinomycin plates with the expected ~800 colony number were then replica-plated onto induction plates (using velvet pads) containing chloramphenicol (60μg/ml). Induction was achieved by addition of arabinose (0.02%) to the plates (which controls T7 RNAP expression in BL21DE3ai cells) and by addition of IPTG (0.5mM, which controls T7-driven expression of both cassettes of the pCDFduet vector, namely, the cDNA clone and the AID-CAT gene). Induction occurred at 30^{o}C overnight (until colonies were visible). We cannot estimate toxicity because death due to toxicity is indistinguishable from death due to lack of solubility and thus lack of chloramphenicol resistance; both scenarios result in a lack of colonies. This is a limitation of this assay; however, we did try to minimize minor toxicity by growing non-induction and induction plates at 30^{o}C overnight. Colonies that survived on induction plates were picked, grown in liquid culture (again under induction conditions and at 30^{o}C overnight). DNA from these cultures was then prepped and sequenced to identify potential interactors.

We aimed for a number of colonies per spectinomycin plate consistent with the expected number of clones per pool (800-1000); however, the number of colonies recovered on induction plates was variable presumably due to the variable number of potential interactors per pool.

Interactors that were recovered multiple times or a few times, but reproducibly in each of three independent screens, are shown in Table 3.1 along with their identity and accession number. Each of these hits was subsequently validated through co-purification of each cDNA of interest (in frame with 6xHis in the pCDFduet vector) with AID-FLAG (cloned in the second cassette of the pCDFduet vector, in lieu of AID-CAT). We were not able to validate the interaction of 6/36 candidates, which corresponds to the 16% false positive rate. This information is also included in the supplemental table in the column titled “Co-IP?”. Those that are marked with a “Y” were able to be validated and those that are marked with an “N” were not.

6.1.6 Bacterial Co-Expression and Co-purification

For co-expression and co-purification experiments, the coding sequence for C- terminal FLAG-tagged AID was initially cloned into the second multiple cloning site of pCDF-Duet (Novagen) with NdeI/XhoI (following the original configuration of the screening setup). The first multiple cloning site was used for cloning the cDNAs of candidate interactors in frame with the hexahistidine tag using the SalI/NotI sites. For further co-purification experiments using AID and RNF126, this configuration was used as well as the opposite (6xHisAID /

RNF126-FLAG, to assuage fears of non-specific precipitation of AID onto FLAG-agarose beads). In all cases, plasmids were transformed into BL21 DE3 cells and protein expression was induced with 1mM IPTG for 12-16 hours at 23°C after the OD_{600} reached 0.8. Bacterial pellets were collected by centrifugation and directly resuspended in buffer A (20 mM HEPES [pH=7.0], 200 mM KCl, 1 mM PMSF), frozen and stored at -80°C. Prior to protein purification, partially lysed cells were thawed, treated with a non-specific nuclease (Benzonase; Novagen) and then fully lysed using a cell disruptor (French press). The lysate was centrifuged and the resulting supernatant was loaded either onto a Talon resin charged with Co^{2+} ions (BD Biosciences) or on anti-FLAG beads (Sigma). Talon resin purification was done as follows: The resin was first washed with buffer A followed by buffer B (buffer A plus 30 mM Imidazole) and eluted with buffer A containing 150mM or 250mM imidazole. The eluted proteins were immediately dialyzed against buffer A to remove the imidazole. Alternatively, lysates were first pre-cleared by incubation with agarose beads and subsequently incubated with anti-FLAG beads for 2 hours at 4°C. Beads were then washed at least five times with buffer A. The captured protein was eluted by incubating the beads with buffer A containing FLAG peptide.

6.1.7 Mammalian Co-expression and Co-purification

293T cells that were plated at 60-75% confluence in a 6 well dish were transfected with 1.5 μg of either pCDNA3.Flag-AID, pCDNA3.HA-RNF126 or both. 48 hr post transfection cells were lysed with 300 μL of Cytosolic Lysis

Buffer (50 mM HEPES, pH 7.2, 10% Glycerol, 1% NP-40, 10 mM KCl, 1 mM $MgCl_2$), sonicated for 10' in 30'' intervals and treated for 1hr with Benzonase (Novagen). The nuclear pellet remaining after this treatment was resuspended and lysed in high salt buffer (Cytosolic Lysis Buffer + 400 mM NaCl). Both lysates were combined and the remaining cellular debris removed by centrifugation at 13.2K RPM. Protein concentrations were normalized between samples prior to performing immunoblots and immunoprecipitations. Approximately 8% of the lysate was loaded as "Input," and the remainder immunoprecipitated with anti-HA (specific for RNF126) or anti-Flag (specific for AID) beads overnight. Each IP was washed 3X with Cytosolic lysis buffer followed by 3X washes with High Salt buffer. Laemmli buffer was added to the beads, which were then boiled and loaded on the gel.

6.1.8 B-Cell Purification and Activation

Naïve splenic B cells from 8-12 week old wildtype and *aicda*$^{-/-}$ C57BL/6 mice were purified by CD43 negative selection on a magnetic column (MACS, Miltenyi Biotec). Purified B cells were plated at a concentration of 0.5 X 10^6 cells/mL and stimulated with 10 ng/mL IL-4 and 25 μg/mL LPS (Sigma).

6.1.9 Quantitative Real Time RT-PCR (Q-PCR)

Total RNA was isolated using TRIzol (Invitrogen) according to the manufacture's instruction. First strand cDNA was synthesized using Superscript III Reverse Transcriptase (Invitrogen). Q-PCR was performed using the SYBR® Green PCR Master Mix (Invitrogen). All samples were analyzed in triplicate and

normalized to levels of CD19. Data shown represents an average of the results from two different primer sets: 1, RNF126 Q1-F/R and 2, RNF126 Q2-F/R. Primers for CD19 are called CD19-QPCR-F/R.

6.1.10 Preparation of mammalian cell extracts

Cells were harvested and the pellets washed with 1X PBS. Cells were lysed with RIPA buffer (50 mM Tris (pH 8), 150 mM NaCl, .5% Na Deoxycholate, .1% SDS, 1% NP-40, .5 mM EDTA, 1 mM DTT and protease inhibitor cocktail (Roche)). Buffer used for the 293T ubiquitination assays was also supplemented with the deubiquitinase inhibitor, N-Ethylmaleimide (NEM, 1 mg/mL, Sigma). Lysates were normalized with the detergent compatible DC-Assay (Biorad). When noted, quantification of protein levels was done using the ImageJ software. Values shown represent the fold change of the protein of interest after normalization to the loading control.

6.1.11 HEK 293T ubiquitination assay

HEK 293T cells were plated in a 6-well dish and transfected with 500 ng-1μg of pCDNA3.AID/FLAG-AID, 500 ng -1μg of pCDNA3.HA-RNF126 and 1 μg of pCDNA3.Flag-Ubiquitin, when indicated, using Lipofectamine 2000 (Invitrogen). 36-48hrs later, cells were lysed as described (Supplemental Methods). AID immunoprecipitations were performed using either 2.5-3.5 μL of anti-AID antibody (Cell Signaling) and protein G-Sepharose beads (Invitrogen) or anti-Flag M2 agarose beads (Sigma). Beads were washed at least 3X with lysis buffer and eluted with Laemmli buffer containing 200 mM DTT and boiling.

Samples were loaded on a 12.5% Tris-HCl Criterion gel (BioRad) transferred and blotted with the antibodies indicated.

6.1.12 *In vitro* ubiquitination assay

His-tagged RNF126 was purified as described. His-tagged Ube1 (E1), His-tagged UbcH5b (E2) and ubiquitin were purchased from BioMol, Lysineless Ubiquitin from Boston Biochem and hAID from Enzymax. All components were resuspended in ubiquitination buffer (50mM Tris HCl (pH 7.5), 2.5 mM $MgCl_2$, 1mM DTT) and incubated at 37°C for 30 minutes to 1 hour. An equal volume of Laemmli buffer containing 200 mM DTT was added and the samples boiled prior to gel loading. Concentrations of components used are as follows: Ube1-50 nM, UbcH5b-500 nM, RNF126-300 ng, AID-300 ng, Ubiquitin-1-4 μg, ATP-1-2 mM, in a total volume of 25 μL. Reactions were loaded on a 12.5% Tris-HCl Criterion gel (Biorad) and membranes were blotted with an anti-AID antibody.

6.1.13 HEK 293T ubiquitination assay using alternate E3 ligases and substrates

When alternate ligases and/or substrates were included in this assay, the protocol remained as described; however, amounts of plasmid DNA transfected in the cells varied depending on the expression level of the given construct. 750ng – 1 μg HA-BCA2 and HA-RNF8 and 200 ng of FLAG-PCNA were transfected in these experiments.

6.1.14 RNF8/PCNA *in vitro* ubiquitination assay

RNF8 and PCNA were used as an alternate E3 Ligase/Substrate pair in this assay. 100 ng of GST-RNF8 (Abnova) and 100 ng of GST-PCNA (Abnova)

were incubated as described in Materials and Methods with 500 nM of either the E2, UbcH5B or the E2, UbcH5C (BioMol). All other components and concentrations remained the same.

6.2 Pertaining to all *in vivo* work: CHAPTER 4

6.2.1 Generation of an RNF126 conditional knockout mouse model

ES cells containing the targeted RNF126 allele were purchased from the EUCOMM International Knockout Mouse Consortium. Targeting was done in the JM8A3.N1 ES cell line, which is derived from a C57BL/6N background and contain an agouti coat color.

ES cells were injected into a C57BL/6 blastocyst by the Rockefeller University Gene Targeting Resource Center. Chimera mice of high chimerism (brown/black in coat color) were selected and bred to C57BL/6 wildtype mice. Offspring were screened for agouti (brown) coat color as an indication of whether the targeted allele was transmitted to the offspring. Several rounds of breeding produced a single brown mouse and southern blotting was used to verify the presence of the RNF126 targeted allele.

