# RNA Exosome Regulated Antisense and Divergent Noncoding RNA Facilitate AID Targeting Throughout the B Cell Genome

Evangelos Pefanis

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2015

# ABSTRACT

## RNA Exosome Regulated Antisense and Divergent Noncoding RNA Facilitate AID Targeting Throughout the B Cell Genome

Evangelos Pefanis

Vertebrate immune systems are armed with the ability to generate highly specific immune responses capable of responding to nearly any foreign molecular threat. One of the major mediators of this response is immunoglobulins (Igs) produced by B lymphocytes. The specificity of individual Igs is created through a tightly orchestrated series of somatic DNA manipulations at Ig encoding loci resulting in functional gene rearrangements and nucleotide substitutions. These events serve to create a pool of naive B cells expressing Igs with distinct specificities, capable of expansion in response to antigen specific selection. Affinity of Ig towards antigen is enhanced through nucleotide substitutions introduced at the antigen binding variable region gene segments through the enzyme activation induced cytidine deaminase (AID) during the process of somatic hypermutation (SHM). AID also generates point mutations within noncoding DNA segments of the *Ig* heavy chain locus that are processed into double strand breaks leading to constant region isotype switching during class switch recombination (CSR).

The Ig diversification processes of SHM and CSR critically depend upon transcriptional activation of the relevant DNA segments. Transcription is thought to facilitate single strand DNA substrate recognition by AID during unwinding of the DNA duplex. The 3'-5' exoribonuclease RNA exosome serves as a transcription dependent cofactor of AID. RNA exosome is comprised of multiple structurally integral core subunits

and associated nuclease subunits. In this work, RNA exosome core subunit *Exosc3* and nuclease *Exosc10* have been targeted for conditional mutagenesis and loss of function analysis in mouse cells. RNA exosome deficient B cells were significantly impaired in AID dependent SHM and CSR *Ig* diversification processes. Transcriptome analyses revealed a striking accumulation of promoter proximal antisense divergent noncoding transcripts (xTSS-RNA) at a subset of genes upon loss of RNA exosome function. xTSS-RNAs mark regions of chromatin containing RNA exosome activity. Multiple known AID target sites including *IgH* and *Myc* were observed to express xTSS-RNA. Furthermore, genomic sites of recurrent AID dependent chromosomal translocations were enriched for xTSS-RNA. In addition to promoter proximal xTSS-RNA, cryptic intragenic antisense noncoding transcripts were found to accumulate at many genomic loci. In fact, multiple translocation hotspots precisely overlap regions of RNA exosome sensitive antisense transcription. AID targeted divergently transcribed promoters containing RNA exosome substrates possessed greater amounts of RNA:DNA hybrids, indicative of frequent transcriptional arrest. Lastly, RNA exosome deficient transcriptomes have revealed a substantial number of novel long intergenic noncoding RNAs and enhancer RNAs, indicating a hidden layer of cellular transcriptional activity.

A model of AID targeting utilizing transcriptional arrest is becoming increasingly apparent. Transcribed chromatin prone to undergo transcriptional arrest, such as *Ig* loci or xTSS-RNA expressing regions, frequently undergoes premature transcription termination coupled to RNA exosome mediated degradation of the nascent transcript. This process results in the creation of AID substrates and serves to stabilize its association with chromatin through multiple interactions involving RNA exosome and transcription complex subunits.

# Table of contents

# List of figures

# Acknowledgements

I begin by expressing sincere gratitude to my thesis advisor Uttiya Basu. It was a privilege conducting science together in such a collegial manner. I greatly appreciate the respectful treatment that I received from Uttiya over the years. The greatest lesson Uttiya has taught me was how to conduct science using a “top-down” approach, with an eye on the big picture first. My approach to science has always been to build up from a sea of details. This would cause me to sometimes lose sight of the forest amongst the trees. Observing and incorporating Uttiya’s approach has made me a better scientist and for that I am most grateful. Lastly, I must thank Uttiya for his genuine concern for me on a personal level. On many occasions he has taken time out of his busy schedule to check on my wellbeing. He provided sound advice when I needed it, space when I was at my limits, and most importantly a kick in the rear when I deserved it!

I thank my thesis committee members Saul Silverstein, Stephen P. Goff, and Sankar Ghosh. Their encouragement and insightful comments facilitated the completion of this work. Special thanks are given to Saul Silverstein for his mentorship during my rotation in his lab many years ago, for facilitating my return to the department, and for his critical reading of this thesis during its construction.

I thank Frederick W. Alt for serving as my external thesis examiner and for what I expect to be nothing less than tremendous input and insight. He is a scientist that I have looked up to for many years. I am truly humbled by his participation.

I thank all past and present members of the Basu lab for their friendship. In particular, I thank Gerson Rothschild for his contributions with ChIP experiments and for making lab a fun place to be around. I especially thank Jiguang Wang, for conducting many of the bioinformatic analyses in this project. I also thank my fellow gene targeter in the Basu lab, Junghyun Lim, for creating the xTSS-RNA CRISPR deleted cell lines. I

thank Jianbo Sun, David Kazadi, and Veronika Grinstein for their comments during lab meetings and being great lab mates. I also thank Jaime Chao for all the fun conversations in lab and for her great help in managing the mouse colony.

I thank the entire Microbiology & Immunology department for fostering an intellectually challenging yet supportive environment. I enjoyed learning from all the great research being conducted within the department. I especially would like to express my appreciation to Lorraine Symington for her support of my work through what I can only surmise were countless scientific discussions with Uttiya. I would also like to thank Ulf Klein for his engaging scientific happy hour conversations over the years. I thank David Fidock and Boris Reizis for their guidance as directors of graduate studies.

I thank countless longtime colleagues at Regeneron. First and foremost I thank George Yancopoulos for providing me with the opportunity to chase my dream. I am forever grateful for his support and the trust he placed in me. I thank Sam Davis for guiding and encouraging me during my inexperienced years. I thank Lynn Macdonald and Andrew Murphy for being excellent colleagues over the years and also for sharing the *Aicda*$^{Cre}$ allele used in this study. I thank Joyce McClain, Nina Schilt, Darshi Persaud, and Kieran Feeley for their friendship and support over many years of working together. I thank my good friend Richard Corpina for sharing in my successes and failures, and for being there when I needed someone to lean on. Most importantly here, I thank Aris Economides for being like a big brother to me. I am deeply appreciative for his genuine pride in my accomplishments, for his selflessness and generosity. I look forward to many more years of working together.

Lastly, I offer my greatest thanks to my dear friends and classmates Sarah Deng, Łucja Grajkowska, and Nilushi De Silva. The good times were epic, the laughs were sometimes painful, and the memories are eternal.

To my father, Ioannis Pefanis, for teaching me the value of hard work, for congratulating me when I did my best, and for motivating me in a positive way to push even a little harder the next time around.

To my mother, Dionysia Pefanis, for always being there, for never failing me even once, I always knew that I could count on you to be there with your unconditional support.

To my daughters, Gianna and Gabriella, I thank you for your encouragement and patience during my years of graduate school. Your love and support has carried me along. You remain, my greatest accomplishment.

# Chapter 1: Introduction

## 1.1 Brief overview of innate and adaptive immune responses

All species of life have evolved various defense mechanisms providing protection against pathogens and other non-self antigens. The majority of species rely almost exclusively on germline encoded non-specific approaches referred to as the innate immune system, which is fully operable prior to pathogen encounter. However, during evolution all vertebrate species acquired an additional mechanism of immunity involving specific recognition of pathogen during the course of an immune response (Hirano et al., 2011). This process is known as the adaptive immune response and involves introduction of somatic genome alterations to generate antigen recognition. Both innate and adaptive arms of the immune system can be further subdivided according to the nature of their mediators. Cell mediated immune responses require direct interactions between host cells and their target whereas humoral immune responses involve soluble factors.

Cellular mediators of the innate immune response include phagocytes such as macrophages and neutrophils. These cell types are capable of engulfing particulate antigens such as immune complexes, opsinized microorganisms, or infected cells and thus are crucial for infection clearance. Cell mediated innate responses also involve natural killer cells and innate lymphoid cells that are involved in direct killing of infected cells and release a wide array of pro-inflammatory cytokines (Spits and Cupedo, 2012). Humoral innate immunity is largely comprised of the complement system, a complex cascade of proteolytic events involving host serum factors eventually leading to opsinization and/or lysis of invading pathogens (Ricklin et al., 2010). Anti-microbial peptides such as defensins are an additional component of humoral innate immunity. These are located primarily on mucosal surfaces and are capable of penetrating microbial cell membranes leading to lysis and/or inhibition of metabolism (Brogden, 2005).

T lymphocytes are the key cell type involved in executing the cell mediated adaptive immune response. Their specificity towards antigen is imparted through the T cell receptor (TCR) that recognizes antigen in the form of membrane bound peptide-MHC complexes on other cells. Their function is largely dependent on receiving activation cues arising from antigen specific interactions with dendritic cells. The dendritic cell in many ways bridges the innate and adaptive immune responses by translating non-specific phagocytosis of antigen into a potent antigen specific activation signal for T cells via cell surface antigen presentation. Activated cytotoxic T cells of the $CD8^+$ lineage drive the cell mediated adaptive immune response by specific killing of infected cells presenting non-self peptides on their MHC-I cell surface molecules. $CD4^+$ T helper cells relay critical activation and survival cues to B cells.

According to current knowledge, immunoglobulin (Ig) produced by B lymphocytes entirely accounts for the humoral adaptive immune response. Ig can bind antigen with high affinity, great specificity, and is a major protein constituent of serum. Interactions between Ig and antigen can lead to neutralization of pathogenic antigens such as viruses and bacterial toxins. Ig accumulation on large antigenic surfaces such as bacterial polysaccharides can induce their phagocytic uptake and degradation through macrophages. This same process can lead to complement-dependent cytotoxicity (CDC), where complement activation occurs on the surfaces of Ig coated microbial pathogens leading to lytic pore formation through the binding of complement initiator C1q with Ig. The accumulation of Ig on cell surfaces can also generate antibody-dependent cell-mediated cytotoxicity (ADCC), where high avidity interactions between Ig and low affinity Fc receptors on natural killer cells can induce release of cytotoxic granules and lead to target cell apoptosis (Vivier et al., 2008). In light of these functions, Igs have repeatedly been leveraged for therapeutic purposes as evidenced by the

successful application of anti-CD20, anti-HER2, anti-VEGF, and anti-CTLA4 in the fight against various cancers (Reichert, 2012).

## 1.2 Primary immunoglobulin diversification – V(D)J recombination

Immunoglobulins are composed of 2 identical ~50 kDa heavy chain polypeptides and 2 identical ~25 kDa light chain polypeptides covalently linked by disulfide bridges. Each polypeptide contains distinct variable and constant regions. The constant regions of heavy and light chains are entirely germline encoded, and the heavy chain constant region determines Ig isotype, which imparts many of the effector functions. The variable regions directly bind with antigen and are not directly encoded within the genome. Instead, variable regions are encoded within discontinuous gene segments requiring assembly through an elaborate mechanism involving combinatorial rearrangement of somatic DNA.

In 1976, Tonegawa provided the initial insight pointing to immunoglobulin production involving somatic rearrangement. In a landmark experiment he observed that a radiolabelled light chain mRNA probe hybridized with two distinct DNA restriction fragments when using early embryo DNA, whereas DNA derived from a plasmacytoma line contained a single restriction fragment hybridizing with the same light chain probe (Hozumi and Tonegawa, 1976). This was correctly interpreted as the variable and constant regions being spaced far apart within the germline but that they undergo a somatic rearrangement bringing them in close proximity during B cell development. It was subsequently determined that light chain variable regions are assembled from two gene segments (variable and joining) (Max et al., 1979; Sakano et al., 1979) and heavy chain variable regions are assembled from three gene segments (variable, diversity, and joining) (Early et al., 1980; Maki et al., 1980) through the process of V(D)J recombination.

V(D)J recombination is the ordered, lineage, and stage specific assembly of variable region antigen receptor gene segments. Lineage specificity is evidenced by the fact that *Ig* loci are fully rearranged only in B cells and not in T cells. Stage specific V(D)J recombination in B cells occurs during development where the heavy chain locus is rearranged prior to the light chain locus. Heavy chain rearrangement occurs during the pro-B cell stage whereas light chain rearrangement occurs during the pre-B cell stage. Joining of $V_H$, $D_H$, and $J_H$ segments during heavy chain rearrangement occurs in a tightly ordered fashion where $D_H$ and $J_H$ segments are rearranged first to form a $D_HJ_H$ fusion segment followed by $V_H$ to $D_HJ_H$ joining (Alt et al., 1984). Ordered rearrangement of gene segments is likely mediated through localized chromatin loops established by CTCF (Guo et al., 2011a; Guo et al., 2011b).

A key event in initiation of V(D)J recombination is induction of germline transcription at unrearranged V, D, and J segments. These germline transcripts were found to coincide with the particular class of gene segments undergoing rearrangement according to the lineage and stage specificity of the cell (Yancopoulos and Alt, 1985). From this observation sprung the accessibility model of V(D)J recombination, which hypothesized that the ordered, lineage, and stage specific rearrangement of variable region gene segments was accomplished through the differential accessibility of gene segments towards the putative recombinase. The epigenetic “opening” of local chromatin was proposed to provide access of given gene segments to transcription and recombinase machineries and helped resolve the long standing question of why antigen receptor gene segments of the B cell lineage are not recombined in the T cell lineage and vice versa. Germline transcripts were subsequently shown to coincide with the developmental onset of V(D)J recombination at all known antigen receptor loci (Abarrategui and Krangel, 2009).