Mice containing one copy of the targeted allele ($RNF126^{ki/+}$) are functionally heterozygous because the gene trap construct inserted in intron 1 disrupts translation. This gene trap cassette is flanked by Flipase recombinase sites (FRT). $RNF126^{ki/+}$ mice were bred to transgenic mice expressing the Flipase recombinase (Jackson Labs # 005703) to reinstate complete translation of the targeted allele. Mice containing a Flipped version of the targeted allele

(RNF126$^{Fl/+}$) were bred to C57BL/6 mice to cross out the Flipase transgene and then to mb-1 Cre expressing mice (Hobeika et al., 2006) to generate the conditional knockout. Through rounds of breeding, RNF126$^{Fl/Fl}$mb1$^{Cre/+}$ mice were generated and validated at the DNA, RNA and protein level as B cell-specific RNF126 conditional knockout mice. RNF126$^{FL/FL}$ mb1$^{+/+}$ were used as wildtype littermate controls in experiments

6.2.2 Genotyping Primers

Table 6.1 presents the genotyping primers used during the generation of the RNF126 conditional knockout mouse, as well as a brief description of where the primers bind and the expected amplicons.

6.2.3 Southern Blot and PCR Analysis of RNF126 conditional knockout mice

Initial confirmation of the correct integration of the construct into the *rnf126* locus was completed by Southern blot analysis. When hybridized to SspI-digested genomic DNA, this probe could differentiate between wildtype (14.1 kb) and targeted (8.5 kb) alleles.

Genomic DNA isolated from tail biopsies or pelleted splenic B cells were used for PCR genotyping. Tails or B cells were incubated overnight at 56°C in 500 μL Tail Lysis Buffer (100 mM Tris-Cl pH 8.5, 5 mM EDTA, .2% SDS, 200 mM NaCl). Samples were centrifuged at 9K RPM for 8 minutes to pellet any remaining tail material and the supernatant was combined with an equal volume

of isopropanol. Samples were centrifuged for 30 minutes at full speed (13.2K RPM) to pellet the DNA, which was then dried and resuspended in water.

Primers used for genotyping and expected amplicon sizes for each genotyping primer set are presented in 6.2.2. KOD Hot Start Polymerase (EMD Millipore) was used for all PCR reactions. For the primer sets Endo-F/Endo-R and Endo-F/RP-R, PCR conditions were 30 cycles of 95°C for 20sec, 54°C for 10sec, and 70°C for 4sec. All four primers for Cre genotyping were included in one reaction. Conditions for this PCR were 30 cycles of 95°C for 20sec, 57°C for 10sec, 70°C for 5sec. Similarly, all four primers for Flipase genotyping were included in one reaction. Conditions for this PCR were 30 cycles of 95°C for 20sec, 58°C for 10sec, 70°C for 10sec.

6.2.4 B Cell Culture Conditions

B Cells (primary and CH12) were cultured in RPMI-1640 + Glutamax (Gibco Life Technologies) supplemented with 10% FBS (BenchMark), 1X Penn Strep (Life Technologies), 55 μM β-mercaptoethanol (Life Technologies), 1mM Sodium Pyruvate (Life Technologies), 5mM HEPES (Life Technologies), 1X MEM Non Essential Amino Acids (Life Technologies). HEK 293T cells were cultured in DMEM media (Gibco Life Technologies) supplemented with 10% FBS (BenchMark), 1X Penn Strep (Life Technologies) and 2mM L-glutamine (Life Technologies).

6.2.5 Splenic B cell Purification and Activation

Naïve splenic B cells from 8-12 week old mice were purified by CD43 negative selection on a magnetic column (MACS, Miltenyi Biotec). Purified B cells were plated at a concentration of 0.5 X 10^6 cells/mL and stimulated with either (1) 10 ng/mL IL-4 (Sigma) and 1 mg/mL anti-CD40 (ebiosciences), (2) 10 ng/mL IL-4 and 25 μg/mL LPS (Sigma) or (3) 25 μg/mL LPS (Sigma).

6.2.6 *In vitro* Class Switch Recombination Assay

Naïve splenic B cells were purified and stimulated as described above. At stated time points post-stimulation (e.g. 72hr, 96hr), cells were collected and FACS analysis was used to determine the percentage of B cells, which have undergone CSR.

6.2.7 B Cell Proliferation Analysis with CFSE

CFSE (Carboxyfluorescein diacetate succinimidyl ester, Invitrogen, CellTrace) was dissolved in DMSO at a final concentration of 5 mM. B cells were resuspended at 1X10^6/mL in PBS/5% FBS. 2 μL of CFSE was added for every 1 mL of cells to achieve a final CFSE concentration of 10 μM. Cells were labeled at room temperature for 10 min. The tube was covered with aluminum foil to prevent bleaching. The reaction was quenched by adding an equal volume of FBS and placing the tube on ice for 5 min. The tube was filled with PBS/5%FBS and centrifuged at 300xg for 5 min at room temperature. Cells were washed two more times in PBS/5%FBS, resuspended in culturing medium supplemented with CSR

stimuli and plated. Cells not loaded with CFSE were used as a negative control for FACS analysis.

6.2.8 sqPCR of RNF126 levels

Splenic B cells were purified from mice stated in the given experiment and stimulated to undergo CSR *in* vitro. RNA was prepared from 72hr activated B cells using the RNeasy Mini Kit (Qiagen). First strand cDNA was synthesized using Superscript III Reverse Transcriptase (Invitrogen) and semi-quantitative PCR performed to assay levels of RNF126 mRNA. Primers for RPL32 were used as a normalization control. Primers used were as follows: RNF126- F: 5'-acggtacttctgccactgct-3' (binds in exon 1) R: 5'-aactggctgtatccctgtgg-3' (binds in exon 4); RPL32- F: 5'-AAGCGAAACTGGCGGAAAC-3' and R: 5'-TAACCGATGTTGGGCATCAG-3'. PFU Turbo (Agilent) was used for PCR amplification using an annealing temperature of 55°C and 30 cycles.

6.2.9 B Cell Development Assay

Bone marrow cells were flushed from the femurs and tibias of mice indicated in the experiment. Flow cytometry was used to assay the relative proportions of pro-, pre- and immature B cells in the bone marrow. Lymphocytes were gated based on their FSC/SSC profiles and B cells were identified as CD19+/IgM-. B cell subsets were determined by size (FSC) and surface CD43 levels. In addition, spleen cells were prepared and IgM/IgD double positive B cells were assayed by Flow Cytometry. Antibodies used are stated below.

6.2.10 Antibodies Used. Anti-RNF126 (Sigma**,** HPA043050), CD19-PE (Clone ID3, BD Pharmingen 557399), CD19-APC (Clone ID3, BD Pharmingen 550992), FAS-PE (BD Pharmingen 554258), GL7-FITC (BD Pharmingen 553666), GL7-Alexa Fluor 647 (eBiosciences 51-5902-80), IgG1-PE (Clone A85-1, BD Pharmingen 550083), IgG1-APC (BD Pharmingen 550874), CD43-PE (BD Pharmingen 553271), IgM-PE (BD Pharmingen 553409), IgD FITC (BD Pharmingen 553439), IgA-PE (eBiosciences 12-4204-82).

6.2.11 Preparation of mammalian cell extracts

Cells were harvested and the pellets washed with 1X PBS. Cells were lysed with RIPA buffer (50mM Tris (pH 8), 150mM NaCl, .5% Na Deoxycholate, .1% SDS, 1% NP-40, .5mM EDTA, 1mM DTT and protease inhibitor cocktail (Roche)). Lysates were normalized with the detergent compatible DC-Assay (Biorad).

6.2.12 Viral Transduction of CH12 Cells

HEK 293T cells were transfected with 1 μg of hairpin expressing pLKO.1 lentiviral vectors with 750 ng of psPAX2 and 250 ng of pMD2.G packaging vectors with Lipofectamine 2000 reagent (Invitrogen). 24hr post-transfection, media was changed to IMDM (Life Technologies) with 5% FBS (BenchMark). 72hr post-transfection, virus-containing supernatant was collected, filtered through a .45μm filter, and supplemented with polybrene at 8μg/mL concentration. $1X10^6$ CH12 cells were resuspended in viral supernatant and spin-

infected at 850xg for 2hr at 20°C. CH12 cells were removed from the centrifuge and 1mL of culturing medium added.

6.2.13 shRNA Knockdown Experiments

CH12 cells were virally transduced with hairpin expressing vectors against GFP, AID, and RNF126. 48hr post-infection, cells were selected in 1μg/mL of puromycin. Cells were stimulated for CSR with 10 ng/mL IL-4 (Sigma), 1 μg/mL anti-CD40 (eBiosciences) and 1 ng/mL TGFβ. CSR to IgA was analyzed by flow cytometry at 48hr post stimulation**.** Hairpin sequences are as follows: shGFP: GCAAGCTGACCCTGAAGTTCA, shAID: CCGGGCGAGATGCATTTCGTATGTTCTCGAGAACATACGAAATGCATCTCGCTTTTTG,
shRNF126-1: CCGGGCTTTGAAATAAATGGACGTTCTCGAGAACGTCCATTTATTTCAAAGCTTTTTG, shRNF126-2: CCGGGCTCCTCAATCAGTTTGAGAACTCGAGTTCTCAAACTGATTGAGGAGCTTTTTG, shRNF126-3: CCGGCCCAGTGTGTAAAGAAGACTACTCGAGTAGTCTTCTTTACACACTGGGTTTTTG.

6.2.14 Immunization of Mice

Mice were immunized intraperitoneally with 50μg of alum-precipitated $NP_{16/17}$-CGG (Biosearch Technologies). Precipitation was carried out at room-temperature using 50mL of a 1 mg/mL solution of NP-CGG and 200μL of Alum.