The mechanism of V(D)J recombination can be largely summarized as a two step process (Figure 1). The first involves site specific recognition and cleavage of V, D, or J gene segments by the lymphocyte specific RAG1/RAG2 endonuclease complex (Oettinger et al., 1990; Schatz et al., 1989). RAG cleavage occurs within recombination signal sequences flanking gene segments, leading to formation of hairpin coding ends. The second phase of V(D)J recombination involves joining of cleaved gene segments and is largely mediated through ubiquitously expressed components of the classical non-homologous end-joining pathway (Alt et al., 2013). Broken coding ends are then held together by the Ku70/Ku80 heterodimer (Walker et al., 2001), which also serves to recruit the DNA-dependent protein kinase catalytic subunit (DNA-PKcs) leading to Artemis endonuclease activation (Lieber, 2010). Artemis will then introduce a nick into the single stranded portion of the coding end hairpins (Ma et al., 2002). The free 3' end can then be processed by terminal deoxynucleotidyl transferase (TdT) to introduce non-templated nucleotides and filled in by DNA polymerase μ or λ (Schatz and Swanson, 2011). In the final step the ligase complex composed of XLF, XRCC4, and DNA ligase IV seals the broken ends (Lieber, 2010) to form the fusion gene segment.

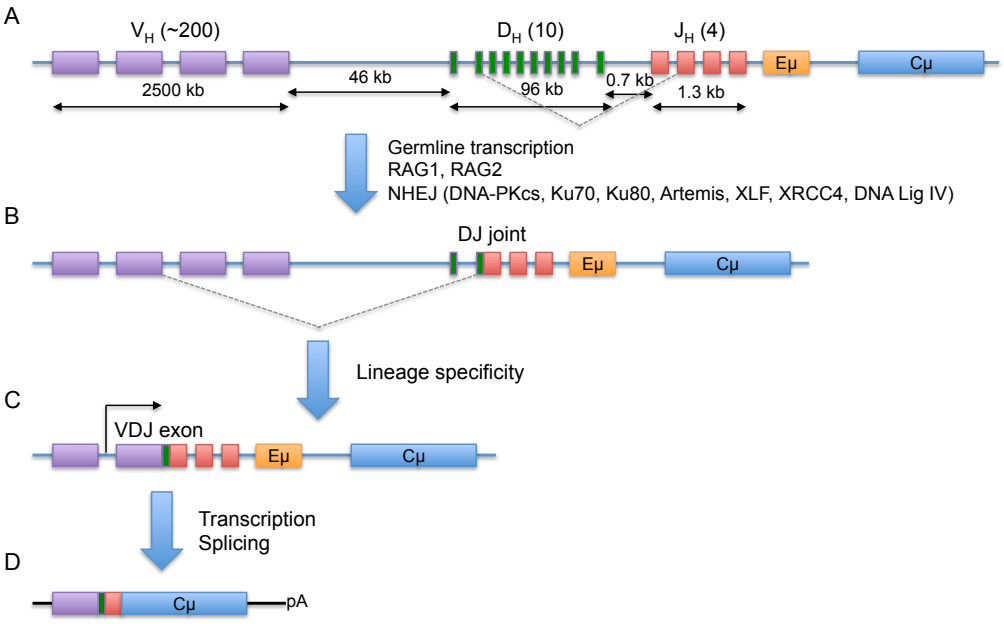


**Figure 1: V(D)J recombination at the mouse *Ig* heavy chain locus**
(A - B) The ~2.7 Mb *Ig* heavy chain locus in germline orientation. Variable ($V_H$) gene segments (~200 in total) are spaced across ~2.5 Mb. Ten diversity ($D_H$) gene segments are spaced across 96 kb. Four joining ($J_H$) gene segments occupy 1.3 kb. The intronic enhancer element (Eμ) lies between the last $J_H$ segment and the Cμ constant region. V(D)J recombination is initiated through germline transcription of $D_H$ and $J_H$ gene segments. RAG endonuclease cleaves the corresponding $D_H$ and $J_H$ segments that are subsequently joined through the NHEJ pathway to produce a fusion $D_HJ_H$ segment.
(B - C) A second recombination event largely following the same mechanism as the first, takes place between a $V_H$ segment and the fusion $D_HJ_H$ segment. The $V_H$ to $D_HJ_H$ recombination event occurs with tight lineage specificity.
(C - D) The $V_HD_HJ_H$ segment serves as the variable region coding exon of the mature heavy chain mRNA.

## 1.3 Secondary Ig diversification – somatic hypermutation & class switch recombination

### 1.3.1 *B cell diversification and selection within the germinal center*

Naive B cells that productively rearrange their *Ig* heavy and light chain loci resulting in functional non-autoreactive surface immunoglobulin (BCR) exit the bone marrow and circulate until they encounter cognate antigen. The fine-tuning of immunoglobulin takes place within a microanatomical structure known as the germinal center, which develops within secondary lymphoid organs upon immunization or antigenic stimulation. It is here where B cells introduce somatic mutations in *Ig* variable regions through the process of somatic hypermutation (SHM) and undergo isotype switching through the process of class switch recombination (CSR). SHM coupled with intercellular B cell selection defines the physiological phenomenon of affinity maturation, a process where the affinity of serum immunoglobulin towards antigen increases over time during an immune response. CSR leads to alteration of immunoglobulin effector function by specifically replacing the germline encoded IgM heavy chain isotype of naive B cells with a different isotype such as IgG, IgE, or IgA. Each of these isotypes differs in their accumulation within various bodily fluids, their molecular stoichiometry, and ability to interact with different cell types according to isotype specific Fc receptor expression.

Naive B cells circulating through secondary lymphoid organs such as spleen or lymph node occasionally encounter BCR-specific antigen and acquire T cell help,

resulting in B cell activation and eventually leading to germinal center formation. Germinal centers are comprised of two distinct histologic domains. A dark zone composed almost exclusively of proliferating B cells, and a light zone containing follicular dendritic cells (FDCs), B cells, and T helper cells (Victora and Nussenzweig, 2012). B cells within the dark zone undergo rapid cell division and express the enzyme activation induced cytidine deaminase (AID) that initiates mutagenesis of actively transcribed *Ig* loci. This process creates a pool of antigen specific B cell variants that undergo affinity based competitive selection involving FDCs and T helper cells within the light zone.

The germinal center is a highly dynamic environment involving extensive cell migration and rapid changes in gene expression (Allen et al., 2007; Hauser et al., 2007; Schwickert et al., 2007). Replicating B cells are maintained in the dark zone due to their high expression of chemokine receptor CXCR4, whereas B cell migration to the light zone is promoted through upregulation of CXCR5 (Allen et al., 2004). Here B cells alter their transcriptional signature from cell division to cell activation (Victora et al., 2010). The activation signal arises from BCR ligation of native antigen-immune complexes retained on the surface of FDCs (Carroll and Isenman, 2012). BCR signaling provides a necessary survival signal for B cells (Lam et al., 1997), and while important for elimination of B cells containing nonsense mutations in either *Ig* locus, is likely not the primary determinant of affinity based selection. A model whereby BCR mediated competition for antigen is solely responsible for selection of high affinity B cells would paradoxically also promote autoreactivity through selection of B cells with higher affinity towards self antigens (Victora and Nussenzweig, 2012). In addition to enhanced BCR signaling, higher BCR affinity towards antigen might also result in greater antigen endocytosis, processing, and presentation onto MHC class II molecules. Accordingly, affinity based B cell survival within the germinal center is thought to operate at the level of T cell help. B cells with the highest affinity towards antigen will engulf more antigen

than lower affinity B cells, and ultimately present more peptide-MHC II on their cell surface. T cells can preferentially synapse with B cells according to surface peptide-MHC II concentration (Depoil et al., 2005), and thereby deliver differential CD40 survival signals. Recent studies using transgenic approaches have formally shown that the amount of antigen captured and presented by germinal center B cells to T follicular helper cells in the light zone regulates B cell division, somatic hypermutation, and Ig affinity maturation (Gitlin et al., 2014). Strong T cell derived CD40 signaling in germinal center B cells leads to NFκB activation, resulting in IRF4 mediated release of Bcl-6 repression of Blimp-1 (De Silva et al., 2012). This signaling pathway ultimately leads to germinal center B cell differentiation toward the plasmablast fate and exit from the germinal center. Memory B cells are also produced during the germinal center reaction, although the signaling mechanisms involved are poorly understood. B cells within the light zone may also migrate back to the dark zone through the upregulation of CXCR4 to repeat the mutagenesis, proliferation, and selection cycle. This presumably provides antigen specific B cells additional opportunities to increase the affinity of their immunoglobulins by repeating the hypermutation process.

An important feature of the selection process is the establishment of a transient pro-apoptotic state in germinal center B cells. In its absence, differences in antigen affinity fail to translate into functional consequences through the process of low affinity B cell apoptosis. In essence, the germinal center reaction is thought to impose a fate of death onto all B cells that must be overcome through acquisition of higher affinity BCRs. The transcription factor Bcl-6 plays a critical role in orchestrating the pro-apoptotic state (Basso and Dalla-Favera, 2010). Bcl-6 actively represses transcription of the anti-apoptotic factor Bcl-2 (Saito et al., 2009). In addition, Bcl-6 is thought to enhance the stringency of B cell survival by downregulating expression of BCR and CD40 signaling mediators (Basso et al., 2010; Shaffer et al., 2000). This likely raises the threshold for

BCR activation that presumably requires a compensatory rise in antigen affinity. Bcl-6 repression of p53 (Phan and Dalla-Favera, 2004) and ATR (Ranuncolo et al., 2007) also promotes a tolerant response towards DNA damage accumulating in germinal center B cells resulting from rapid DNA replication and AID action. Bcl-6 also plays a role in cell fate determination within the germinal center by repressing expression of the plasma cell differentiation factor Blimp-1 (Shaffer et al., 2000), thereby prolonging the residency time of B cells within the germinal center reaction.

### 1.3.2 *Regulation of AID expression and activity*

AID is a protein expressed in germinal center B cells that is strictly required during SHM and CSR in both humans and mice (Muramatsu et al., 2000; Revy et al., 2000). In its absence, both diversification processes are entirely ablated. AID was discovered through subtractive hybridization screening as a novel factor highly induced in a B lymphoma cell line upon CSR stimulation (Muramatsu et al., 1999). The cDNA for AID was found to encode a 24 kDa protein homologous to the RNA editing cytidine deaminase APOBEC-1. On the basis of this homology AID was predicted to deaminate the mRNA of a yet undiscovered DNA mutator leading to its functional activation (Muramatsu et al., 2000). Many lines of biochemical and genetic evidence have now determined that the substrate of AID is single stranded DNA (ssDNA). One of the first indications that DNA was the target of AID came from an experiment showing that AID expression in *E. coli* leads to accumulation of C to T transitions within the *E. coli* chromosome (Petersen-Mahrt et al., 2002). This was difficult to reconcile with the hypothesis that a specific eukaryotic mRNA served as the substrate of AID. In vitro experiments later indicated that although able to bind to both RNA and ssDNA, only on ssDNA can AID deaminate deoxycytidine residues to deoxyuridine (Bransteitter et al., 2003; Dickerson et al., 2003). RNA, duplex DNA, and RNA:DNA hybrids were poor

substrates of AID in these assays. Recently, AID dependent deoxyuridine residues have been detected at multiple *Ig* loci through biochemical methods (Maul et al., 2011).

Constitutive expression of AID in mice through the use of a ubiquitously expressed transgene leads to widespread tumorigenesis (Okazaki et al., 2003). Given the highly mutagenic potential of AID it is not surprising that its expression and activity are tightly regulated in B cells through multiple mechanisms. Transcription of the *Aicda* locus (encoding AID) is inhibited in naive B cells through binding of c-Myb and E2F family transcriptional repressors to *Aicda* intron 1 (Huong le et al., 2013; Tran et al., 2010). Binding of transcriptional activators Pax5 and E2A to this same region partially counters the inhibitory activity of c-Myb and E2F (Gonda et al., 2003; Tran et al., 2010). The *Aicda* locus also contains binding sites for transcriptional activators NFκB, STAT6, and HoxC4, all of whose activities are rapidly induced during conditions of *Aicda* induction such as CD40 ligation or LPS and IL-4 treatment (Dedeoglu et al., 2004; Park et al., 2009). AID mRNAs are also negatively regulated through binding of microRNAs. miR-155 (Dorsett et al., 2008; Teng et al., 2008) and miR-181b (de Yebenes et al., 2008) bind to the 3' untranslated region of the AID transcript and promote its degradation.

AID mutagenic activity can also be regulated through active subcellular partitioning of the protein. The majority of steady state AID is cytoplasmic (Rada et al., 2002a; Schrader et al., 2005). AID gains access to the nucleus through its N-terminal nuclear localization signal (NLS) that binds to importin alpha in the cytoplasm and is shuttled into the nucleus through the nuclear pore complex (Ito et al., 2004; Patenaude et al., 2009). Nuclear occupancy of AID is partly limited through active nuclear export. The CRM1 nuclear export factor can bind to a nuclear export signal (NES) at the C-terminal end of AID and ferry the protein back into the cytoplasm (Ito et al., 2004; McBride et al., 2004). AID also contains a cytoplasmic retention domain near the C-

terminus, distinct from its NES (Patenaude et al., 2009), which is responsible for cytoplasmic sequestration with translation elongation factor 1 alpha (eEF1A) (Hasler et al., 2011). Factors influencing the subcellular stability of AID have also been identified. Nuclear AID is polyubiquitinated and has a shorter half-life compared to cytoplasmic AID (Aoufouchi et al., 2008). The nuclear factor REG-γ can accelerate proteasome mediated degradation of AID (Uchimura et al., 2011). Conversely, Hsp90 chaperones can bind to and stabilize cytoplasmic AID (Orthwein et al., 2010). This intricate balance between active AID subcellular localization and stability is critical for mediating AID's immunoglobulin diversification functions while limiting potentially deleterious mutagenesis events.