6.2.15 Mutation Analysis

Splenic germinal center B cells were FACS-sorted using the surface markers CD19, FAS and GL7 and genomic DNA was prepared. The following regions were amplified with a nested PCR reaction using PFUTurbo (Agilent) and the stated primers: V186.2 Exon (351 bp) (5'-TCTTTACAGTTACTGAGCACACAGGAC-3' and 5'-GGGTCTAGAGGTGTCCCTAGTCCTTCATGACC-3' followed by 5'-CAGTAGCAGGCTTGAGGTCTGGAC-3' and 5'-GGGTCTAGAGGTGTCCCTAGTCCTTCATGACC-3'), JH4 Intron (357 bp) (5'-AGCCTGACATCTGAGGAC-3' and 5'-TAGTGTGGAACATTCCTCAC-3' followed by 5'-CTGACATCTGAGGACTCTGC-3' and 5'-GCTGTCACAGAGGTGGTCCTG-3') and V186.2 Upstream Region (295 bp) (5'-GGCTCTAATGTTACATCCATAGCCTCAAC-3' and 5'-GGGTCTAGAGGTGTCCCTAGTCCTTCATGACC-3' followed by 5'-CAGACAAGATGAGGACTTGGGCTTCAGTATCC-3' and 5'-GTCCAGACCTCAAGCCTGCTACTG-3'). PCR products were blunt cloned into the pSC-B vector provided with the StrataClone Blunt PCR Cloning Kit (Agilent) and colonies sequenced with a T3 primer. Blat was run locally to align sequences to a consensus and trim vector sequence information external to the PCR product. For each region analyzed, all sequences from one mouse were combined and mutation analysis was conducted with the aid of SHMTool (Maccarthy et al., 2009).

6.2.16 NP Elisa

Serum from immunized mice was prepared from peripheral blood collected from the facial vein using Serum Seperator Tubes (BD Biosciences). Serum dilutions (prepared in blocking buffer) were incubated with either NP_3-BSA or NP_{30}-BSA (5μg/mL) coated wells of a flat-bottom EIA/RIA plate (Costar). Serum was incubated at 37°C for 1.5 hours after coated plates had been blocked in blocking buffer at room temperature for 1 hour. Wells were washed 3X with washing buffer after blocking and 5X after incubation with serum. A 1:1000 dilution of secondary antibody was prepared in blocking buffer, loaded in each well and incubated for 1 hour at 37°C. Again, plates were washed 5X with washing buffer. Plates were developed using a 1 mg/mL solution of 4-Nitrophenyl phosphate disodium salt hexahydrate (PNPP, Sigma) resuspended in ELISA buffer. Substrate was incubated for 30 minutes prior to measuring the OD at 405 nm. NP-specific serum IgG1 was detected using an alkaline phosphatase (AP) conjugated secondary antibody (goat anti-mouse IgG1 antibody, Southern Biotech, 1070-04). Total IgG1 levels for the purpose of normalization were determined by coating plates with a Goat anti-mouse IgG antibody (Jackson Labs, 115-006-071) and detection with the same secondary (anti-mouse IgG1) antibody. Buffer conditions are as follows: Blocking Buffer (1% BSA, 1X Borate Buffer), Washing Buffer (.05% Tween 20 in 1X PBS), 1X ELISA Buffer (.1M Glycine (from a stock of 1M Glycine at pH 10.4), 1mM ZnCl2, 1mM MgCl2), 1X Borate Buffer (100 mM Boric Acid, 25 mM Sodium Borate, 75 mM Sodium

Chloride). Serum NP_3 IgG1 and NP_{30} IgG1 levels are normalized to total IgG1 levels. Affinity Maturation is determined as the ratio of NP_3/NP_{30}.

Table 6.1 Primers used to genotype the RNF126 conditional knockout mouse. Primer sequences, descriptions and expected amplicon sizes are presented.

	Primer Name	Sequence	Description	Amplicon
RNF126 Locus	R2-F	TCT ATA GTC GCA GTA GGC GG	Construct Specific Sequence, 5' of Targeted Region	See Figure 4.2
	Endo-F	CTG CCT GCT CTA CTC TTG TC	Binds Intron 1	
	Endo-R	GTA GGA CGT GGA CAG CTA GG	Binds Intron 1	
	RP-R	TGA ACT GAT GGC GAG CTC AGA CC	Construct Specific Sequence, 3' of Targeted Region	
FLIPASE Transgene	oIMR1348	CAC TGA TAT TGT AAG TAG TTT GC	FLIPASE Transgene Detection	725 bp
	oIMR1349	CTA GTG CGA AGT AGT GAT CAG G		
	oIMR7338	CTA GGC CAC AGA ATT GAA AGA TCT	Internal Control-Forward	324 bp
	oIMR7339	GTA GGT GGA AAT TCT AGC ATC ATC C	Internal Control-Reverse	
mb-1 CRE	hCRE-F	CCCTGTGGATGCCACCTC	Mb1 Cre Allele-Forward	450 bp
	hCRE-R	GTCCTGGCATCTGTCAGAG	Mb1 Cre Allele-Reverse	
	oIMR8744	CAAATGTTGCTTGTCTGGTG	Internal Control-Forward	200 bp
	oIMR8745	GTCAGTCGAGTGCACAGTTT	Internal Control-Reverse	

REFERENCES

Abdouni, H., King, J.J., Suliman, M., Quinlan, M., Fifield, H., and Larijani, M. (2013). Zebrafish AID is capable of deaminating methylated deoxycytidines. Nucleic Acids Res, 1-12.

Aida, M., Hamad, N., Stanlie, A., Begum, N.A., and Honjo, T. (2013). Accumulation of the FACT complex, as well as histone H3.3, serves as a target marker for somatic hypermutation. Proceedings of the National Academy of Sciences *110*, 7784–7789.

Alber, F., Dokudovskaya, S., Veenhoff, L.M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B.T., et al. (2007a). Determining the architectures of macromolecular assemblies. Nature *450*, 683–694.

Alber, F., Dokudovskaya, S., Veenhoff, L.M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B.T., et al. (2007b). The molecular architecture of the nuclear pore complex. Nature *450*, 695–701.

Alt, F.W., Zhang, Y., Meng, F.-L., Guo, C., and Schwer, B. (2013). Mechanisms of programmed DNA lesions and genomic instability in the immune system. Cell *152*, 417–429.

Amemiya, Y., Azmi, P., and Seth, A. (2008). Autoubiquitination of BCA2 RING E3 Ligase Regulates Its Own Stability and Affects Cell Migration. Molecular Cancer Research *6*, 1385–1396.

Aoufouchi, S., Faili, A., Zober, C., D'Orlando, O., Weller, S., Weill, J.-C., and Reynaud, C.-A. (2008). Proteasomal degradation restricts the nuclear lifespan of AID. J. Exp. Med. *205*, 1357–1368.

Arakawa, H., Moldovan, G.-L., Saribasak, H., Saribasak, N.N., Jentsch, S., and

Buerstedde, J.-M. (2006). A role for PCNA ubiquitination in immunoglobulin hypermutation. PLoS Biol. *4*, e366.

Bachl, J., Carlson, C., Gray-Schopfer, V., Dessing, M., and Olsson, C. (2001). Increased transcription levels induce higher mutation rates in a hypermutating cell line. J. Immunol. *166*, 5051–5057.

Bachl, J., Ertongur, I., and Jungnickel, B. (2006). Involvement of Rad18 in somatic hypermutation. Proc. Natl. Acad. Sci. U.S.a. *103*, 12081–12086.

Bacopulos, S., Amemiya, Y., Yang, W., Zubovits, J., Burger, A., Yaffe, M., and Seth, A.K. (2012). Effects of partner proteins on BCA2 RING ligase activity. BMC Cancer *12*, 63.

Barbaric, I., Miller, G., and Dear, T.N. (2007). Appearances can be deceiving: phenotypes of knockout mice. Brief Funct Genomic Proteomic *6*, 91–103.

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.

Basu, U., Chaudhuri, J., Alpert, C., Dutt, S., Ranganath, S., Li, G., Schrum, J.P., Manis, J.P., and Alt, F.W. (2005). The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. Nature *438*, 508–511.

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.

Basu, U., Wang, Y., and Alt, F.W. (2008). Evolution of phosphorylation-dependent regulation of activation-induced cytidine deaminase. Molecular Cell *32*, 285–291.

Bates, P.A., and DeLuca, N.A. (1998). The polyserine tract of herpes simplex virus ICP4 is required for normal viral gene expression and growth in murine trigeminal ganglia. J. Virol. *72*, 7115–7124.

Begum, N.A., Stanlie, A., Doi, T., Sasaki, Y., Jin, H.W., Kim, Y.S., Nagaoka, H., and Honjo, T. (2009). Further evidence for involvement of a noncanonical function of uracil DNA glycosylase in class switch recombination. Proceedings of the National Academy of Sciences *106*, 2752–2757.

Besmer, E., Market, E., and Papavasiliou, F.N. (2006). The transcription elongation complex directs activation-induced cytidine deaminase-mediated DNA deamination. Molecular and Cellular Biology *26*, 4378–4385.

Betz, B.C., Jordan-Williams, K.L., Wang, C., Kang, S.G., Liao, J., Logan, M.R., Kim, C.H., and Taparowsky, E.J. (2010). Batf coordinates multiple aspects of B and T cell function required for normal antibody responses. Journal of Experimental Medicine *207*, 933–942.

Bhutani, N., Decker, M.N., Brady, J.J., Bussat, R.T., Burns, D.M., Corbel, S.Y., and Blau, H.M. (2012). A critical role for AID in the initiation of reprogramming to induced pluripotent stem cells. Faseb J *27* (3), 1107-1113.

Bienko, M., Green, C.M., Sabbioneda, S., Crosetto, N., Matic, I., Hibbert, R.G., Begovic, T., Niimi, A., Mann, M., Lehmann, A.R., et al. (2010). Regulation of translesion synthesis DNA polymerase eta by monoubiquitination. Molecular Cell *37*, 396–407.

Blagodatski, A., Batrak, V., Schmidl, S., Schoetz, U., Caldwell, R.B., Arakawa, H., and Buerstedde, J.-M. (2009). A cis-acting diversification activator both necessary and sufficient for AID-mediated hypermutation. PLoS Genet. *5*, e1000332.

Boboila, C., Oksenych, V., Gostissa, M., Wang, J.H., Zha, S., Zhang, Y., Chai, H., Lee, C.-S., Jankovic, M., Saez, L.-M.A., et al. (2012). Robust chromosomal DNA repair via alternative end-joining in the absence of X-ray repair cross-complementing protein 1 (XRCC1) Proceedings of the National Academy of Sciences U.S.A *109* (7), 2473.

Boboila, C., Yan, C., Wesemann, D.R., Jankovic, M., Wang, J.H., Manis, J., Nussenzweig, A., Nussenzweig, M., and Alt, F.W. (2010). Alternative end-joining catalyzes class switch recombination in the absence of both Ku70 and DNA ligase 4 Journal of Experimental Medicine *207*, (2), 417-427.

Boehm, T. (2011). Design principles of adaptive immune systems. Nat Rev Immunol *11*, 307–317.

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. Journal of Experimental Medicine *207*, 855–865.

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. Molecular Cell *42*, 319–329.

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.

Brar, S.S., Watson, M., and Diaz, M. (2004). Activation-induced cytosine deaminase (AID) is actively exported out of the nucleus but retained by the induction of DNA breaks. J. Biol. Chem. *279*, 26395–26401.

Braun, T., Rudnicki, M.A., Arnold, H.H., and Jaenisch, R. (1992). Targeted inactivation of the muscle regulatory gene Myf-5 results in abnormal rib development and perinatal death. Cell *71*, 369–382.

Buckley, K.M., Munshaw, S., Kepler, T.B., and Smith, L.C. (2008a). The 185/333 gene family is a rapidly diversifying host-defense gene cluster in the purple sea urchin Strongylocentrotus purpuratus. J. Mol. Biol. *379*, 912–928.

Buckley, K.M., Terwilliger, D.P., and Smith, L.C. (2008b). Sequence variations in 185/333 messages from the purple sea urchin suggest posttranscriptional modifications to increase immune diversity. J. Immunol. *181*, 8585–8594.

Burger, A., Amemiya, Y., Kitching, R., and Seth, A.K. (2006). Novel RING E3 ubiquitin ligases in breast cancer. Neoplasia *8*, 689–695.

Burnet, F.M. (1976). A modification of Jerne's theory of antibody production using the concept of clonal selection. CA Cancer J Clin *26*, 119–121.

Cabantous, S., and Waldo, G.S. (2006). In vivo and in vitro protein solubility assays using split GFP. Nat. Methods *3*, 845–854.

Cadwell, K., and Coscoy, L. (2005). Ubiquitination on nonlysine residues by a viral E3 ubiquitin ligase. Science *309*, 127–130.

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.

Cheng, H.-L., Vuong, B.Q., Basu, U., Franklin, A., Schwer, B., Astarita, J., Phan, R.T., Datta, A., Manis, J., Alt, F.W., et al. (2009). Integrity of the AID serine-38 phosphorylation site is critical for class switch recombination and somatic hypermutation in mice. Proceedings of the National Academy of Sciences *106*, 2717–2722.

Conticello, S.G., Ganesh, K., Xue, K., Lu, M., Rada, C., and Neuberger, M.S. (2008). Interaction between antibody-diversification enzyme AID and spliceosome-associated factor CTNNBL1. Molecular Cell *31*, 474–484.

Crouch, E.E., Li, Z., Takizawa, M., Fichtner-Feigl, S., Gourzi, P., Montaño, C., Feigenbaum, L., Wilson, P., Janz, S., Papavasiliou, F.N., et al. (2007). Regulation of AID expression in the immune response. J. Exp. Med. *204*, 1145–1156.

Daniels, G.A., and Lieber, M.R. (1995). RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. Nucleic Acids Res. *23*, 5006–5011.

Danilova, N. (2012). The evolution of adaptive immunity. Adv. Exp. Med. Biol. *738*, 218–235.

de Yébenes, V.G., Belver, L., Pisano, D.G., González, S., Villasante, A., Croce, C., He, L., and Ramiro, A.R. (2008). miR-181b negatively regulates activation-induced cytidine deaminase in B cells. J. Exp. Med. *205*, 2199–2206.

Dedeoglu, F., Horwitz, B., Chaudhuri, J., Alt, F.W., and Geha, R.S. (2004). Induction of activation-induced cytidine deaminase gene expression by IL-4 and CD40 ligation is dependent on STAT6 and NFkappaB. Int. Immunol. *16*, 395–404.

Delbos, F., Aoufouchi, S., Faili, A., Weill, J.-C., and Reynaud, C.-A. (2007). DNA polymerase eta is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse. J. Exp. Med. *204*, 17–23.

Delbos, F., De Smet, A., Faili, A., Aoufouchi, S., Weill, J.-C., and Reynaud, C.-A. (2005). Contribution of DNA polymerase eta to immunoglobulin gene hypermutation in the mouse. J. Exp. Med. *201*, 1191–1196.

Delker, R.K., Fugmann, S.D., and Papavasiliou, F.N. (2009). A coming-of-age story: activation-induced cytidine deaminase turns 10. Nat. Immunol. *10*, 1147–1153.

Delker, R.K., Zhou, Y., Strikoudis, A., Stebbins, C.E., and Papavasiliou, F.N. (2012). Solubility-based genetic screen identifies RING finger protein 126 as an E3 ligase for activation-induced cytidine deaminase. Proceedings of the National Academy of Sciences *110* (3), 1029-1034.

Di Virgilio, M., Callen, E., Yamane, A., Zhang, W., Jankovic, M., Gitlin, A.D., Feldhahn, N., Resch, W., Oliveira, T.Y., Chait, B.T., et al. (2013). Rif1 Prevents Resection of DNA Breaks and Promotes Immunoglobulin Class Switching. Science *339* (6120), 711-715.

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.

Dinkelmann, M., Spehalski, E., Stoneham, T., Buis, J., Wu, Y., Sekiguchi, J.M., and Ferguson, D.O. (2009). Multiple functions of MRN in end-joining pathways during isotype class switching. Nat. Struct. Mol. Biol. *16*, 808–813.

Doil, C., Mailand, N., Bekker-Jensen, S., Menard, P., Larsen, D.H., Pepperkok, R., Ellenberg, J., Panier, S., Durocher, D., Bartek, J., et al. (2009). RNF168 binds and amplifies ubiquitin conjugates on damaged chromosomes to allow accumulation of repair proteins. Cell *136*, 435–446.

Dorsett, Y., McBride, K.M., Jankovic, M., Gazumyan, A., Thai, T.-H., Robbiani, D.F., Di Virgilio, M., Reina San-Martin, B., Heidkamp, G., Schwickert, T.A., et al. (2008). MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation. Immunity *28*, 630–638.

Du, L., Dunn-Walters, D.K., Chrzanowska, K.H., Stankovic, T., Kotnis, A., Li, X., Lu, J., Eggertsen, G., Brittain, C., Popov, S.W., et al. (2008). A regulatory role for NBS1 in strand-specific mutagenesis during somatic hypermutation. PLoS ONE *3*, e2482.

Duke, J.L., Liu, M., Yaari, G., Khalil, A.M., Tomayko, M.M., Shlomchik, M.J., Schatz, D.G., and Kleinstein, S.H. (2013). Multiple Transcription Factor Binding Sites Predict AID Targeting in Non-Ig Genes. J. Immunol *190* (8), 3878-3888.

Dunnick, W., Hertz, G.Z., Scappino, L., and Gritzmacher, C. (1993). DNA sequences at immunoglobulin switch region recombination sites. Nucleic Acids Res. *21*, 365–372.

Durandy, A., Revy, P., and Fischer, A. (2004). Human models of inherited immunoglobulin class switch recombination and somatic hypermutation defects (hyper-IgM syndromes). Adv. Immunol. *82*, 295–330.

Ehrenstein, M.R., and Neuberger, M.S. (1999). Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation. Embo J. *18*, 3484–3490.

Ehrenstein, M.R., Rada, C., Jones, A.M., Milstein, C., and Neuberger, M.S. (2001). Switch junction sequences in PMS2-deficient mice reveal a microhomology-mediated mechanism of Ig class switch recombination. Proc. Natl. Acad. Sci. U.S.a. *98*, 14553–14558.

Faili, A., Aoufouchi, S., Guéranger, Q., Zober, C., Léon, A., Bertocci, B., Weill, J.-C., and Reynaud, C.-A. (2002). AID-dependent somatic hypermutation occurs as a DNA single-strand event in the BL2 cell line. Nat. Immunol. *3*, 815–821.

Faili, A., Aoufouchi, S., Weller, S., Vuillier, F., Stary, A., Sarasin, A., Reynaud, C.-A., and Weill, J.-C. (2004). DNA polymerase eta is involved in hypermutation occurring during immunoglobulin class switch recombination. J. Exp. Med. *199*, 265–270.

Feil, R., Wagner, J., Metzger, D., and Chambon, P. (1997). Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. Biochem. Biophys. Res. Commun. *237*, 752–757.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245–246.

Fineran, P.C., and Charpentier, E. (2012). Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. Virology *434*, 202–209.

Flajnik, M.F., and Kasahara, M. (2010). Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat. Rev. Genet. *11*, 47–59.

Fritz, E.L., and Papavasiliou, F.N. (2010). Cytidine deaminases: AIDing DNA demethylation? Genes Dev. *24*, 2107–2114.

Fritz, E.L., Rosenberg, B.R., Lay, K., Mihailović, A., Tuschl, T., and Papavasiliou, F.N. (2013). A comprehensive analysis of the effects of the deaminase AID on the transcriptome and methylome of activated B cells. Nat. Immunol *14*, 749-755.

Fugmann, S.D., Messier, C., Novack, L.A., Cameron, R.A., and Rast, J.P. (2006). An ancient evolutionary origin of the Rag1/2 gene locus. Proc. Natl. Acad. Sci. U.S.a. *103*, 3728–3733.

Fukita, Y., Jacobs, H., and Rajewsky, K. (1998). Somatic hypermutation in the heavy chain locus correlates with transcription. Immunity *9*, 105–114.

Gavin, A.-C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dümpelfeld, B., et al. (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature *440*, 631–636.

Gazumyan, A., Timachova, K., Yuen, G., Siden, E., Di Virgilio, M., Woo, E.M., Chait, B.T., Reina San-Martin, B., Nussenzweig, M.C., and McBride, K.M. (2011). Amino-terminal phosphorylation of activation-induced cytidine deaminase suppresses c-myc/IgH translocation. Molecular and Cellular Biology *31*, 442–449.

Ge, H., and Roeder, R.G. (1994). Purification, cloning, and characterization of a human coactivator, PC4, that mediates transcriptional activation of class II genes. Cell *78*, 513–523.