### 1.3.3 *Transcription and AID function at Ig loci*

As the substrate specificity of AID is towards ssDNA, a key question was the context in which AID substrates become accessible. Preceding the discovery of AID, transcription of the target sequence was implicated as a critical step in the hypermutation process. It was shown that the variable region exon of a germline rearranged heavy chain transgene in mouse B cells contained far fewer mutations when its promoter was deleted (Fukita et al., 1998). However, targeting of SHM to variable regions was not specifically determined by the nature of the promoter, as transgenes containing heterologous promoters were still capable of undergoing SHM (Betz et al., 1994; Tumas-Brundage and Manser, 1997). Later experiments using drug inducible transgenes in a hypermutating cell line indicated that the rates of SHM and CSR directly correlated with the rate of transgene transcription (Bachl et al., 2001; Lee et al., 2001). Another critical experiment provided evidence that the nature of the transcribed sequence, variable region exons in the case of SHM, does not determine targeting specificity. Here, SHM was evaluated in B cell hybridomas containing a rearranged light

chain transgene where the Vκ promoter was duplicated such that the VκJκ exon and the Cκ region were transcribed separately (Peters and Storb, 1996). Under these conditions both VκJκ and the Cκ regions underwent SHM, but the intervening sequence did not. In addition to showing that SHM was not determined by the variable region sequence per se, it also provided strong evidence that SHM was specifically linked to transcription initiation. This was consistent with previous findings showing that *Ig* locus mutations are clustered within the variable region while the constant region is spared, despite transcription taking place across both regions arising from the variable region promoter (Both et al., 1990; Lebecque and Gearhart, 1990).

This functional association between SHM and transcription prompted experiments to determine whether AID physically interacts with the transcription machinery. Indeed, chromatin immunoprecipitation (ChIP) studies showed that AID occupies transcriptionally induced *IgH* loci during CSR (Nambu et al., 2003), thus implicating AID at sites of RNA polymerase II (RNAP II) activity. This same study also showed through co-immunoprecipitation experiments that AID and RNAP II physically interact within a complex (Nambu et al., 2003). More recently Spt5, an evolutionarily conserved component of the transcription machinery, was found to interact with AID (Pavri et al., 2010). Depletion of Spt5 through shRNA knockdown reduced AID occupancy at the *IgH* locus, and inhibited CSR by ~60% in primary B cells (Pavri et al., 2010). Spt5 is one half of the heterodimeric DSIF complex (Wada et al., 1998). DSIF inhibits transcription elongation and is critical for promoter proximal pausing of RNAP II (Renner et al., 2001; Yamaguchi et al., 1999). Consequently, genomic occupancy of Spt5 is centered around transcription start sites (TSSs) (Pavri et al., 2010; Rahl et al., 2010). Structural studies of DSIF/RNAP complexes have shown Spt5 interacts with a region of RNAP very close to the transcription bubble (Klein et al., 2011a; Martinez-Rucobo et al., 2011), prompting the suggestion that Spt5 may tether AID onto RNAP II in

close proximity to ssDNA (Gazumyan et al., 2012). Interestingly, genome wide ChIP studies have also found AID concentrated near TSSs (Yamane et al., 2011). Co-localization of RNAP II, Spt5, and AID near TSSs has led several groups to propose an RNAP II stalling model of AID function whereby stalled transcription complexes provide a platform to recruit AID to ssDNA target sites (Gazumyan et al., 2012; Kenter, 2012; Storb, 2014; Sun et al., 2013b).

The elucidation of AID as a ssDNA specific deaminase reconciled well with the known role of transcription in hypermutation. One of the defining features of transcription is the unwinding of the DNA duplex during RNA synthesis within the transcription bubble formed by the elongating RNA polymerase (Gnatt et al., 2001). Within the transcription bubble the template DNA strand is buried within RNA polymerase and hybridized with the 3' most 8-9 nucleotides of nascent RNA (Nudler et al., 1997). In contrast, the non-template DNA strand is partially exposed in the context of the transcription bubble and is susceptible to enzymatic digestion (Korzheva et al., 2000; Wang and Landick, 1997). Therefore, the act of transcription can generate exposed ssDNA that could potentially allow access of AID to the DNA duplex. Indeed, using biochemical and genetic approaches, transcription was shown to create AID substrates largely restricted to the non-template DNA strand (Chaudhuri et al., 2003; Ramiro et al., 2003).

The mechanism of AID action within the *Ig* loci of B cells has turned out to be far more complex and involving many downstream events beyond the initial cytidine deamination. For instance, incorporation of *Ig* mutations at A:T base pairs and recombination of $C_H$ genes during CSR, are both AID dependent processes but cannot be explained solely through cytidine deamination. AID utilizes many cofactors during *Ig* diversification and their collective actions in the context of CSR are illustrated below.

CSR is the process leading to Ig isotype switch from IgM to IgG, IgE, or IgA. Each isotype is determined by the corresponding heavy chain $C_H$ gene. The mouse $C_H$

genes are contained within a ~200 kb locus downstream of the $J_H$ gene segments and organized as Cμ, Cδ, $Cγ_3$, $Cγ_1$, $Cγ_{2b}$, $Cγ_{2a}$, Cε, and Cα (Shimizu et al., 1982). In addition to serving as exons of the mature heavy chain mRNA arising from the rearranged $V_H$ promoter, each $C_H$ region can also function as a distinct transcriptional unit. Each $C_H$ gene is accompanied by its own promoter that can be induced during specific stimulation conditions leading to production of noncoding germline transcripts (Figure 2) (Chaudhuri et al., 2007). Germline transcription is important for directing CSR machinery to the $C_H$ genes that will participate in the isotype switch. This was hypothesized based on the observation that transformed B cell lines with a tendency to switch to a particular isotype also express germline transcripts of the same $C_H$ gene (Stavnezer-Nordgren and Sirlin, 1986; Yancopoulos et al., 1986). In much the same way that germline transcription of unrearranged V segments is thought to provide differential accessibility between TCR and IgH loci during V(D)J recombination, germline transcription of $C_H$ genes also provides differential accessibility between the different $C_H$ genes during CSR.

Immediately upstream of $C_H$ genes (except Cδ) lies a highly repetitive, 1-12 kb G-rich sequence referred to as the switch (S) region (Figure 2). $C_H$ germline transcription initiates with a short noncoding “I exon”, proceeds through the S region, and terminates downstream of the last $C_H$ coding exon (Lennon and Perry, 1985). The S region is entirely intronic and therefore absent from the mature $C_H$ germline transcript. Incidentally, splicing of germline transcript S region introns is critical for CSR as $C_H$ alleles containing mutant splice acceptor/donor sequences are highly defective for CSR (Hein et al., 1998; Lorenz et al., 1995), although the mechanistic underpinnings are poorly understood. For CSR to proceed, simultaneous transcription of donor and acceptor $C_H$ genes is required. The donor $C_H$ gene is invariantly Cμ and is constitutively transcribed in naive B cells (Li et al., 1994), in accordance with its role as the universal

donor $C_H$ gene during CSR. Transcription of downstream acceptor $C_H$ genes is induced through specific mitogen and/or cytokine signaling pathways (Keim et al., 2013).

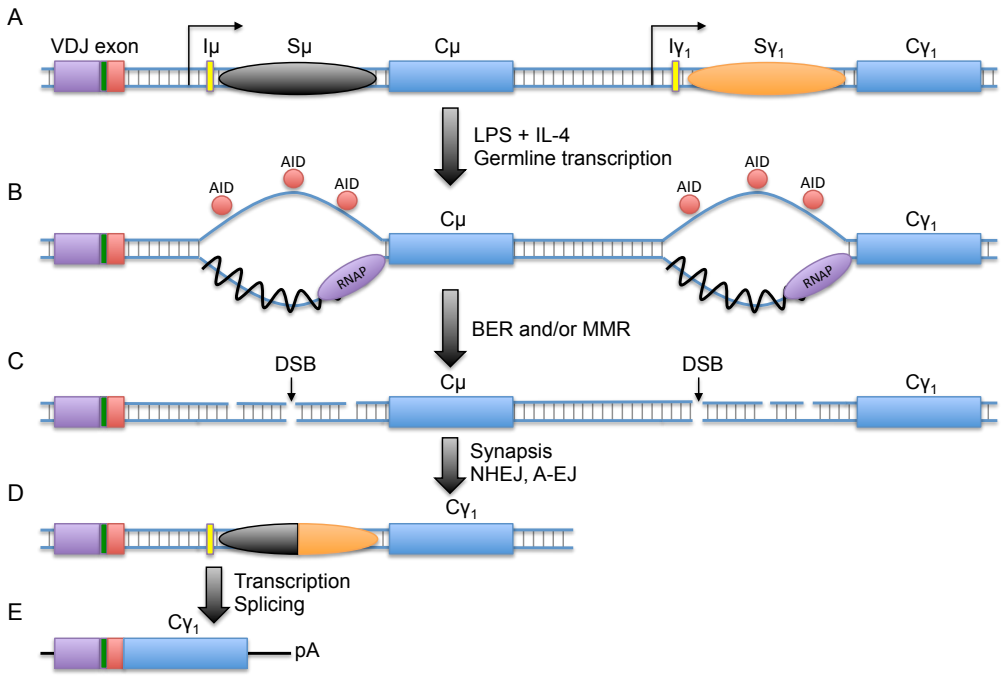



**Figure 2: Molecular overview of class switch recombination**
(A) Multiple $C_H$ genes (only Cμ and $C\gamma_1$ are illustrated here) lie downstream of the rearranged heavy chain variable region VDJ exon. Each $C_H$ gene contains its own cognate promoter. Cμ is constitutively expressed whereas downstream $C_H$ genes such as $C\gamma_1$ are controlled through inducible promoters. Upstream of each $C_H$ gene is a 1-12 kb highly repetitive, guanosine rich intronic switch sequence.
(B) Specific cytokine stimulation results in transcriptional activation of a downstream acceptor $C_H$ gene. Transcription through switch regions leads to DNA secondary structures that promote formation of ssDNA substrates of AID. AID carries out deamination of switch region deoxycytidine residues to deoxyuridine.
(C) Base excision and mismatch repair pathways process DNA incorporated deoxyuridine residues into ssDNA breaks. Closely apposed ssDNA breaks can lead to blunt ended or staggered DNA double strand breaks.
(D) DNA double strand breaks at donor and acceptor switch regions initiate a DNA damage response leading to synapsis and long-range end joining between switch regions, resulting in juxtaposition of the donor $C_H$ gene with the VDJ exon. The intervening Cμ containing DNA segment is excised as a circular DNA fragment (not illustrated).
(E) Transcription and splicing of heavy chain transcripts arising from the $V_H$ promoter lead to mRNA containing the same variable region exons but with new isotype switched $C_H$ exons.

S regions are required for CSR. Germline deletion of the donor Sμ region leads to a nearly complete block in CSR towards all isotypes (Khamlichi et al., 2004). The highly repetitive and G-rich nature of S regions makes them prone to co-transcriptional R-loop formation (Figure 2). Typically the nascent $C_H$ germline transcript will hybridize with the C-rich template strand forming a stable RNA:DNA hybrid leaving a displaced G-

rich non-template strand. R-loop accumulation at transcribed S regions was first observed in vitro where S region containing plasmids adopted RNase H sensitive conformers following transcription (Daniels and Lieber, 1995; Reaban and Griffin, 1990). Subsequent studies confirmed the presence of S region R-loops in chromosomes of activated B cells (Yu et al., 2003). Although the precise mechanism leading to formation of S region R-loops is unknown, nucleic acid thermodynamic properties are likely involved. For instance, the base composition of S region RNA:DNA hybrids is heavily biased towards base pairing between guanosine (rG) and deoxycytidine (dC). Physical studies with nucleic acids have demonstrated that rG:dC base pairs (as found in S region R-loops) are more stable than dG:dC base pairs present in duplex DNA (Roberts and Crothers, 1992). This may be an important factor in maintaining the S region RNA:DNA hybrid in the presence of continuous competition from the non-template strand. The importance of this asymmetric distribution of dG nucleotides between template and non-template strands at S region DNA has been confirmed in vivo. Inversion of $S\gamma_1$, such that the G-rich strand now serves as the template strand leads to a substantial, although not complete, defect in CSR to IgG1 (Shinkura et al., 2003). In addition, G-rich ssDNA such as the displaced S region non-template strand, can adopt stable intramolecular non-B form DNA secondary structures involving Hoogsteen base pairing (Bochman et al., 2012). For example, guanine can form a hydrogen bonded planar arrangement of 4 guanines known as a G-quartet. Stacking of G-quartets through the looping of ssDNA can create stable G-quadruplex structures. Therefore, involving multiple mechanisms, S regions appear to be inherently capable of promoting formation of co-transcriptional AID substrate ssDNA structures.