Ge, H., Zhao, Y., Chait, B.T., and Roeder, R.G. (1994). Phosphorylation negatively regulates the function of coactivator PC4. Proc. Natl. Acad. Sci. U.S.a. *91*, 12691–12695.

Geisberger, R., Rada, C., and Neuberger, M.S. (2009). The stability of AID and its function in class-switching are critically sensitive to the identity of its nuclear-export sequence. Proceedings of the National Academy of Sciences *106*, 6736–6741.

Ghavidel, A., Cagney, G., and Emili, A. (2005). A skeleton of the human protein interactome. Cell *122*, 830–832.

Gonda, H., Sugai, M., Nambu, Y., Katakai, T., Agata, Y., Mori, K.J., Yokota, Y., and Shimizu, A. (2003). The balance between Pax5 and Id2 activities is the key to AID gene expression. J. Exp. Med. *198*, 1427–1437.

Gourzi, P., Leonova, T., and Papavasiliou, F.N. (2006). A Role for Activation-Induced Cytidine Deaminase in the Host Response against a Transforming Retrovirus. Immunity *24*, 779–786.

Gu, X. (2003). Evolution of duplicate genes versus genetic robustness against null mutations. Trends Genet. *19*, 354–356.

Gu, Z., Steinmetz, L.M., Gu, X., Scharfe, C., Davis, R.W., and Li, W.-H. (2003). Role of duplicate genes in genetic robustness against null mutations. Nature *421*, 63–66.

Hamilton, C.E., Papavasiliou, F.N., and Rosenberg, B.R. (2010). Diverse functions for DNA and RNA editing in the immune system. RNA Biol *7*, 220–228.

Han, L., and Yu, K. (2008). Altered kinetics of nonhomologous end joining and class switch recombination in ligase IV–deficient B cells. J. Exp. Med *205* (12), 2745-2753.

Häsler, J., Rada, C., and Neuberger, M.S. (2011). Cytoplasmic activation-induced cytidine deaminase (AID) exists in stoichiometric complex with translation elongation factor 1α (eEF1A). Proceedings of the National Academy of Sciences *108*, 18366–18371.

He, B., Qiao, X., and Cerutti, A. (2004). CpG DNA induces IgG class switch DNA recombination by activating human B cells through an innate pathway that requires TLR9 and cooperates with IL-10. The Journal of Immunology *173* (7), 4479-4491.

Hein, K., Lorenz, M.G., Siebenkotten, G., Petry, K., Christine, R., and Radbruch, A. (1998). Processing of switch transcripts is required for targeting of antibody class switch recombination. J. Exp. Med. *188*, 2369–2374.

Hibbert, R.G., Huang, A., Boelens, R., and Sixma, T.K. (2011). E3 ligase Rad18 promotes monoubiquitination rather than ubiquitin chain formation by E2 enzyme Rad6. Proceedings of the National Academy of Sciences *108*, 5590–5595.

Hobeika, E., Thiemann, S., Storch, B., Jumaa, H., Nielsen, P.J., Pelanda, R., and Reth, M. (2006). Testing gene function early in the B cell lineage in mb1-cre mice. Proc. Natl. Acad. Sci. U.S.a. *103*, 13789–13794.

Hoege, C., Pfander, B., Moldovan, G.-L., Pyrowolakis, G., and Jentsch, S. (2002). RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. Nature *419*, 135–141.

Hsu, Y., Jubelin, G., Taieb, F., Nougayrède, J.-P., Oswald, E., and Stebbins, C.E. (2008). Structure of the cyclomodulin Cif from pathogenic Escherichia coli. J. Mol. Biol. *384*, 465–477.

Hu, Y., Ericsson, I., Torseth, K., Methot, S.P., Sundheim, O., Liabakk, N.B., Slupphaug, G., Di Noia, J.M., Krokan, H.E., and Kavli, B. (2012). A Combined Nuclear and Nucleolar Localization Motif in Activation-Induced Cytidine Deaminase (AID) Controls Immunoglobulin Class Switching. J. Mol. Biol. *425* (2), 424-443.

Hunter, T. (2007). The age of crosstalk: phosphorylation, ubiquitination, and beyond. Molecular Cell *28*, 730–738.

Ise, W., Kohyama, M., Schraml, B.U., and Zhang, T. (2011). The transcription factor BATF controls the global regulators of class-switch recombination in both B cells and T cells. Nature *6*, 536-543.

Ito, S., Nagaoka, H., Shinkura, R., Begum, N., Muramatsu, M., Nakata, M., and Honjo, T. (2004). Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. Proc. Natl. Acad. Sci. U.S.A. *101*, 1975–1980.

Iwasato, T., Shimizu, A., Honjo, T., and Yamagishi, H. (1990). Circular DNA is excised by immunoglobulin class switch recombination. Cell *62*, 143–149.

Jackson, S.P., and Durocher, D. (2013). Regulation of DNA Damage Responses by Ubiquitin and SUMO. Molecular Cell *49*, 795–807.

Jansen, J.G., Langerak, P., Tsaalbi-Shtylik, A., van den Berk, P., Jacobs, H., and de Wind, N. (2006). Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. J. Exp. Med. *203*, 319–323.

Jeevan-Raj, B.P., Robert, I., Heyer, V., Page, A., Wang, J.H., Cammas, F., Alt, F.W., Losson, R., and Reina San-Martin, B. (2011). Epigenetic tethering of AID to the donor switch region during immunoglobulin class switch recombination. J. Exp. Med. *208*, 1649–1660.

Johnston, R.J., and Desplan, C. (2008). Stochastic neuronal cell fate choices. Curr. Opin. Neurobiol. *18*, 20–27.

Johnston, R.J., and Desplan, C. (2010). Stochastic mechanisms of cell fate specification that yield random or robust outcomes. Annu. Rev. Cell Dev. Biol. *26*, 689–719.

Kaelin, W.G. (2012). Use and Abuse of RNAi to Study Mammalian Gene Function. Science *337*, 421–422.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., et al. (2003). Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature *421*, 231–237.

Klein, I.A., Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D.F., et al. (2011). Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. Cell *147*, 95–106.

Klein, U., and Dalla-Favera, R. (2008). Germinal centres: role in B-cell physiology and malignancy. Nat Rev Immunol *8*, 22–33.

Kobayashi, M., Aida, M., Nagaoka, H., Begum, N.A., Kitawaki, Y., Nakata, M., Stanlie, A., Doi, T., Kato, L., Okazaki, I.-M., et al. (2009). AID-induced decrease in topoisomerase 1 induces DNA structural alteration and DNA cleavage for class switch recombination. Proceedings of the National Academy of Sciences *106*, 22375–22380.

Kodgire, P., Mukkawar, P., Ratnam, S., Martin, T.E., and Storb, U. (2013). Changes in RNA polymerase II progression influence somatic hypermutation of Ig-related genes by AID. Journal of Experimental Medicine *210* (7), 1481-1492.

Kohli, R.M., Abrams, S.R., Gajula, K.S., Maul, R.W., Gearhart, P.J., and Stivers, J.T. (2009). A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase. J. Biol. Chem. *284*, 22898–22904.

Kothapalli, N., Norton, D.D., and Fugmann, S.D. (2008). Cutting edge: a cis-acting DNA element targets AID-mediated sequence diversification to the chicken Ig light chain gene locus. J. Immunol. *180*, 2019–2023.

Kretzschmar, M., Kaiser, K., Lottspeich, F., and Meisterernst, M. (1994). A novel mediator of class II gene transcription with homology to viral immediate-early transcriptional regulators. Cell *78*, 525–534.

Krijger, P.H.L., Langerak, P., van den Berk, P.C.M., and Jacobs, H. (2009). Dependence of nucleotide substitutions on Ung2, Msh2, and PCNA-Ub during somatic hypermutation. J. Exp. Med. *206*, 2603–2611.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., et al. (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature *440*, 637–643.

Kumar, R., Dimenna, L., Schrode, N., Liu, T.-C., Franck, P., Muñoz-Descalzo, S., Hadjantonakis, A.-K., Zarrin, A.A., Chaudhuri, J., Elemento, O., et al. (2013). AID stabilizes stem-cell phenotype by removing epigenetic memory of pluripotency genes. Nature.

Kuwahara, K., Yoshida, M., Kondo, E., Sakata, A., Watanabe, Y., Abe, E., Kouno, Y., Tomiyasu, S., Fujimura, S., Tokuhisa, T., et al. (2000). A novel nuclear phosphoprotein, GANP, is up-regulated in centrocytes of the germinal center and associated with MCM3, a protein essential for DNA replication. Blood *95*, 2321–2328.

Kuwahara, K., Fujimura, S., Takahashi, Y., Nakagata, N., Takemori, T., Aizawa, S., and Sakaguchi, N. (2004). Germinal center-associated nuclear protein contributes to affinity maturation of B cell antigen receptor in T cell-dependent responses. Proc. Natl. Acad. Sci. U.S.a. *101*, 1010–1015.

Langerak, P., Krijger, P.H.L., Heideman, M.R., van den Berk, P.C.M., and Jacobs, H. (2009). Somatic hypermutation of immunoglobulin genes: lessons from proliferating cell nuclear antigenK164R mutant mice. Philos. Trans. R. Soc. Lond., B, Biol. Sci. *364*, 621–629.

Larson, E.D., Cummings, W.J., Bednarski, D.W., and Maizels, N. (2005). MRE11/RAD50 cleaves DNA in the AID/UNG-dependent pathway of immunoglobulin gene diversification. Molecular Cell *20*, 367–375.

Lazzaro, B.P., Sceurman, B.K., and Clark, A.G. (2004). Genetic basis of natural variation in D. melanogaster antibacterial immunity. Science *303*, 1873–1876.

Lebecque, S.G., and Gearhart, P.J. (1990). Boundaries of somatic mutation in rearranged immunoglobulin genes: 5“ boundary is near the promoter, and 3” boundary is approximately 1 kb from V(D)J gene. J. Exp. Med. *172*, 1717–1727.

LEDERBERG, J. (1959). Genes and antibodies. Science *129*, 1649–1653.

Lee, C.G., Kinoshita, K., Arudchandran, A., Cerritelli, S.M., Crouch, R.J., and Honjo, T. (2001). Quantitative regulation of class switch recombination by switch region transcription. J. Exp. Med. *194*, 365–374.

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.

Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. Nature *445*, 168–176.

Leslie, A.G., Moody, P.C., and Shaw, W.V. (1988). Structure of chloramphenicol acetyltransferase at 1.75-A resolution. Proc. Natl. Acad. Sci. U.S.A. *85*, 4133–4137.

Li, G., Borjeson, T., Boboila, C., and Alt, F.W. (2008). DNA-PKcs and Artemis function in the end-joining phase of immunoglobulin heavy chain class switch recombination. The Journal of Experimental Medicine *205* (3), 557-64.

Li, M., Brooks, C.L., Wu-Baer, F., Chen, D., Baer, R., and Gu, W. (2003). Mono- versus polyubiquitination: differential control of p53 fate by Mdm2. Science *302*, 1972–1975.

Li, S.C., Rothman, P.B., Zhang, J., Chan, C., Hirsh, D., and Alt, F.W. (1994). Expression of I mu-C gamma hybrid germline transcripts subsequent to immunoglobulin heavy chain class switching. Int. Immunol. *6*, 491–497.

Li, S., Zhao, Y., and Wang, J.-Y. (2012). Analysis of Ig gene hypermutation in Ung(-/-)Polh(-/-) mice suggests that UNG and A:T mutagenesis pathway target different U:G lesions. Mol. Immunol. *53*, 214–217.

Li, Z., Zhao, C., Iglesias-Ussel, M.D., Polonskaya, Z., Zhuang, M., Yang, G., Luo, Z., Edelmann, W., and Scharff, M.D. (2006). The mismatch repair protein Msh6 influences the in vivo AID targeting to the Ig locus. Immunity *24*, 393–403.

Liang, G., Kitamura, K., Wang, Z., Liu, G., Chowdhury, S., Fu, W., Koura, M., Wakae, K., Honjo, T., and Muramatsu, M. (2013). RNA editing of hepatitis B virus transcripts by activation-induced cytidine deaminase. Proceedings of the National Academy of Sciences *110*, 2246–2251.

Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B., and Ohgi, K.A. (2009). Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell *139* (6), 1069-1083.

Litman, G.W., Cannon, J.P., and Dishaw, L.J. (2005a). Reconstructing immune phylogeny: new perspectives. Nat Rev Immunol *5*, 866–879.

Litman, G.W., Cannon, J.P., and Rast, J.P. (2005b). New insights into alternative mechanisms of immune receptor diversification. Adv. Immunol. *87*, 209–236.

Litman, G.W., Rast, J.P., and Fugmann, S.D. (2010). The origins of vertebrate adaptive immunity. Nat Rev Immunol *10*, 543–553.

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.

Longerich, S., Basu, U., Alt, F., and Storb, U. (2006). AID in somatic hypermutation and class switch recombination. Curr. Opin. Immunol. *18*, 164–174.

Longerich, S., Tanaka, A., Bozek, G., Nicolae, D., and Storb, U. (2005). The very 5' end and the constant region of Ig genes are spared from somatic mutation because AID does not access these regions. J. Exp. Med. *202*, 1443–1454.

Lorenz, M., Jung, S., and Radbruch, A. (1995). Switch transcripts in immunoglobulin class switching. Science *267*, 1825–1828.

Lowder, M.A., Appelbaum, J.S., Hobert, E.M., and Schepartz, A. (2011). Visualizing protein partnerships in living cells and organisms. Curr Opin Chem Biol *15*, 781–788.

Lumsden, J.M., McCarty, T., Petiniot, L.K., Shen, R., Barlow, C., Wynn, T.A., Morse, H.C., Gearhart, P.J., Wynshaw-Boris, A., Max, E.E., et al. (2004). Immunoglobulin class switch recombination is impaired in Atm-deficient mice. J. Exp. Med. *200*, 1111–1121.

Maccarthy, T., Roa, S., Scharff, M.D., and Bergman, A. (2009). SHMTool: a webserver for comparative analysis of somatic hypermutation datasets. DNA Repair (Amst.) *8*, 137–141.

MacDuff, D.A., Neuberger, M.S., and Harris, R.S. (2006). MDM2 can interact with the C-terminus of AID but it is inessential for antibody diversification in DT40 B cells. Mol. Immunol. *43*, 1099–1108.

Maeda, K., Singh, S.K., Eda, K., Kitabatake, M., Pham, P., Goodman, M.F., and Sakaguchi, N. (2010). GANP-mediated recruitment of activation-induced cytidine deaminase to cell nuclei and to immunoglobulin variable region DNA. J. Biol. Chem. *285*, 23945–23953.

Mailand, N., Bekker-Jensen, S., Faustrup, H., Melander, F., Bartek, J., Lukas, C., and Lukas, J. (2007). RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins. Cell *131*, 887–900.

Market, E., and Papavasiliou, F.N. (2003). V(D)J recombination and the evolution of the adaptive immune system. PLoS Biol. *1*, E16.

Martin, A., and Scharff, M.D. (2002). Somatic hypermutation of the AID transgene in B and non-B cells. Proc. Natl. Acad. Sci. U.S.a. *99*, 12304–12308.

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.

Masani, S., Han, L., and Yu, K. (2013). Apurinic/apyrimidinic endonuclease 1 is the essential nuclease during immunoglobulin class switch recombination. Molecular and Cellular Biology *33*, 1468–1473.

Matheson, L.S., and Corcoran, A.E. (2012). Local and global epigenetic regulation of V(D)J recombination. Curr. Top. Microbiol. Immunol. *356*, 65–89.

Matthews, A.G.W., Kuo, A.J., Ramón-Maiques, S., Han, S., Champagne, K.S., Ivanov, D., Gallardo, M., Carney, D., Cheung, P., Ciccone, D.N., et al. (2007). RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. Nature *450*, 1106–1110.

Mattiroli, F., Vissers, J.H.A., van Dijk, W.J., Ikpa, P., Citterio, E., Vermeulen, W., Marteijn, J.A., and Sixma, T.K. (2012). RNF168 Ubiquitinates K13-15 on H2A/H2AX to Drive DNA Damage Signaling. Cell *150*, 1182–1195.

Maul, R.W., Saribasak, H., Martomo, S.A., McClure, R.L., Yang, W., Vaisman, A., Gramlich, H.S., Schatz, D.G., Woodgate, R., Wilson, D.M., et al. (2011). Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions. Nat. Immunol. *12*, 70–76.

Mayorov, V.I., Rogozin, I.B., Adkison, L.R., and Gearhart, P.J. (2005a). DNA polymerase eta contributes to strand bias of mutations of A versus T in immunoglobulin genes. J. Immunol. *174*, 7781–7786.

Mayorov, V.I., Rogozin, I.B., Adkison, L.R., Frahm, C., Kunkel, T.A., and Pavlov, Y.I. (2005b). Expression of human AID in yeast induces mutations in context similar to the context of somatic hypermutation at G-C pairs in immunoglobulin genes. BMC Immunol. *6*, 10.

McBride, K.M., Barreto, V., Ramiro, A.R., Stavropoulos, P., and Nussenzweig, M.C. (2004). Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. J. Exp. Med. *199*, 1235–1244.

McBride, K.M., Gazumyan, A., Woo, E.M., Barreto, V.M., Robbiani, D.F., Chait, B.T., and Nussenzweig, M.C. (2006). Regulation of hypermutation by activation-induced cytidine deaminase phosphorylation. Proc. Natl. Acad. Sci. U.S.a. *103*, 8798–8803.

McBride, K.M., Gazumyan, A., Woo, E.M., Schwickert, T.A., Chait, B.T., and Nussenzweig, M.C. (2008). Regulation of class switch recombination and somatic mutation by AID phosphorylation. J. Exp. Med. *205*, 2585–2594.

Medzhitov, R., and Janeway, C. (2000). Innate immunity. N. Engl. J. Med. *343*, 338–344.

Mehta, A., Kinter, M.T., Sherman, N.E., and Driscoll, D.M. (2000). Molecular cloning of apobec-1 complementation factor, a novel RNA-binding protein involved in the editing of apolipoprotein B mRNA. Molecular and Cellular Biology *20*, 1846–1854.

Miau, L.H., Chang, C.J., Tsai, W.H., and Lee, S.C. (1997). Identification and characterization of a nucleolar phosphoprotein, Nopp140, as a transcription factor. Molecular and Cellular Biology *17*, 230–239.

Milstein, C., Neuberger, M.S., and Staden, R. (1998). Both DNA strands of antibody genes are hypermutation targets. Proc. Natl. Acad. Sci. U.S.a. *95*, 8791–8794.

Min, I.M., Schrader, C.E., Vardo, J., Luby, T.M., D'Avirro, N., Stavnezer, J., and Selsing, E. (2003). The Smu tandem repeat region is critical for Ig isotype switching in the absence of Msh2. Immunity *19*, 515–524.

Morgan, H.D., Dean, W., Coker, H.A., and Reik, W. (2004). Aid deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues-implications for epigenetic reprogramming. Journal of Biological Chemistry *279* (50), 52353-60.

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.

Nakamura, M., Kondo, S., Sugai, M., Nazarea, M., Imamura, S., and Honjo, T. (1996). High frequency class switching of an IgM+ B lymphoma clone CH12F3 to IgA+ cells. Int. Immunol. *8*, 193–201.

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.

Neuberger, M.S., Ehrenstein, M.R., Klix, N., Jolly, C.J., Yélamos, J., Rada, C., and Milstein, C. (1998). Monitoring and interpreting the intrinsic features of somatic hypermutation. Immunol. Rev. *162*, 107–116.

Neuberger, M.S. (2008). Antibody diversification by somatic mutation: from Burnet onwards. Immunol. Cell Biol. *86*, 124–132.

Niida, H., Katsuno, Y., Sengoku, M., Shimada, M., Yukawa, M., Ikura, M., Ikura, T., Kohno, K., Shima, H., Suzuki, H., et al. (2010). Essential role of Tip60-dependent recruitment of ribonucleotide reductase at DNA damage sites in DNA repair during G1 phase. Genes Dev. *24*, 333–338.

Nishana, M., and Raghavan, S.C. (2012). Role of recombination activating genes in the generation of antigen receptor diversity and beyond. Immunology *137*, 271–281.