### 1.3.4 *Processing of AID generated DNA lesions during SHM and CSR*

However, R-loops cannot entirely account for generation of AID substrate ssDNA. First, variable region exons are not particularly G-rich on the non-template strand. Therefore AID mediated deamination during SHM is thought to occur through a mechanism that does not involve R-loops. Secondly, amphibian S regions display A:T richness, in contrast to G:C rich mammalian S regions (Mussmann et al., 1997). In vitro transcription of *Xenopus* Sμ does not produce R-loops, yet can partially substitute for murine $Sγ_1$ in vivo (Zarrin et al., 2004). Furthermore, in contrast to mammalian S regions, A:T rich Xenopus Sμ does not display an orientation dependence during CSR (Zarrin et al., 2004). These observations suggest that additional mechanisms beyond R-loop formation may be responsible for creating AID substrate ssDNA during CSR.

Replication protein A (RPA) was isolated through a biochemical screen aimed at identifying cofactors that promote AID access to transcribed DNA (Chaudhuri et al., 2004). RPA and AID together, but neither protein alone, can efficiently deaminate non-R-loop forming transcribed DNA in vitro. RPA is a heterotrimeric ssDNA binding protein involved in DNA replication and damage responses (Oakley and Patrick, 2010). Stretches of ssDNA as short as 4 nucleotides can physically interact with RPA (Matsunaga et al., 1996), consistent with a possible role in promoting AID function during SHM in addition to CSR. On the basis of these observations it was proposed that RPA stabilizes ssDNA structures during SHM and CSR, thereby promoting their access to AID.

The interaction between AID and RPA is controlled through serine 38 phosphorylation of AID by protein kinase A (PKA) (Basu et al., 2005; Basu et al., 2008; Pasqualucci et al., 2006). Although only a minor fraction of total AID exists in the phosphorylated state, this fraction is substantially enriched on chromatin (McBride et al., 2006). Furthermore, mutagenesis studies involving S38A knock-in mutants of AID have

confirmed that the integrity of the serine 38 residue is critical for SHM and CSR in vivo (Cheng et al., 2009; McBride et al., 2008). Recent studies suggest that phosphorylation of AID on chromatin may involve a positive feedback loop involving DNA breaks and CSR (Vuong et al., 2013). AID phosphorylation was found to be stimulated in response to DNA breaks. Likewise, AID phosphorylation enhances the rate of DNA break formation. Consistent with this model, PKA is recruited to S regions during CSR, resulting in RPA recruitment and AID phosphorylation (Vuong et al., 2009). Therefore, the post-translationally regulated interaction between RPA and AID provides an additional mechanism for transcription coupled ssDNA access to AID.

In addition to providing ssDNA substrates of AID, cellular factors are also involved in processing and amplification of the initial AID generated lesion. The biochemical reaction catalyzed by AID, deamination of deoxycytidine to deoxyuridine, cannot directly account for mutations observed at A:T base pairs during SHM and CSR. Likewise, the specific activity of AID alone will not lead to a DNA break, a necessary intermediate for CSR. AID action on DNA creates a damage signal that is recognized by cellular DNA repair machinery. Base excision repair (BER) and mismatch repair (MMR) are the two major processes that cooperate with AID (Di Noia and Neuberger, 2007). Paradoxically, during Ig diversification BER and MMR lead to error-prone DNA repair, thus enhancing the spectrum of mutations within *Ig* loci. Deoxyuridine (dU) present in chromosomal DNA can be detected by both BER and MMR. It is thought that BER activity leads to diversification of the dU containing base pair, whereas MMR can diversify neighboring base pairs as well (Figure 3). As a result, MMR activity is required for mutagenesis of non-AID targeted A:T base pairs.

Some AID generated dU:dG base pairs can escape BER or MMR and be carried into the S-phase of the cell cycle. Faithful DNA replication of dU:dG results in one daughter cell harboring a C>T or G>A transition, according to which DNA strand

contained the dU nucleotide (Figure 3). Therefore, DNA replication of AID deaminated nucleotides is one means of providing a limited source of mutation during SHM and CSR.

Non-native nucleotides incorporated in chromosomal DNA can be detected by DNA glycosylases involved in BER. dU is specifically recognized by uracil DNA glycosylase (UNG), which excises the uracil nitrogenous base from the deoxyribose-phosphate backbone leaving behind an abasic site (Figure 3). Such residues are non-informative and cannot be used as a template during faithful DNA replication. Instead, abasic sites can be bypassed by translesion DNA polymerases that can incorporate any of the four canonical deoxynucleotides onto the nascent chain (Figure 3). Abasic sites generated by UNG can also be excised from the DNA backbone by apurinic/apyrimidinic endonuclease 1 (APE1) creating a ssDNA break (Figure 3).

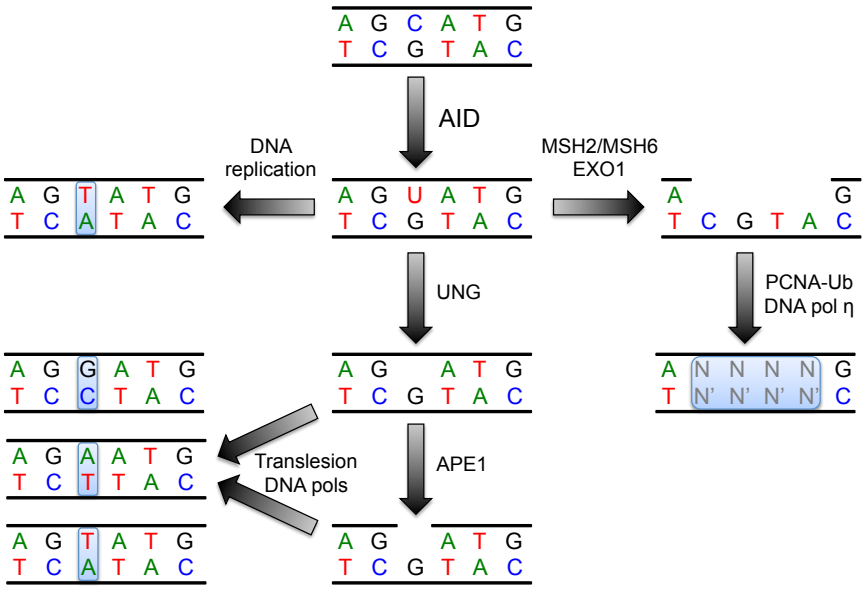



**Figure 3: Outline of downstream SHM mechanism involving BER and MMR pathways**
AID generated dU residues can be processed in 3 different ways. First, dU can be replicated by high fidelity DNA polymerases during S-phase to yield C>T transitions. Secondly, DNA incorporated dU can be recognized by the base excision repair enzyme uracil DNA glycosylase (UNG). UNG can excise the nitrogenous base from the dU nucleotide to produce an abasic site. The abasic residue can be removed by apurinic/apyrimidinic endonuclease 1 (APE1) to create a ssDNA break. Translesion DNA polymerases can bypass abasic templates by incorporating any deoxynucleotide. Thirdly, the dU:dG base pair can be recognized by the mismatch repair heterodimer MSH2/MSH6 and recruit exonuclease 1 (EXO1) to resect a short patch on the dU containing strand. MSH2/MSH6 can recruit mono-ubiquitylated PCNA and DNA polymerase η leading to error prone fill-in.

Consistent with a role for BER in Ig diversification, UNG deficiency leads to a severe defect in CSR and a stark alteration in the spectrum of mutations during SHM (Di Noia and Neuberger, 2002; Imai et al., 2003; Rada et al., 2002b). Transversion mutations at C:G base pairs produced during SHM are nearly absent when UNG activity is ablated. Instead, only transition mutations are observed at C:G base pairs, consistent with DNA replication over the unprocessed AID generated dU:dG base pair. Furthermore, the activity of UNG appears to be restricted to C:G base pairs as UNG deficiency does not alter the spectrum or frequency of mutations at A:T base pairs. Mutations in translesion DNA polymerases acting downstream of UNG also lead to alterations in the pattern of SHM. Rev1 is a translesion DNA polymerase possessing deoxycytidyl transferase activity that can bypass an abasic template residue by adding a dC nucleotide onto the nascent strand (Nelson et al., 1996). Accordingly, Rev1 deficient mice exhibit a strong reduction in C>G and G>C transversion frequency during SHM (Jansen et al., 2006). Also, APE1 deficiency was shown to impair CSR (Guikema et al., 2007; Masani et al., 2013).

The MMR pathway serves as a parallel mechanism of dU recognition and is critical for introducing mutations at A:T base pairs during SHM and CSR. Under normal cellular circumstances MMR helps preserve the integrity of genomic information by removing incorrect nucleotides from the newly synthesized strand following DNA replication. dU:dG base pairs are typically recognized by the MutSα heterodimer composed of MSH2 and MSH6 (Chahwan et al., 2012). During error-free MMR, MutSα recruits MutLα (composed of MLH1 and PMS2) to the mismatch where it then introduces a nick into the DNA strand containing the misincorporated nucleotide. The nick is then resected by exonuclease 1 (EXO1) leaving behind a short single stranded gap in the DNA that is subsequently filled in by DNA polymerase δ (Constantin et al., 2005).

SHM employs an error-prone version of MMR. While the AID generated dU:dG base pair is recognized by MSH2/MSH6, current evidence suggests that the MLH1/PMS2 nickase fails to be recruited (Chahwan et al., 2012). The mechanism behind this exclusion is poorly understood. However, evidence does exist for a nick occurring during SHM because EXO1 can still create a single stranded gap downstream of MSH2/MSH6 recruitment (Bardwell et al., 2004). Incidentally, the nick used by EXO1 is likely not provided by the BER enzyme APE1, as UNG deficient mice exhibit normal SHM at A:T base pairs. MSH2/MSH6 can recruit mono-ubiquitylated PCNA and DNA polymerase η to the dU:dG base pair (Chahwan et al., 2012; Wilson et al., 2005). DNA polymerase η exhibits a high mutation rate at A:T base pairs (Matsuda et al., 2001).

The importance of MMR in SHM was gleaned from the phenotypes of MSH2 and MSH6 deficient mice. B cells from *Msh2*$^{-/-}$ mice exhibit reduced overall SHM, with an exacerbation at A:T base pairs (Frey et al., 1998; Rada et al., 1998). The SHM pattern of *Msh6*$^{-/-}$ mice phenocopies that of *Msh2*$^{-/-}$ (Wiesendanger et al., 2000). Mutations at A:T base pairs are reduced by up to 90% in *Msh2* and *Msh6* null B cells, indicating that the vast majority of mutations at these residues are introduced through the MMR pathway. The remaining ~10% of mutations at A:T base pairs in *Msh2* and *Msh6* null B cells are very likely mediated through BER. When both BER and MMR pathways were ablated in *Ung*$^{-/-}$ *Msh2*$^{-/-}$ double deficient mice A:T mutations were entirely eliminated (Rada et al., 2004). In fact, only C>T and G>A transition mutations were observed at *Ig* loci in *Ung*$^{-/-}$ *Msh2*$^{-/-}$ mice. This strongly implies these mutations arise from S-phase DNA replication over all AID generated dU:dG lesions in *Ung*$^{-/-}$ *Msh2*$^{-/-}$ B cells, thereby providing a genetic footprint of AID action. K164 mono-ubiquitylated PCNA cooperates with MSH2/MSH6 to recruit error-prone DNA polymerase η to the EXO1 processed ssDNA gap near the dU:dG base pair. *Pcna*$^{K164R/K164R}$ B cells incapable of PCNA mono-ubiquitylation are strongly impaired in A:T mutagenesis (Arakawa et al., 2006; Langerak

et al., 2007). Likewise, EXO1 or DNA polymerase η deficiency also results in a nearly complete loss of mutations at A:T base pairs (Bardwell et al., 2004; Delbos et al., 2007; Masuda et al., 2007).

### 1.3.5 *DNA damage response and end joining during CSR*

As CSR involves DNA rearrangement of $C_H$ genes (Kataoka et al., 1980), a key event is formation of DNA double strand breaks (DSBs) at donor and acceptor S regions. Following the discovery of AID's critical role in CSR, it was quickly discovered that AID was required for generation of DSBs at S regions and that this initiated a DNA damage response (Petersen et al., 2001). As outlined above, the processing of AID generated dU:dG base pairs in S regions through the actions of APE1 and/or EXO1 can introduce ssDNA breaks during CSR. Current models propose a stochastic mechanism for DSB formation during CSR whereby temporally apposed ssDNA breaks on both strands leads to a DSB. This is supported by the observation that S region staggered DSBs vastly outnumber blunt DSBs produced during CSR (Rush et al., 2004). DSBs generated at S regions are recognized by cellular components of the DNA damage response that coordinate with end joining pathways. This process leads to ligation of broken donor and acceptor S regions and deletion of Cμ and other intervening $C_H$ genes.