Nonaka, T., Doi, T., Toyoshima, T., Muramatsu, M., Honjo, T., and Kinoshita, K. (2009). Carboxy-terminal domain of AID required for its mRNA complex formation in vivo. Proceedings of the National Academy of Sciences *106*, 2747–2751.

Nowak, U., Matthews, A.J., Zheng, S., and Chaudhuri, J. (2011). The splicing regulator PTBP2 interacts with the cytidine deaminase AID and promotes binding of AID to switch-region DNA. Nat. Immunol. *12*, 160–166.

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

Okazaki, I.-M., Okawa, K., Kobayashi, M., Yoshikawa, K., Kawamoto, S., Nagaoka, H., Shinkura, R., Kitawaki, Y., Taniguchi, H., Natsume, T., et al. (2011). Histone chaperone Spt6 is required for class switch recombination but not somatic hypermutation. Proceedings of the National Academy of Sciences *108*, 7920–7925.

Oliveira, T.Y., Resch, W., Jankovic, M., Casellas, R., Nussenzweig, M.C., and Klein, I.A. (2012). Translocation capture sequencing: a method for high throughput mapping of chromosomal rearrangements. J. Immunol. Methods *375*, 176–181.

Orthwein, A., Patenaude, A.-M., Affar, E.B., Lamarre, A., Young, J.C., and Di Noia, J.M. (2010). Regulation of activation-induced deaminase stability and antibody gene diversification by Hsp90. J. Exp. Med. *207*, 2751–2765.

Park, S.-R., Zan, H., Pal, Z., Zhang, J., Al-Qahtani, A., Pone, E.J., Xu, Z., Mai, T., and Casali, P. (2009). HoxC4 binds to the promoter of the cytidine deaminase AID gene to induce AID expression, class-switch DNA recombination and somatic hypermutation. Nat. Immunol. *10*, 540–550.

Pasqualucci, L., Kitaura, Y., Gu, H., and Dalla-Favera, R. (2006). PKA-mediated phosphorylation regulates the function of activation-induced deaminase (AID) in B cells. Proc. Natl. Acad. Sci. U.S.a. *103*, 395–400.

Patenaude, A.-M., and Di Noia, J.M. (2010). The mechanisms regulating the subcellular localization of AID. Nucleus *1*, 325–331.

Patenaude, A.-M., Orthwein, A., Hu, Y., Campo, V.A., Kavli, B., Buschiazzo, A., and Di Noia, J.M. (2009a). Active nuclear import and cytoplasmic retention of activation-induced deaminase. Nat. Struct. Mol. Biol. *16*, 517–527.

Patenaude, A.-M., Orthwein, A., Hu, Y., Campo, V.A., Kavli, B., Buschiazzo, A., and Di Noia, J.M. (2009b). Active nuclear import and cytoplasmic retention of activation-induced deaminase. Nat. Struct. Mol. Biol. *16*, 517–527.

Pauklin, S., Sernández, I.V., Bachmann, G., Ramiro, A.R., and Petersen-Mahrt, S.K. (2009). Estrogen directly activates AID transcription and function. Journal of Experimental Medicine *206*, 99–111.

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.

Perlot, T., Li, G., and Alt, F.W. (2008). Antisense transcripts from immunoglobulin heavy-chain locus V(D)J and switch regions. Proceedings of the National Academy of Sciences *105*, 3843–3848.

Peters, A., and Storb, U. (1996). Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. Immunity *4*, 57–65.

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.

Phizicky, E.M., and Fields, S. (1995). Protein-protein interactions: methods for detection and analysis. Microbiol. Rev. *59*, 94–123.

Pickart, C.M. (2001a). Mechanisms underlying ubiquitination. Annu. Rev. Biochem. *70*, 503–533.

Pickart, C.M. (2001b). Ubiquitin enters the new millennium. Molecular Cell *8*, 499–504.

Pinaud, E., Khamlichi, A.A., Le Morvan, C., Drouet, M., Nalesso, V., Le Bert, M., and Cogné, M. (2001). Localization of the 3' IgH locus elements that effect long-distance regulation of class switch recombination. Immunity *15*, 187–199.

Pinaud, E., Marquet, M., Fiancette, R., Péron, S., Vincent-Fabert, C., Denizot, Y., and Cogné, M. (2011). The IgH locus 3' regulatory region: pulling the strings from behind. Adv. Immunol. *110*, 27–70.

Pone, E.J., Zhang, J., Mai, T., White, C.A., Li, G., Sakakura, J.K., Patel, P.J., Al-Qahtani, A., Zan, H., Xu, Z., et al. (2012). BCR-signalling synergizes with TLR-signalling for induction of AID and immunoglobulin class-switching through the non-canonical NF-κB pathway. Nat Commun *3*, 767–.

Prabakaran, S., Lippens, G., Steen, H., and Gunawardena, J. (2012). Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. Wiley Interdiscip Rev Syst Biol Med *4*, 565–583.

Rada, C., and Milstein, C. (2001). The intrinsic hypermutability of antibody heavy and light chain genes decays exponentially. Embo J. *20*, 4570–4576.

Rada, C. (2009). AID and RPA: PKA makes the connection local. Nat. Immunol. *10*, 367–369.

Rada, C., Jarvis, J.M., and Milstein, C. (2002a). AID-GFP chimeric protein increases hypermutation of Ig genes with no evidence of nuclear localization. Proc. Natl. Acad. Sci. U.S.a. *99*, 7003–7008.

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

Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., and Jones, D.A. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell *135* (7), 1201-12.

Ramachandran, S., Chahwan, R., Nepal, R.M., Frieder, D., Panier, S., Roa, S., Zaheen, A., Durocher, D., Scharff, M.D., and Martin, A. (2010). The RNF8/RNF168 ubiquitin ligase cascade facilitates class switch recombination. Proceedings of the National Academy of Sciences *107*, 809–814.

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.

Rawlings, D.J., Schwartz, M.A., Jackson, S.W., and Meyer-Bahlburg, A. (2012). Integration of B cell responses through Toll-like receptors and antigen receptors. Nat Rev Immunol *12*, 282–294.

Reina San-Martin, B., Difilippantonio, S., Hanitsch, L., Masilamani, R.F., Nussenzweig, A., and Nussenzweig, M.C. (2003). H2AX is required for recombination between immunoglobulin switch regions but not for intra-switch region recombination or somatic hypermutation. J. Exp. Med. *197*, 1767–1778.

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A., et al. (2000). Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). Cell *102*, 565–575.

Rivera-Munoz, P., and Soulas-Sprauel, P. (2009). Reduced immunoglobulin class switch recombination in the absence of Artemis. Blood *114* (17), 3601-9.

Roa, S., Avdievich, E., Peled, J.U., Maccarthy, T., Werling, U., Kuang, F.L., Kan, R., Zhao, C., Bergman, A., Cohen, P.E., et al. (2008). Ubiquitylated PCNA plays a role in somatic hypermutation and class-switch recombination and is required for meiotic progression. Proceedings of the National Academy of Sciences *105*, 16248–16253.

Roa, S., Li, Z., Peled, J.U., Zhao, C., Edelmann, W., and Scharff, M.D. (2010). MSH2/MSH6 complex promotes error-free repair of AID-induced dU:G mispairs as well as error-prone hypermutation of A:T sites. PLoS ONE *5*, e11182.

Roach, J.C., Glusman, G., Rowen, L., Kaur, A., Purcell, M.K., Smith, K.D., Hood, L.E., and Aderem, A. (2005). The evolution of vertebrate Toll-like receptors. Proc. Natl. Acad. Sci. U.S.a. *102*, 9577–9582.

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. Molecular Cell *36*, 631–641.

Robert, I., Dantzer, F., and Reina San-Martin, B. (2009). Parp1 facilitates alternative NHEJ, whereas Parp2 suppresses IgH/c-myc translocations during immunoglobulin class switch recombination. Journal of Experimental Medicine *206* (5), 1047-56.

Rocha, P.P., Micsinai, M., Kim, J.R., Hewitt, S.L., Souza, P.P., Trimarchi, T., Strino, F., Parisi, F., Kluger, Y., and Skok, J.A. (2012). Close proximity to Igh is a contributing factor to AID-mediated translocations. Molecular Cell *47*, 873–885.

Rogozin, I.B., and Kolchanov, N.A. (1992). Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. Biochim. Biophys. Acta *1171*, 11–18.

Rogozin, I.B., Iyer, L.M., Liang, L., Glazko, G.V., Liston, V.G., Pavlov, Y.I., Aravind, L., and Pancer, Z. (2007). Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. Nat. Immunol. *8*, 647–656.

Roux, K.J., Kim, D.I., Raida, M., and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J. Cell Biol. *196*, 801–810.

Rubio, M.A.T., Pastar, I., Gaston, K.W., Ragone, F.L., Janzen, C.J., Cross, G.A.M., Papavasiliou, F.N., and Alfonzo, J.D. (2007). An adenosine-to-inosine tRNA-editing enzyme that can perform C-to-U deamination of DNA. Proc. Natl. Acad. Sci. U.S.a. *104*, 7821–7826.

Rudnicki, M.A., Braun, T., Hinuma, S., and Jaenisch, R. (1992). Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development. Cell *71*, 383–390.

Rudnicki, M.A., Schnegelsberg, P.N., Stead, R.H., Braun, T., Arnold, H.H., and Jaenisch, R. (1993). MyoD or Myf-5 is required for the formation of skeletal muscle. Cell *75*, 1351–1359.

Sale, J.E., and Neuberger, M.S. (1998). TdT-accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line. Immunity *9*, 859–869.

Sayegh, C.E., Quong, M.W., Agata, Y., and Murre, C. (2003). E-proteins directly regulate expression of activation-induced deaminase in mature B cells. Nat. Immunol. *4*, 586–593.

Schatz, D.G., and Baltimore, D. (1988). Stable expression of immunoglobulin gene V(D)J recombinase activity by gene transfer into 3T3 fibroblasts. Cell *53*, 107–115.

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

Schatz, D.G., and Baltimore, D. (2004). Uncovering the V(D)J recombinase. Cell *S116*, S103-106.

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

Schmidt-Supprian, M., and Rajewsky, K. (2007). Vagaries of conditional gene targeting. Nat. Immunol. *8*, 665–668.