One of the earliest events occurring after a DSB is binding of the Mre11-Rad50-Nbs1 (MRN) complex to either end of the DSB (Carney et al., 1998). MRN tethering to the DSB recruits ATM kinase leading to its activation (Lee and Paull, 2005). ATM is a key component of the DNA damage response. It phosphorylates the p53 tumor suppressor, thus inducing the p53-dependent G1/S checkpoint to allow for sufficient time to repair the break before initiation of DNA replication (Boboila et al., 2012). ATM also phosphorylates histone variant H2AX (known as γH2AX) (Burma et al., 2001) and

histone binding protein 53BP1 (Anderson et al., 2001; Rappold et al., 2001) leading to assembly of large DNA damage associated chromatin foci proximal to the DSB. These foci can exceed 2 Mb in DNA length and are thought to serve as scaffolds to assist in recruitment of DNA repair factors (Boboila et al., 2012). Indeed, CSR is impaired to varying degrees in B cells from mice deficient for various DNA damage response factors. *Nbn* null B cells (Nbs1 deficient) are mildly impaired in CSR, displaying a ~50% reduction in CSR efficiency (Kracker et al., 2005; Reina-San-Martin et al., 2005). A slightly greater impairment in CSR is observed in either *Atm* null (Lumsden et al., 2004; Reina-San-Martin et al., 2004) or *H2afx* null (H2AX deficient) B cells (Petersen et al., 2001; Reina-San-Martin et al., 2003). In contrast to other DNA damage response factors, *Trp53bp1* gene deletion (53BP1 deficient) leads to a severe CSR phenotype. B cells from *Trp53bp1*$^{-/-}$ mice display approximately 5% CSR efficiency relative to wild type controls (Manis et al., 2004; Ward et al., 2004). This indicates a potential role for 53BP1 in CSR beyond its recognized function in DNA damage response.

Sequence analysis of S region joins from isotype switched B cells showed that the majority of CSR breaks are joined with little or no homology (Dunnick et al., 1993). This is consistent with S region DSBs being joined primarily via the classical non-homologous end joining (C-NHEJ) pathway, as C-NHEJ can join blunt or staggered DNA ends containing 1-4 nucleotides of microhomology (Boboila et al., 2012). Likewise, CSR is impaired in mature B cells from C-NHEJ deficient mice expressing rearranged heavy and light chain “knock-in” alleles. However, unlike V(D)J recombination, which is strictly dependent on C-NHEJ, a significant CSR activity remains in mature B cells deficient in core C-NHEJ factors. Xrcc4 and DNA ligase IV are two such factors involved in the final end joining step of C-NHEJ. Consistent with a role for C-NHEJ in CSR end joining, *Xrcc4* or *Lig4* deficient mature B cells display isotype switching impairment (Yan et al., 2007). Surprisingly, the level of CSR in *Xrcc4* or *Lig4* deficient mature B cells ranged between

25-50% relative to control B cells, demonstrating the presence of a robust alternative end joining mechanism driving CSR in the absence of C-NHEJ (Yan et al., 2007). Early reports on CSR in the absence of Ku70 or Ku80 suggested that CSR was entirely abrogated in the absence of either Ku component (Casellas et al., 1998; Manis et al., 1998). However, a reexamination of the role of Ku in CSR using more robust B cell stimulation conditions showed that relative to control cells, up to 50% of CSR efficiency could be observed in Ku deficient mature B cells (Boboila et al., 2010). Indeed, analysis of S region joins produced in the absence of Xrcc4, Lig4, or Ku70 indicated a bias towards usage of longer microhomologies in the joining of S region DSBs (Boboila et al., 2010; Yan et al., 2007). Collectively, current evidence supports the notion that end joining of S region DSBs is primarily mediated through C-NHEJ with a secondary, possibly backup pathway, operating through a microhomology annealing mechanism.

Another important feature of DSB processing during CSR is the synapsis of broken S regions. This is of critical importance because in its absence chromosome loss or translocations could occur. γH2AX and 53BP1 that accumulate over long stretches of chromatin proximal to DSBs are implicated in synapsis of broken S regions. As mentioned previously, CSR is impaired in the absence of H2AX and 53BP1. However, S region breaks during CSR can be joined either within the same S region, leading to an intra-switch deletion (ISD), or between different S regions, and leading to isotype switching. Interestingly, while H2AX and 53BP1 deficiency impairs long-range CSR, it does not reduce the rate of local ISD (Reina-San-Martin et al., 2007; Reina-San-Martin et al., 2003). This suggests that γH2AX and 53BP1 are involved in promoting long-range end joining during CSR through the synapsis of broken S regions. Another important function of synapsis may be to suppress resection of broken DNA ends, to support C-NHEJ. Indeed, H2AX and 53BP1 are required to suppress end resection of broken S regions during CSR (Bothmer et al., 2011; Bothmer et al., 2010). Consistent with this

notion, accumulation of the ssDNA binding protein RPA has been observed at S regions in H2AX or 53BP1 deficient B cells during CSR (Yamane et al., 2013).

### 1.4 AID activity beyond *Ig* loci

AID's mutagenic activity poses a significant threat to the genomic integrity of B cells. Although B cells have developed a remarkable ability to target AID activity to the *Ig* loci, the rest of the genome is not entirely spared. In fact, the notion of aberrant SHM introducing mutations beyond *Ig* genes predates the discovery of AID. One of the earliest examples implicating SHM as a mechanism responsible for introducing mutations in bystander genes was the discovery of *BCL6* mutations in germinal center B cells (Pasqualucci et al., 1998; Shen et al., 1998). Sequence analysis of *BCL6* in human B cell malignancies of various developmental origins indicated that *BCL6* mutations are largely restricted to tumors of germinal center and post-germinal center origin (Pasqualucci et al., 1998). Subsequently, additional genes including *MYC*, *PIM1*, *PAX5*, and *CD79B* were found to contain mutations in a large percentage of germinal center derived diffuse large B-cell lymphoma samples (Gordon et al., 2003; Pasqualucci et al., 2001). A relatively broader approach towards identifying the extent of AID generated mutations across the B cell genome later provided evidence that AID activity was far less restricted than previously thought. 1 kb segments of DNA downstream of the major TSS were systematically sequenced for 118 expressed genes from mouse germinal center B cells (Liu et al., 2008). Approximately 25% of assessed genes accumulated statistically significant levels of AID dependent mutations (Liu et al., 2008). When this same analysis was performed in the DNA repair deficient *Ung*$^{-/-}$ *Msh2*$^{-/-}$ background, up to half of the genes analyzed contained a significantly higher mutation load compared to wild type cells (Liu et al., 2008). This indicates that although AID activity is widespread, it is frequently countered by high fidelity DNA repair at regions outside *Ig* loci. Curiously, this

is the opposite phenomenon of what is observed at *Ig* loci, where DNA repair deficiency results in impairment of SHM. A very small number of assayed genes that were significantly mutated showed nearly identical mutation frequencies in wild type and *Ung*$^{-/-}$ *Msh2*$^{-/-}$ background (Liu et al., 2008), suggesting that DNA repair may be excluded at certain loci targeted by AID. This rare behavior was observed at *Bcl6* and may have contributed to it being the first identified AID "off-target".

In addition to point mutations arising from aberrant SHM, human B cell lymphomas frequently harbor chromosomal translocations involving *Ig* variable and S regions (Nussenzweig and Nussenzweig, 2010). Many of these translocations involve improper DSB resolution arising from RAG and AID mediated Ig diversification processes. A common consequence is the juxtaposition of strong *Ig* transcriptional regulatory elements onto proto-oncogene containing loci. This frequently leads to deregulation and/or overexpression of the participating proto-oncogene, resulting in cellular transformation. The earliest molecular characterization of a translocation involving *Ig* loci and a proto-oncogene was the discovery of reciprocal chromosomal rearrangements involving *MYC* and *IGH* in human Burkitt's lymphoma (Dalla-Favera et al., 1982; Taub et al., 1982). Determining the role of AID in the genesis of *Myc/IgH* translocations was greatly assisted by the development of mouse models that could reproducibly give rise to this particular DNA rearrangement (Potter and Wiener, 1992). Specifically, using an IL-6 transgenic mouse model of plasmacytoma formation, *Myc/IgH* translocations were reported to be entirely AID dependent (Ramiro et al., 2004). Subsequent studies went on to show that Ung, a BER protein critical for S region DSB formation during CSR, was also required for *Myc/IgH* translocation (Ramiro et al., 2006). Furthermore, AID was specifically found to be required for creation of DSBs at both *Myc* and *IgH* (Robbiani et al., 2008), indicating a direct involvement of AID in *Myc/IgH* translocation.

In addition to *Myc*, AID was found to generate DSBs at many non-*Ig* loci across the B cell genome (Robbiani et al., 2009; Staszewski et al., 2011). In recent years, two independent groups have reported their attempts to catalog the extent of genome-wide AID dependent DSBs and translocations using a novel deep sequencing approach (Chiarle et al., 2011; Klein et al., 2011b). Briefly, I-SceI meganuclease recognition sequences were engineered into either *Myc* or *IgH* loci (Robbiani et al., 2008; Wang et al., 2009a; Zarrin et al., 2007). B cells harboring either of these modified alleles, on *Aicda*$^{WT/WT}$ or *Aicda*$^{-/-}$ backgrounds, were stimulated ex vivo to induce AID expression and subsequently transduced with I-SceI encoding retroviruses (Chiarle et al., 2011; Klein et al., 2011b). Under these conditions, one artificial DSB is intentionally introduced in either *Myc* or *IgH* through I-SceI cleavage of its target sequence, while undefined DSBs are generated in other parts of the genome through AID mediated mechanisms. I-SceI induced DSBs resolved through translocations ultimately produce DNA rearrangements containing known and unknown sequences on either side of the junction. Adaptor PCR techniques enabled these junctions to be identified through deep sequencing approaches. Importantly, these experiments involved short-term culture conditions to allow for unbiased identification of rare translocations and to prevent the selective outgrowth of oncogenic translocations. These experiments identified genomic sites of recurrent AID dependent DSBs and translocations, which included many genes frequently translocated in various human B cell lymphomas (Chiarle et al., 2011; Klein et al., 2011b).

Genome mapping of these translocation sites uncovered several themes underlying the nature of recurrent B cell translocations. First, translocation frequency is vastly greater within a few kilobases of the I-SceI site, although this phenomenon was observed in both *Aicda*$^{WT/WT}$ and *Aicda*$^{-/-}$ backgrounds, indicating that proximity is an important factor in DSB resolution in the absence of AID (Chiarle et al., 2011; Klein et

al., 2011b). Furthermore, even over longer distances such as hundreds of kilobases from the I-SceI site, intra-chromosomal rearrangement was strongly preferred (Chiarle et al., 2011; Klein et al., 2011b), suggesting that nuclear architecture and chromosomal boundaries might be involved in DSB ligation.

AID dependent translocation junctions frequently mapped to transcribed genic regions (Chiarle et al., 2011; Klein et al., 2011b). Specifically, AID dependent translocations were significantly enriched near TSS regions, often within 2 kb of the TSS (Chiarle et al., 2011; Klein et al., 2011b). This is consistent with the notion of TSSs as areas of RNAP II accumulation and ssDNA generation, which would presumably be enriched in AID substrates. Surprisingly, transcription per se was insufficient to produce a recurrent AID dependent translocation site as multiple highly transcribed genes did not participate in translocations (Chiarle et al., 2011; Klein et al., 2011b). This suggests that other transcription coupled processes may be critically involved in the generation of AID dependent DSBs leading to translocations. More recent studies have begun to investigate variables involved in recurrent AID dependent translocations. Using chromosome conformation capture experiments and genome-wide RPA occupancy as readouts of genomic contact frequency and DNA damage, respectively, the relative contributions of physical proximity and AID generated DNA damage were assessed at recurrent translocation sites (Hakim et al., 2012). This study concluded that in the absence of AID, proximity and nuclear organization are important factors in determining translocation frequency (Hakim et al., 2012). In contrast, DNA damage resulting from AID activity correlates more strongly with translocation frequency than physical contact frequency (Hakim et al., 2012). Therefore, while both DNA damage and proximity are important variables in the generation of translocations, lesions generated by AID activity appear to be the driving force behind recurrent B cell translocations.

## 1.5 The conundrum of AID activity on both strands of DNA

A requisite of AID dependent CSR and chromosomal translocations is the creation of DNA breaks on both strands to produce a DSB. Statistical analyses of hundreds of antibody sequences provided early evidence that the SHM process is targeted to both DNA strands. A comparison of all triplet sequences with their respective inverse complements revealed similar mutation frequencies at the middle nucleotides and also between the first and third nucleotides (Milstein et al., 1998). Following the discovery of AID, mutational analysis of *Ig* loci under DNA repair deficient conditions that genetically reveal the AID deamination footprint, it was confirmed that AID activity is targeted nearly equivalently to both DNA strands (Rada et al., 2004; Xue et al., 2006). However, AID deaminase activity is specific for ssDNA, and currently proposed models of AID activity imply that each strand of the DNA duplex is targeted by individual AID complexes. Furthermore, transcription coupled AID deamination assays performed in vitro mainly result in deamination of the non-template strand (Basu et al., 2005; Chaudhuri et al., 2004). Ectopic expression of AID in bacteria or yeast also results in a strong bias towards non-template strand mutations (Gomez-Gonzalez and Aguilera, 2007; Ramiro et al., 2003). Restriction of mutations to the non-template strand in these assays is likely due to occlusion of the template strand by the elongating RNAP complex. Furthermore, such assays are likely to not fully recapitulate all aspects of AID function as they occur in B cells. A major question therefore arises. How is AID activity targeted to the template strand in the physiological contexts of SHM and CSR?

A biochemical screen was performed in an attempt to identify potentially uncharacterized cellular factors that may promote AID access to the template strand during transcription coupled deamination (Basu et al., 2011). Namely, SHM substrate duplex DNA was transcribed in vitro in the presence of AID containing cellular extracts from the human B cell lymphoma Ramos cell line (Basu et al., 2011). A series of

chromatography steps including size exclusion, charge, density, and affinity purification were utilized to enrich for macromolecular complexes associated with DNA and AID. Importantly, fractions from each chromatography step were assayed for deamination activity to enrich for AID stimulation activity in subsequent purification steps (Basu et al., 2011). Final affinity purification fractions eluted from an anti-AID column contained an activity that could stimulate AID deamination by ~100 fold in transcription coupled in vitro assays (Basu et al., 2011). Using mass spectrometry multiple subunits of the RNA exosome complex were identified in the AID stimulating fractions (Basu et al., 2011). Physical interaction between AID and RNA exosome was verified in primary mouse B cells stimulated for CSR as well as human and mouse B lymphoma cell lines through co-immunoprecipitation experiments (Basu et al., 2011).

Functional consequences of RNA exosome association with AID were explored in the CH12F3 mouse B lymphoma cell line. These cells are capable of AID induction and directed class switching to IgA upon stimulation with anti-CD40, IL-4, and TGFβ (Nakamura et al., 1996). shRNA mediated knockdown of the Rrp40 subunit of RNA exosome in CH12F3 cells reduced CSR efficiency between 50-70% (Basu et al., 2011). Consistent with a role in CSR, Rrp40 recruitment to IgH Sμ and Sα regions was induced in stimulated CH12F3 cells (Basu et al., 2011). A similar pattern of RNA exosome recruitment was observed at Sμ and $S\gamma_1$ in primary mouse B cells stimulated towards IgG1 class switching (Basu et al., 2011). Strikingly, RNA exosome recruitment to S regions during CSR was found to be AID dependent, as Rrp40 occupancy at Sμ and $S\gamma_1$ was not observed in *Aicda*$^{-/-}$ (AID deficient) B cells using identical stimulation conditions (Basu et al., 2011). This indicates that germline transcription of S regions is insufficient to recruit RNA exosome and that AID likely functions genetically upstream of RNA exosome during CSR. Finally, strand specific in vitro transcribed AID deamination assays indicated that RNA exosome, purified from either 293T human embryonic kidney

cells or reconstituted from recombinant bacterially expressed and purified subunits (Greimann and Lima, 2008), could direct AID activity towards both template and non-template strands of multiple transcribed SHM substrates (Basu et al., 2011). Collectively, these data strongly implicate RNA exosome as the missing factor involved in directing AID activity to both strands of the DNA duplex during transcription coupled SHM and CSR at *Ig* loci.

### 1.6 RNA exosome structure and function

The RNA exosome is an ancient molecular complex whose existence can be traced throughout evolution. Its general structure and catalytic activity have been conserved across all kingdoms of life, although some of the mechanistic details have diverged. All RNA exosome related complexes studied to date possess 3'-5' exoribonucleolytic activity and are comprised of multiple subunits that assemble into a toroid quaternary structure (Lykke-Andersen et al., 2009). As such, RNA exosome complexes possess a characteristic central channel traversing the length of the complex, through which single stranded RNA is funneled.

Eukaryotic RNA exosome complexes contain either 10 or 11 distinct subunits (Allmang et al., 1999b). The structure's core is made up of 9 subunits arranged in 2 layers of rings. The larger ring is a heterohexamer composed of individual copies of Rrp41, Rrp42, Rrp43, Rrp45, Rrp46, and Mtr3 (Januszyk and Lima, 2014). Each of these subunits is homologous to bacterial RNase PH and thereby contains the characteristic RNase PH domain. Intersubunit interactions involving the RNase PH domains allow these 6 subunits to assemble into a toroid. In contrast to archaeal RNA exosome (Lorentzen et al., 2005), the hexamer ring of the eukaryotic RNA exosome is unstable on its own and requires 3 additional subunits to stabilize the complex (Liu et al., 2006). Rrp40, Rrp4, and Csl4 form a secondary ring that sits on "top" of the hexamer

and function as a cap, holding the hexamer together. Each of the 3 cap subunits make extensive contacts with a different pair of hexamer subunits (Liu et al., 2006) resulting in a stabilized 9 subunit RNA exosome core structure. In addition, each cap subunit contains a pair of RNA binding domains that are important during substrate binding (Chlebowski et al., 2013). Namely, Rrp40 and Rrrp4 possess S1 and KH RNA binding domains, whereas Csl4 contains a KH and zinc ribbon domain.

A distant structural homolog of the eukaryotic RNA exosome is the bacterial enzyme polynucleotide phosphorylase (PNPase). Similar to RNA exosome, bacterial PNPase contains 3'-5' exoribonucleolytic activity and adopts a toroid structure (Symmons et al., 2000). However, the PNPase structure is composed of one repeating unit that assembles into a homotrimer. Each PNPase subunit contains two RNase PH domains (analogous to the RNA exosome hexamer) and a set of S1 and KH RNA binding domains (analogous to the RNA exosome cap) (Symmons et al., 2000).

Archaeal RNA exosome complexes are structured similarly to bacterial PNPase and eukaryotic RNA exosomes in that they assemble into a toroid and contain 3'-5' exoribonucleolytic activity. However, the RNA exosome complex of archaea contains an intermediate degree of compositional complexity. The 6 RNase PH domains that collectively go into formation of the toroid are contributed via 3 copies of the Rrp41/Rrp42 heterodimer that form the archaeal RNA exosome hexamer (Lorentzen et al., 2005). Similar to eukaryotes, archaea possess distinct cap subunits that interact with and help further stabilize the hexamer. Although only Rrp4 and Csl4 cap subunits have been observed in archaea, the stoichiometry between cap and hexamer subunits is conserved. As such, 3 copies of Rrp4 and/or Csl4 interact with the archaeal hexamer (Buttner et al., 2005; Lorentzen and Conti, 2005). Due to the symmetry of the archaeal hexamer imposed by the trimer of Rrp41/Rrp42 dimers, multiple combinations of Rrp4 and Csl4 are present in archaeal RNA exosome complexes (Chlebowski et al., 2013).

The relevance of this structural diversity in archaeal RNA exosome complexes is poorly understood.

In addition to evolutionary divergence in RNA exosome structural composition, a dramatic divergence in the mechanism of catalysis has also occurred. Bacterial PNPase and archaeal RNA exosome contain active sites that project into the lumen of their respective central channels (Chlebowski et al., 2013). In stark contrast, the 9 subunit eukaryotic RNA exosome core is catalytically inert (Dziembowski et al., 2007), largely because of non-synonymous substitutions at critical active site residues. Nucleolytic activity is imparted on the eukaryotic RNA exosome core in trans through association with active nuclease subunits Rrp6 and Dis3 (known as Rrp44 in yeast). Rrp6 is a distributive 3'-5' exoribonuclease (Januszyk et al., 2011), whereas Dis3 contains both processive 3'-5' exo- (Lorentzen et al., 2008) and endoribonuclease activities contained within distinct domains (Lebreton et al., 2008; Schaeffer et al., 2009). Therefore, unlike distantly related complexes, the eukaryotic RNA exosome has evolved to possess 3 distinct nucleolytic activities through association of complimentary nuclease subunits with the catalytically inert RNA exosome core. Furthermore, Rrp6 and Dis3 binding to the RNA exosome core does not appear to be positively or negatively codependent. As such, different RNA exosome isoforms can exist containing Rrp6 alone, Dis3 alone, or together with Rrp6 and Dis3 (Tomecki et al., 2010). Cellular localization studies have shown that different RNA exosome isoforms are concentrated in distinct subcellular compartments. Specifically, human RNA exosome exists in the nucleus as an 11 subunit holoenzyme containing both Rrp6 and Dis3 nucleases. Nucleolar and cytoplasmic RNA exosome complexes contain 10 subunits but differ in their nuclease composition. Cytoplasmic human RNA exosome complexes are associated with Dis3L, a homolog of Dis3 that contains 3'-5' exoribonuclease activity but lacks endonuclease activity (Staals et al., 2010). A distinct RNA exosome isoform also exists in the nucleolus that contains

Rrp6 but lacks a Dis3 subunit. These subcellular compositional differences in nuclease subunits presumably result in compartmentalization of distinct nucleolytic activities within the cell. Nuclear RNA exosome complexes, containing both types of nucleases, exhibit processive and distributive 3'-5' exoribonuclease activity as well as endonuclease activity. Cytoplasmic and nucleolar complexes lack endonuclease activity but possess processive and distributive 3'-5' exoribonuclease activities, respectively. How the cell orchestrates formation of distinct RNA exosome isoforms within different subcellular compartments is entirely unknown.

Recent biochemical studies on RNA exosome complexes incorporating active site mutants of Rrp6 and Dis3 have shed light on the catalytic mechanisms performed by the holoenzyme. Importantly, Rrp6 and Dis3 were found to influence the activity of one another. Rrp6 exonuclease activity is severely impaired within 11 subunit holoenzymes reconstituted with exonuclease dead Dis3 (Wasmuth and Lima, 2012). Likewise, Rrp6 robustly accelerates the activity of Dis3, as 10 subunit RNA exosome complexes containing Dis3 are far less active when compared to 11 subunit complexes containing both nuclease subunits (Wasmuth and Lima, 2012). Surprisingly, the ability of Rrp6 to stimulate Dis3 within the 11 subunit holoenzyme is also observed with exonuclease dead mutants of Rrp6. These observations provided evidence for a functional interaction between Rrp6 and Dis3. Further studies uncovered a critical role for the catalytically inert RNA exosome central channel in modulating Rrp6 and Dis3 function within the complex. RNA exosome channel mutants containing various length amino acid insertions predicted to project into and occlude the central channel were used to elucidate the influence of the RNA exosome core on nuclease subunit function. Dis3 activity is strongly impaired within these channel occlusion mutants (Wasmuth and Lima, 2012). As mentioned previously, Rrp6 activity is poisoned by exonuclease dead Dis3 within the 11 subunit complex. However, Rrp6 function in the presence of exonuclease dead Dis3

can be partially restored by occlusion of the RNA exosome central channel (Wasmuth and Lima, 2012). In fact the activity of Rrp6 in the context of channel occlusion appears to be largely independent of Dis3 function. Furthermore, RNA crosslinking studies indicate that channel occlusion preferentially diverts the RNA path away from Dis3 and towards Rrp6 (Wasmuth and Lima, 2012). Collectively, these results suggest the existence of alternate, but mutually exclusive, RNA paths within the RNA exosome. The dominant path appears to traverse the entire length of the central channel. Here, RNA 3' ends enter the RNA exosome central channel from one end and exit from the opposite end where they interact with Dis3. This path presumably blocks an alternate path leading to Rrp6. A recent structural analysis of an RNA exosome core complex bound to Rrp6 and RNA has revealed a widening of the Rrp40-Rrp4-Csl4 entry ring when bound to Rrp6, presumably providing RNA greater access through the central channel towards Dis3 in the presence of Rrp6 (Wasmuth et al., 2014). Analysis of the RNA exosome core crystal structure has uncovered 4 potential side channels that largely bypass the central channel and may provide direct access of RNA to Rrp6 (Wasmuth and Lima, 2012). Therefore, it appears that the eukaryotic RNA exosome core has evolved to modulate the activities of associated nucleases.

Initial structural studies performed on partial RNA exosome preparations revealed that Rrp6 and Dis3 bind to opposite ends of the complex. Cryo-electron microscopy of 10 subunit RNA exosome complexes containing Rrp6 alone suggested that Rrp6 binds near the "top" of the complex, adjacent to the entry pore Rrp40-Rrp4-Csl4 cap (Cristodero et al., 2008). X-ray crystallography of Dis3 bound to the Rrp41-Rrp45 fragment of the RNA exosome hexamer ring indicated that Dis3 lies on the "bottom" of the complex near the exit pore (Bonneau et al., 2009). More recently, the crystal structure of an RNA bound 11 subunit RNA exosome complex was reported (Makino et al., 2013). Consistent with previous results, Rrp6 and Dis3 are located near

the entry and exit pores, respectively. Furthermore, the structure reveals the long RNA path that traverses the length of the central channel leading to Dis3. Approximately 25 nucleotides of single stranded RNA is needed to reach the exonuclease site of Dis3 (Makino et al., 2013). In addition, Dis3 adopts a strikingly different conformation in the RNA bound 11 subunit complex (Makino et al., 2013) in comparison to the RNA free Dis3-Rrp41-Rrp45 structure (Bonneau et al., 2009). Specifically, the Dis3 exonuclease domain exhibits a ~90 degree rotation between the two structures (Bonneau et al., 2009; Makino et al., 2013). In the RNA bound structure, the exonuclease domain of Dis3 aligns with the exit pore in a position that allows it to capture the RNA 3' end (Makino et al., 2013). However, recent cryo-electron microscopy studies on Dis3 containing RNA exosome complexes suggest that RNA can also directly access the Dis3 exonuclease domain. RNA molecules that are not long enough to traverse the central channel can still bind the exonuclease domain of Dis3, but do not generate the conformational change in Dis3 associated with RNA threading through the central channel (Liu et al., 2014).

Collectively, these structural studies suggest that RNA exosome utilizes a ratcheting like mechanism of catalysis during RNA degradation. RNA passing through the central channel is captured by Dis3 due to a conformational change in the exonuclease domain. It has been proposed that following RNA binding, "breathing" of the Dis3 exonuclease domain between the apo- and RNA-bound forms may generate a pulling force on the RNA resulting in processive nucleolysis (Liu et al., 2014). Further studies will undoubtedly reveal additional mechanistic insights involving the multitude of interactions between RNA and the RNA exosome. Currently, no less than 3 different RNA paths have been described. RNA substrates can be threaded through the entire length of the central channel to reach Dis3 (Makino et al., 2013). RNA can pass through the entry pore but be diverted to Rrp6 through side channels between the cap and hexamer ring, thereby bypassing the exit pore (Wasmuth and Lima, 2012). Lastly, short

RNA can bypass the RNA exosome core entirely, through interactions with the Dis3 exonuclease domain in the apo state (Liu et al., 2014). It is likely that these distinct RNA paths are involved in RNA substrate discrimination and processing fates.