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., Guikema, J.E.J., Linehan, E.K., Selsing, E., and Stavnezer, J. (2007). Activation-induced cytidine deaminase-dependent DNA breaks in class switch recombination occur during G1 phase of the cell cycle and depend upon mismatch repair. J. Immunol. *179*, 6064–6071.

Serizawa, S., Miyamichi, K., and Sakano, H. (2004). One neuron-one receptor rule in the mouse olfactory system. Trends Genet. *20*, 648–653.

Shinkura, R., Tian, M., Smith, M., Chua, K., Fujiwara, Y., and Alt, F.W. (2003). The influence of transcriptional orientation on endogenous switch region function. Nat. Immunol. *4*, 435–441.

Simpson, L.J., and Sale, J.E. (2003). Rev1 is essential for DNA damage tolerance and non-templated immunoglobulin gene mutation in a vertebrate cell line. Embo J. *22*, 1654–1664.

Singh, S.K., Maeda, K., Eid, M.M.A., Almofty, S.A., Ono, M., Pham, P., Goodman, M.F., and Sakaguchi, N. (2013). GANP regulates recruitment of AID to immunoglobulin variable regions by modulating transcription and nucleosome occupancy. Nat Commun *4*, 1830.

Smith, C.J., Berry, D.M., and McGlade, C.J. (2013). The E3 ubiquitin ligases RNF126 and Rabring7 regulate endosomal sorting of the Epidermal Growth Factor Receptor. J. Cell. Sci. *126*, 1366-80.

Soulas-Sprauel, P., Le Guyader, G., Rivera-Munoz, P., Abramowski, V., Olivier-Martin, C., Goujet-Zalc, C., Charneau, P., and de Villartay, J.-P. (2007). Role for DNA repair factor XRCC4 in immunoglobulin class switch recombination. The Journal of Experimental Medicine *204* (7), 1717-27.

Sprinzak, E., Sattath, S., and Margalit, H. (2003). How reliable are experimental protein-protein interaction data? J. Mol. Biol. *327*, 919–923.

Stanlie, A., Aida, M., Muramatsu, M., Honjo, T., and Begum, N.A. (2010). Histone3 lysine4 trimethylation regulated by the facilitates chromatin transcription complex is critical for DNA cleavage in class switch recombination. Proceedings of the National Academy of Sciences *107*, 22190–22195.

Stavnezer, J. (1996). Immunoglobulin class switching. Curr. Opin. Immunol. *8*, 199–205.

Stavnezer, J., and Schrader, C.E. (2006). Mismatch repair converts AID-instigated nicks to double-strand breaks for antibody class-switch recombination. Trends Genet. *22*, 23–28.

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.

Storb, U., Shen, H.M., Longerich, S., Ratnam, S., Tanaka, A., Bozek, G., and Pylawka, S. (2007). Targeting of AID to immunoglobulin genes. Adv. Exp. Med. Biol. *596*, 83–91.

Tashiro, J., Kinoshita, K., and Honjo, T. (2001). Palindromic but not G-rich sequences are targets of class switch recombination. Int. Immunol. *13*, 495–505.

Teng, G., and Papavasiliou, F.N. (2007). Immunoglobulin somatic hypermutation. Annu. Rev. Genet. *41*, 107–120.

Teng, G., Hakimpour, P., Landgraf, P., Rice, A., Tuschl, T., Casellas, R., and Papavasiliou, F.N. (2008). MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. Immunity *28*, 621–629.

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

Tran, T.H., Nakata, M., Suzuki, K., Begum, N.A., Shinkura, R., Fagarasan, S., Honjo, T., and Nagaoka, H. (2009). B cell–specific and stimulation-responsive enhancers derepress Aicda by overcoming the effects of silencers. Nat. Immunol. *11*, 148–154.

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

Uchimura, Y., Barton, L.F., Rada, C., and Neuberger, M.S. (2011). REG-γ associates with and modulates the abundance of nuclear activation-induced deaminase. Journal of Experimental Medicine *208*, 2385–2391.

Unniraman, S., and Schatz, D.G. (2007). Strand-biased spreading of mutations during somatic hypermutation. Science *317*, 1227–1230.

Vincent-Fabert, C., Fiancette, R., Pinaud, E., Truffinet, V., Cogné, N., Cogné, M., and Denizot, Y. (2010). Genomic deletion of the whole IgH 3' regulatory region (hs3a, hs1,2, hs3b, and hs4) dramatically affects class switch recombination and Ig secretion to all isotypes. Blood *116*, 1895–1898.

Vuong, B.Q., and Chaudhuri, J. (2012). Combinatorial mechanisms regulating AID-dependent DNA deamination: interacting proteins and post-translational modifications. Semin. Immunol. *24*, 264–272.

Vuong, B.Q., Lee, M., Kabir, S., Irimia, C., Macchiarulo, S., McKnight, G.S., and Chaudhuri, J. (2009). Specific recruitment of protein kinase A to the immunoglobulin locus regulates class-switch recombination. Nat. Immunol. *10*, 420–426.

Wang, X., Herr, R.A., Chua, W.-J., Lybarger, L., Wiertz, E.J.H.J., and Hansen, T.H. (2007). Ubiquitination of serine, threonine, or lysine residues on the cytoplasmic tail can induce ERAD of MHC-I by viral E3 ligase mK3. J. Cell Biol. *177*, 613–624.

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

Watson, F.L., Püttmann-Holgado, R., Thomas, F., Lamar, D.L., Hughes, M., Kondo, M., Rebel, V.I., and Schmucker, D. (2005). Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science *309*, 1874–1878.

Wei, W., Ba, Z., Gao, M., Wu, Y., Ma, Y., Amiard, S., White, C.I., Danielsen, J.M.R., Yang, Y.-G., and Qi, Y. (2012). A Role for Small RNAs in DNA Double-Strand Break Repair. Cell 1–12.

Weigert, M.G., Cesari, I.M., Yonkovich, S.J., and Cohn, M. (1970). Variability in the lambda light chain sequences of mouse antibody. Nature *228*, 1045–1047.

Wiesendanger, M., Scharff, M.D., and Edelmann, W. (1998). Somatic hypermutation, transcription, and DNA mismatch repair. Cell *94*, 415–418.

Winter, D.B., Sattar, N., Mai, J.J., and Gearhart, P.J. (1997). Insertion of 2 kb of bacteriophage DNA between an immunoglobulin promoter and leader exon stops somatic hypermutation in a kappa transgene. Mol. Immunol. *34*, 359–366.

Wu, X., and Stavnezer, J. (2007). DNA polymerase beta is able to repair breaks in switch regions and plays an inhibitory role during immunoglobulin class switch recombination. J. Exp. Med. *204*, 1677–1689.

Xu, Z., Fulop, Z., Wu, G., Pone, E.J., Zhang, J., Mai, T., Thomas, L.M., Al-Qahtani, A., White, C.A., Park, S.-R., et al. (2010). 14-3-3 adaptor proteins recruit AID to 5“-AGCT-3-”rich switch regions for class switch recombination. Nat. Struct. Mol. Biol. *17*, 1124–1135.

Xu, Z., Zan, H., Pone, E.J., Mai, T., and Casali, P. (2012). Immunoglobulin class-switch DNA recombination: induction, targeting and beyond. Nat Rev Immunol *12*, 517–531.

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.

Yabuki, M., Fujii, M.M., and Maizels, N. (2005). The MRE11-RAD50-NBS1 complex accelerates somatic hypermutation and gene conversion of immunoglobulin variable regions. Nat. Immunol. *6*, 730–736.

Yadav, A., Olaru, A., Saltis, M., Setren, A., Cerny, J., and Livák, F. (2006). Identification of a ubiquitously active promoter of the murine activation-induced cytidine deaminase (AICDA) gene. Mol. Immunol. *43*, 529–541.

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.

Yamane, A., Robbiani, D.F., Resch, W., Bothmer, A., Nakahashi, H., Oliveira, T., Rommel, P.C., Brown, E.J., Nussenzweig, A., Nussenzweig, M.C., et al. (2013). RPA accumulation during class switch recombination represents 5“-3” DNA-end resection during the S-G2/M phase of the cell cycle. Cell Rep *3*, 138–147.

Yin, F.F., Bailey, S., Innis, C.A., Ciubotaru, M., Kamtekar, S., Steitz, T.A., and Schatz, D.G. (2009). Structure of the RAG1 nonamer binding domain with DNA reveals a dimer that mediates DNA synapsis. Nat. Struct. Mol. Biol. *16*, 499–508.

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.

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.

Zarrin, A.A., Alt, F.W., Chaudhuri, J., Stokes, N., Kaushal, D., Pasquier, Du, L., and Tian, M. (2004). An evolutionarily conserved target motif for immunoglobulin class-switch recombination. Nat. Immunol. *5*, 1275–1281.

Zarrin, A.A., Del Vecchio, C., Tseng, E., Gleason, M., Zarin, P., Tian, M., and Alt, F.W. (2007). Antibody class switching mediated by yeast endonuclease-generated DNA breaks. Science *315*, 377–381.

Zeng, X., Winter, D.B., Kasmer, C., Kraemer, K.H., Lehmann, A.R., and Gearhart, P.J. (2001). DNA polymerase eta is an A-T mutator in somatic hypermutation of immunoglobulin variable genes. Nat. Immunol. *2*, 537–541.

Zeng, X., Negrete, G.A., Kasmer, C., Yang, W.W., and Gearhart, P.J. (2004). Absence of DNA polymerase eta reveals targeting of C mutations on the nontranscribed strand in immunoglobulin switch regions. J. Exp. Med. *199*, 917–924.

Zhang, S., Chea, J., Meng, X., Zhou, Y., Lee, E.Y.C., and Lee, M.Y.W.T. (2008). PCNA is ubiquitinated by RNF8. Cell Cycle *7*, 3399–3404.

Zhi, X., Zhao, D., Wang, Z., Zhou, Z., Wang, C., Chen, W., Liu, R., and Chen, C. (2013). E3 ubiquitin ligase RNF126 promotes cancer cell proliferation by targeting the tumor suppressor p21 for ubiquitin-mediated degradation. Cancer Res. *73*, 385–394.