The RNA exosome is responsible for degradation and/or processing of a wide array of cellular transcripts. In fact prior to the discovery of the RNA exosome complex, the Rrp4 subunit was cloned on the basis of its ability to complement a temperature sensitive 5.8S ribosomal RNA (rRNA) 3' end processing defect in yeast (Mitchell et al., 1996). Subsequent fractionation and purification studies revealed that Rrp4 is a component of a much larger 3'-5' exoribonuclease complex involved in 3' end maturation of 5.8S rRNA from 7S rRNA precursors (Mitchell et al., 1997). Loss of function mutations in either RNA exosome nuclease subunit, Rrp6 or Dis3, leads to readily detectable accumulation of 7S rRNA precursors in yeast (Briggs et al., 1998; Dziembowski et al., 2007). More recently, shRNA mediated knockdown of multiple RNA exosome subunits in human HeLa cell lines has confirmed the involvement of RNA exosome in 5.8S rRNA 3' end maturation (Tafforeau et al., 2013). In contrast to yeast studies, RNA exosome shRNA knockdown in HeLa cells has provided evidence that RNA exosome, in particular the Rrp6 nuclease subunit, may be involved in a late 3' end processing step during human 18S rRNA maturation (Tafforeau et al., 2013).

In addition to rRNA, RNA exosome is also involved in processing of yeast small nucleolar RNA (snoRNA) and small nuclear RNA (snRNA) precursors (Allmang et al., 1999a; van Hoof et al., 2000). These findings were confirmed in genome-wide transcriptome analyses of Rrp6 and Dis3 mutant yeast. Specifically, Rrp6 or Dis3 depletion leads to widespread accumulation of snoRNA and snRNA precursors (Gudipati et al., 2012). However, this processing appears to be much more dependent on Rrp6 than Dis3 as accumulation of pre-snoRNA and pre-snRNA is far greater in Rrp6 versus

Dis3 single mutants, and is not exacerbated upon depletion of both Rrp6 and Dis3 compared to Rrp6 alone (Gudipati et al., 2012).

Mature mRNAs are not robustly targeted by RNA exosome. Approximately 10% of mRNAs exhibit significant differential accumulation between wild type and Rrp6 or Dis3 mutant yeast (Gudipati et al., 2012). However, intron containing pre-mRNAs can serve as substrates of RNA exosome (Bousquet-Antonelli et al., 2000). In genome-wide studies, up to half of all intronic regions display elevated read counts in Dis3 mutants owing to accumulation of pre-mRNAs rather than spliced introns (Gudipati et al., 2012). While Rrp6 deletion did not significantly alter intronic read counts (Gudipati et al., 2012), RNA-protein crosslinking studies indicate significant and largely similar preferential binding of Rrp6 and Dis3 to intron-exon junctions (pre-mRNA) compared to exon-exon junctions (spliced mRNA) (Schneider et al., 2012). Thus, the specific roles of Rrp6 and Dis3 in pre-mRNA degradation are not yet fully understood.

In addition to normal cellular transcripts, RNA exosome is also critical for rapid elimination of distinct types of aberrant transcripts. In particular, mRNA transcripts that lack a stop codon are degraded via the “nonstop mediated mRNA decay” pathway. These transcripts can arise through mutation, errors in transcription, or premature polyadenylation. Nonstop mRNA decay substrates are rapidly targeted for degradation by the cytoplasmic RNA exosome in a process requiring ribosomal stalling at the 3’ end of the transcript (Frischmeyer et al., 2002; van Hoof et al., 2002). RNA exosome is also involved in cytoplasmic 3’-5’ directed “nonsense mediated mRNA decay” of aberrant mRNA molecules containing premature stop codons (Mitchell and Tollervey, 2003; Takahashi et al., 2003). However, the role of RNA exosome in nonsense mediated mRNA decay is limited, as degradation from the opposite end of the transcript involving decapping and Xrn1 nuclease mediated 5’-3’ directed degradation is more robust (Schaeffer et al., 2010). Aberrant “read-through” transcripts that bypass transcription

termination signals can be targeted for RNA exosome mediated degradation (Rondon et al., 2009; Torchet et al., 2002). Translationally stalled transcripts observed in “no-go decay” also are subject to RNA exosome mediated degradation following endonucleolytic cleavage (Doma and Parker, 2006). In summary, RNA exosome can target a multitude of cellular transcripts. How RNA exosome identifies transcripts destined for degradation is poorly understood. It is likely that cofactors play an important role in recruiting RNA exosome to its targets. One of the better understood pathways involves co-transcriptional targeting of RNA exosome to a particularly unstable class of widespread cryptic transcripts in yeast.

Cryptic unstable transcripts (CUTs) encompass a class of RNA molecules deriving from intergenic regions originally discovered in yeast mutants of Rrp6 (Wyers et al., 2005). A very short half-life limits their abundance in wild type cells to nearly undetectable levels in the steady state due to rapid RNA exosome mediated degradation. Ablation of RNA exosome function typically results in stabilization of CUTs (Wyers et al., 2005). Genome-wide analyses of CUT biogenesis have revealed their length to be approximately 300 bp on average (Neil et al., 2009; Xu et al., 2009). CUTs primarily initiate from nucleosome depleted regions of bidirectional promoters and are often transcribed divergent to stable mRNA transcripts (Neil et al., 2009; Xu et al., 2009). Unlike relatively stable mRNA transcripts, CUTs undergo a distinct form of transcription termination that is coupled to RNA degradation. In yeast, a complex composed of RNA binding proteins Nrd1 and Nab3 along with helicase Sen1 (NNS complex) is responsible for CUT transcriptional termination (Arigo et al., 2006; Thiebaut et al., 2006). The NNS complex interacts with transcriptionally engaged RNAP II molecules (Gudipati et al., 2008; Vasiljeva et al., 2008) and recruits TRAMP, which catalyzes 3’ end oligoadenylation of the nascent transcript (Grzechnik and Kufel, 2008). TRAMP is a nuclear polyadenylation complex composed of non-canonical polyA polymerase Trf4,

either RNA binding protein Air1 or Air2, and RNA helicase Mtr4 (LaCava et al., 2005; Wyers et al., 2005). Importantly, TRAMP strongly stimulates RNA exosome function by adding unstructured 3' tails onto transcripts destined for RNA exosome mediated degradation. In fact, RNA exosome targeted snoRNAs undergo NNS mediated transcriptional termination (Kim et al., 2006) and accumulate extended 3' oligoadenylated tails in RNA exosome mutant yeast (LaCava et al., 2005). Collectively, the NNS and TRAMP complexes help to recruit and stimulate RNA exosome activity at co-transcriptional targets.

RNA exosome studies performed in human cell lines have uncovered similarities as well as important differences between yeast and human. CUT like "promoter upstream transcripts" (PROMPTs) are stabilized in human HeLa cells upon shRNA mediated knockdown of RNA exosome subunits (Preker et al., 2008). Similar to CUTs, PROMPTs are highly unstable and are primarily transcribed from bidirectional promoters divergent to stable mRNA transcripts (Preker et al., 2008). In addition, PROMPTs can reach 2 kb in length, which is substantially longer than CUTs (Preker et al., 2008). Similarly performed studies in mouse embryonic stem cells indicate that RNA exosome sensitive upstream antisense RNA (uaRNA) are transcriptionally regulated species. Similar to many mRNA transcripts, uaRNA engaged RNAP II can undergo promoter proximal pausing (Flynn et al., 2011). Paused nascent uaRNA transcription complexes can be released through P-TEFb mediated phosphorylation of RNAP II to subsequently undergo RNA exosome mediated degradation (Flynn et al., 2011).

When compared with yeast, mammalian cells may also utilize different mechanisms of RNA exosome recruitment to co-transcriptional targets. For instance MTR4, the human homolog of yeast TRAMP subunit Mtr4, exists in two distinct nuclear complexes. "Human TRAMP" complex resembles yeast TRAMP in its composition and contains MTR4, TRF4, and ZCCHC7 (human homolog of yeast Air1/2) (Lubas et al.,

2011). The nuclear exosome targeting (NEXT) complex, is made up of MTR4 and RNA binding proteins RBM7 and ZCCHC8 (Lubas et al., 2011). Surprisingly, NEXT complex is required for RNA exosome mediated degradation of PROMPTs, whereas human TRAMP complex is not (Lubas et al., 2011). However, it is unclear how NEXT complex is recruited to RNA exosome co-transcriptional substrates as obvious homologs of yeast NNS complex subunits Nrd1 or Nab3 appears to be absent in humans (Chlebowski et al., 2013). Recent studies indicate that the cap-binding complex (CBC) physically associates with NEXT complex and RNA exosome (Andersen et al., 2013). Consistent with a role for CBC in RNA exosome targeting to co-transcriptional substrates, PROMPT stabilization resulting from individual knockdown of NEXT complex or RNA exosome subunits is exacerbated when paired with shRNA mediated knockdown of CBC subunits (Andersen et al., 2013).

# Chapter 2: Materials and Methods

## 2.1 Bacterial homologous recombination

DH10B strain *E. coli* clones were obtained from a mouse genomic DNA BAC library and electroporated with plasmid pTSEBG-rev. This plasmid contains the lambda phage *exo*, *bet*, and *gam* open reading frames as a single polycistronic unit under control of an arabinose inducible promoter. pTSEBG-rev also contains the temperature sensitive pSC101 origin of replication. Single colonies carrying both the BAC clone and pTSEBG-rev were grown to mid-log phase at 30°C. The cultures were then induced with L-arabinose (Sigma) to a final concentration of 0.2% and grown at 37°C for 45 minutes. Competent cells were immediately prepared from these cultures and electroporated with linearized donor vector. Cells were then cultured in LB broth for an additional 4 hours and plated overnight at 37°C in the presence of antibiotics specific for the recombinant BAC.

## 2.2 Pulsed-field gel electrophoresis (PFGE)

BAC DNA (500 ng) was digested using restriction enzymes for 4 hours. DNA fragments were embedded in 1% low melting temperature agarose (SeaPlaque; Lonza). The agarose plugs were allowed to solidify in the wells of a 1% PFGE-grade agarose gel (SeaKem Gold; Lonza) prepared in 0.5X TBE buffer. Electrophoresis was performed using the CHEF-DR II PFGE system (Bio-Rad). PFGE running conditions were in 0.5X TBE buffer at 6 volts, for 11 hours at 12°C, using 1 sec and 6 sec initial and final switch times, respectively. The gel was soaked in 0.5 µg/ml ethidium bromide for 30 minutes, washed several times with $dH_2O$, and visualized over UV light.

## 2.3 Cell culture and class switch recombination

Splenic B cells from sex matched littermate mice were prepared using anti-CD43 microbead (Miltenyi Biotec) negative selection and cultured in RPMI 1640 media

containing 15% FBS. Ex vivo cultures utilizing the *ROSA26*$^{CreERt2}$ allele were cultured for 16 hours with 100 nM 4-hydroxytamoxifen (Sigma) and 20 μg/ml LPS (Sigma) followed by addition of IL-4 (R&D Systems) at 20 ng/ml, and cultured for an additional 72 hours in a 37°C incubator. Cells were collected, resuspended in cold PBS containing 2% FBS and incubated with anti-CD16/CD32 neutralizing antibody (BD Biosciences) for 15 minutes on ice. Cell suspensions were subsequently stained using fluorescently labeled anti-B220 and anti-IgG1 antibodies (BD Biosciences) for 30 minutes on ice with periodic mixing. Data were acquired on a FACSAria cell sorter (BD Biosciences) and analyzed using FloJo software (Tree Star).

### 2.4 Western blot analysis

Cells were collected, washed once with cold PBS buffer, and centrifuged. Supernatant was discarded and cell pellets were resuspended in 1 ml cold RIPA buffer containing proteinase inhibitor cocktail (Roche) per $10^7$ cells and incubated on ice with periodic mixing for 20 minutes. Lysates were then centrifuged at top speed in a table top centrifuge for 15 minutes. The clarified lysates were collected and stored at -80°C. Protein concentrations were determined using the bicinchoninic acid (BCA) reagent method (Pierce). Ten μg of total protein was used for each sample during SDS-PAGE. Antibodies used during immunoblotting were anti-Exosc3 (Santa Cruz Biotechnology), anti-Exosc10 (Sigma), anti-AID (Chaudhuri et al., 2003), anti-Actin (Abcam), and goat anti-rabbit IgG HRP (Jackson ImmunoResearch).

### 2.5 Chromatin immunoprecipitation

Crosslinking was performed on cultured B cells using formaldehyde. Sonication was carried out on ice using a Branson Sonifier 250 apparatus for 20 cycles, each cycle comprising 20 sec of sonication at duty cycle 30% followed by a 2 min rest period.

Lysates were pre-cleared for 1 hr. Lysates were incubated with indicated antibodies overnight at 4 °C. Protein A/G conjugated sepharose beads were added for 90 min with continued rotation. Beads were subsequently treated to the standard series of washes (low salt, high salt, LiCl, and TE) and ChIP products were eluted followed by RNase A treatment overnight at 60 °C and proteinase K treatment for 2 hr at 55 °C. ChIP DNA was recovered using ethanol precipitation. Primers used for ChIP quantitative PCR were as follows, *Myc*: CGGTTGATCACCCTCTATCACTC and GCTCCACACAATACGCCATGTAC; *Pim1*: CCCAGGATCTAGCCCACATAACATC and AGCGTAGCAAGTTGTGAGAAATGG; *Pax5*: CTGCTAGGATGGTTCTGCTTGG and CAACTCAATTGCAACCTCCATAGGTC; *Cd79b*: TGCTGATTGAGAAGGTTGGTGTG and GGAAGGGGTTGCTCCTGAATC; *Cd83*: AGATCTCCCTTGCTCAAACAACG and GACCTGCTACTCTCCAGATTTTGTG; *Cmas*: GGAAAACGGAAAGAGGCTGGAG and TGAGCTCAGAGGAGCCTCTAG; *Atp13a2*: CAGCCTGTCCTTTTCCGTCTATC and AGCTCGCTGAGATCTTGATGC; *March2*: GCAGCAAGTCTACAGCCAGAG and GCCTCTGAGTATCATCTGCCAATC; *Fam107b*: GACACCTTCCATTAGACAGGTGAC and AGATGAGAGCTCTGGATCCTTGG.

### 2.6 DNA:RNA hybrid immunoprecipitation (DRIP)

DRIP was performed on cultured B cells according to a previously described method (Ginno et al., 2012) with minor modifications involving restriction endonuclease selections. Briefly, cells were digested overnight in TE buffer with proteinase K, followed by phenol-chloroform extraction and ethanol precipitation of nucleic acids. Genomic DNA was digested using *Bso*BI, *Nhe*I, *Nco*I, and *Stu*I in the presence or absence of RNase H (New England Biolabs). Purified DNA was subsequently immunoprecipitated using the RNA:DNA hybrid specific S9.6 monoclonal antibody (kind gift of F. Chédin). DRIP and ChIP primers used for quantitative PCR were the same.

## 2.7 CRISPR/Cas9 mediated deletion of *Cd83* and *Pim1* xTSS-RNA regions

Guide RNA sequences were designed using an online tool (http://tools.genome-engineering.org). Guide RNA encoding oligonucleotides (*Cd83*, AGTGCCCAACACTACCTAAT and TTCCGAAGCCTCAGGGCGCG; *Pim1*, ATCAGACACATTCCGAGAAG and CTCTGTGTTTCCCGGAGATT) were cloned into the BbsI site of pSpCas9(BB)-2A-GFP (pX458 Addgene) as described (Ran et al., 2013). Guide RNA/Cas9 expression vectors were electroporated into CH12F3 cells using Amaxa Nucleofector (Lonza). Cells were cloned using limiting dilution 3 days after electroporation. Individual clones were screened for homozygous deletion of xTSS-RNA encoding regions using PCR. Screening primers for *Cd83* xTSS-RNA deletion were CCATGCTACAATGCACAGACCTAC and CAGCCTAGAAACAGGAGCTGGAG. Screening primers for *Pim1* xTSS-RNA deletion were CCAGGGATCAAACCTAGGATTTTC and CAGAAGACGCCCTATTTGCATAAGG. AID ChIP primers were as follows: *Pim1*, CTCGCTCCGCCGCCGCTGCTG and CGCAGGTGGGCCAGGGAGTTGAT; *Cd83*, GCCTCCAGCTCCTGTTTCTA and TCGGAGCAAGCCACCGTCAC.

## 2.8 Somatic hypermutation

Peyer's patches were excised from 2.5 – 3.5 month old paired littermates and gently dissociated by passage through a 70 µm cell strainer. Germinal center B cells were stained with anti-B220 (BD Biosciences) and peanut agglutinin (Vector Laboratories). DAPI stain was added just prior to cell sorting to exclude dead cells. Cells were directly sorted into lysis buffer containing proteinase K (Viagen) using a FACSAria cell sorter (BD Biosciences) and incubated at 55°C overnight. 15 µg GlycoBlue (Life Technologies) was added to the lysates and genomic DNA was purified via ethanol precipitation. A portion of the JH4 intron was amplified by PCR using LA Taq (Takara)

and a primer pair (J558FR3Fw: 5'-GCCTGACATCTGAGGACTCTGC-3' and JH4intronRv: 5'-CCTCTCCAGTTTCGGCTGAATCC-3') that requires a VDJ rearrangement of the IgH locus for amplification to occur (Jolly et al., 1997). JH4 amplicons were ligated into the pCR2.1-TOPO vector (Life Technologies) and individual clones were sequenced using M13 primers. Mutation analysis at *Cd83* in CH12F3 cells was performed on a 488 base pair sequence located ~150 base pairs downstream to the *Cd83* mRNA TSS between PCR primers GCCTCCAGCTCCTGTTTCTA and TGTTGCTTTCACTGCAGCTCTC.

### 2.9 Cell proliferation analysis using intracellular fluorescent tracer

Splenic B cells were prepared using anti-CD43 microbead negative selection. At the completion of the purification protocol, cells were washed twice with serum free PBS and resuspended at a cell concentration of 2 x $10^6$ per ml. An equivalent volume of 2 µM Violet Proliferation Dye 450 (VPD450) (BD Biosciences) was added to the cells and incubated at 37°C for 15 minutes. Extracellular VPD450 was quenched by adding 5 volumes of RPMI 1640 media containing 15% FBS. Cells were centrifuged, resuspended in fresh growth media at 0.2 x $10^6$ per ml and cultured in a 37°C incubator. VPD450 and GFP fluorescence was monitored on a daily basis using flow cytometry.

### 2.10 RNA preparation and quantitative RT-PCR

Total RNA was isolated from B cells using Trizol reagent (Life Technologies). RNA was resuspended in nuclease-free water and quantified using a NanoDrop spectrophotometer. Total RNA (1.5 µg) was converted to cDNA using random hexamer oligonucleotides and the Superscript III First-Strand Synthesis System (Life Technologies). *Exosc3*, *Exosc10*, and *Aicda* mRNA measurements were performed using the TaqMan Gene Expression Assay (Applied Biosystems). All other qRT-PCR

experiments were performed using SYBR Green Master Mix (Roche Applied Science). Expression levels for individual transcripts were normalized against beta-actin and/or cyclophilin A with similar results. Fold changes in transcript levels were calculated as:

Fold change = $2^{(Ct^{WT,GOI} - Ct^{mutant,\ GOI})} / 2^{(Ct^{WT,actin} - Ct^{mutant,actin})}$

### 2.11 RNA sequencing and transcriptome assembly

Total RNA was depleted of ribosomal RNA using the Ribo-Zero rRNA removal kit (Epicentre). Libraries were prepared with Illumina TruSeq or TruSeq Stranded total RNA sample prep kits, and then sequenced with 50 to 60 million, 2×100 bp paired raw filter passing reads on an Illumina HiSeq 2000 V3 instrument at the Columbia Genome Center. To assemble each transcriptome, all total RNA reads were first mapped onto the mouse reference genome (mm9) with TopHat version 1.3.2 (Trapnell et al., 2009). Cufflinks version 1.2.1 was subsequently applied to assemble the whole transcriptome and to identify all possible transcripts (Trapnell et al., 2010). To retain short RNA information, the Cufflinks overlap-radius was set to one and biological replicates were merged. Assembled transcripts were further annotated by overlapping with transcript coordinates of annotated mRNA/lncRNA (UCSC mouse genome browser), lincRNA (Guttman et al., 2009), and snoRNA (Yoshihama et al., 2013). The transcript overlap window was set to 200 bp of the annotated transcription start and end sites. Assembled transcripts possessing no overlap with any annotated RNA located within 2 kb upstream of an annotated transcript were classified as gapped upstream transcripts.

### 2.12 Statistical analysis

To test the significance of the difference of two vectors, two-sided nonparametric Wilcoxon rank sum test was applied to calculate p-values. To test the difference between two proportions, the following equation was used:

$$Z = \frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $p_1$ and $p_2$ are two proportions, $p$ is the expected value, and $n_1$ and $n_1$ are the population size. P-values were generated by normal distribution. DEseq pipeline (Anders and Huber, 2010) was applied for the normalization of library size and differential gene expression analysis.

### 2.13 Analysis of B cell translocation capture sequencing (TC-Seq) datasets

Genome-wide B cell translocation breakpoint datasets were downloaded under SRA accession number SRA039959 (http://www.ncbi.nlm.nih.gov/sra). Mapping of the breakpoints was performed according to a previously described computational workflow (Oliveira et al., 2012). All alignments were performed using the Burrows-Wheeler Aligner (Li and Durbin, 2009) with default parameters. Read counting was performed using BEDTools (Quinlan and Hall, 2010).

### 2.14 Statistical model for predicting translocation breakpoint probability

To predict the probability of observing breakpoints in a given genomic region, we assume the number of translocation (TC-Seq) reads within a given region follows a negative binomial model with parameters *r* and *p*. A maximal likelihood method was applied to estimate the parameters. At each given genomic region the probability of translocation breakpoint occurrence was defined as the probability to harbor more than 0 TC-Seq reads. Two types of genomic regions were evaluated separately. For the analysis of translocation hotspot genes each base pair spanning 2 kb upstream from the TSS to the TES were binned into 13 tiers according to antisense transcript expression

levels. For the genome-wide analysis, annotated genes that do not harbor known antisense transcripts (28,947 loci) were binned into 46 antisense expression tiers.

### 2.15 Mouse embryonic stem cell derivation and cell culture

*Exosc3* conditional null ES cells were derived through mating of female *Exosc3*$^{COIN/+}$ and male *Exosc3*$^{COIN/+}$ *ROSA26*$^{CreERt2/CreERt2/+}$ mice. Likewise, *Exosc10* conditional null ES cells were derived by crossing female *Exosc10*$^{COIN/+}$ with male *Exosc10*$^{LacZ/+}$ *ROSA26*$^{CreERt2/CreERt2/+}$ mice. Three week old female mice were superovulated by intraperitoneal injection of 5 units pregnant mare's serum gonadotropin (PMSG) followed 47 hours later by intraperitoneal injection of 5 units human chorionic gonadotropin (hCG). Immediately following delivery of hCG, female mice were paired individually with rested stud male mice. Mating pairs were separated the following morning. Females were euthanized 3 days later for blastocyst isolation. Individual ES cell clones were generated with the assistance of the Transgenic Mouse Shared Resource (Herbert Irving Comprehensive Cancer Center). ES cells were maintained in DMEM supplemented with 15% FBS, 1 mM sodium pyruvate, 1X non-essential amino acids, 1X L-glutamine, 1X penicillin/streptomycin, 100 μM 2-mercaptoethanol, and 1000 units/mL murine leukemia inhibitory factor (Millipore). ES cell cultures were typically maintained on irradiated CF-1 mouse embryonic fibroblast (Global Stem) feeder layers and received daily media changes. For *Exosc3* or *Exosc10* ablation studies, ES cell cultures were treated with 4-OHT for 2 days, followed by 2 additional days without 4-OHT. GFP$^{+}$ (*Exosc3*) or FP635$^{+}$ (*Exosc10*) ES cells were isolated using cell sorting.

# Chapter 3: Results

The results presented in this chapter have been partly published in,

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Chao, J., Rabadan, R., Economides, A.N., and Basu, U. (2014). Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. Nature *514*, 389-393.

### 3.1 Selection of *Exosc3* and *Exosc10* for B cell studies

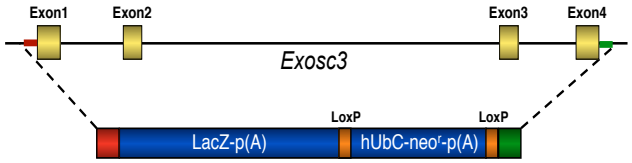
To explore the role of RNA exosome in AID mediated immunoglobulin diversification mechanisms and B lymphocyte transcriptional regulation we undertook a path to mutagenize *Exosc3* and *Exosc10* in mouse cells. The rationale for focusing on these two subunits was that it might allow us to identify exosome mediated co-transcriptional RNA processing activities that may be key to AID related functions. Loss of *Exosc3*, encoding the RRP40 structural subunit necessary for assembly of the core exosome complex (Liu et al., 2006), would likely translate into a complete loss of RNA exosome function. *Exosc10* encodes the RRP6 3'-5' exoribonuclease subunit of the exosome complex. In contrast to RRP40, RRP6 is not required for assembly of the core exosome complex (Liu et al., 2006) and its localization is greatly enriched in the nucleus (Tomecki et al., 2010). As transcription coupled RNA processing and AID mediated deamination of DNA are strictly nuclear events we hypothesized that *Exosc10* ablation may result in perturbation of these activities owing to nuclear specific exosome defects. An important caveat to this prediction is potential redundancy between *Exosc10* and *Dis3*, which encodes the RRP44 nuclease subunit found in both nuclear and cytoplasmic exosome complexes (Tomecki et al., 2010). *Exosc3* deficiency however, is expected to disrupt nuclear as well as cytoplasmic functions of the exosome complex.

### 3.2 Initial attempts at creating and analyzing RNA exosome developmentally deficient mice

We sought to generate RNA exosome developmentally deficient mice using embryonic stem cells containing targeted, monoallelic LacZ cassette based whole gene replacements of *Exosc3* (Figure 4A) or *Exosc10*. F1 generation heterozygotes of *Exosc3* or *Exosc10* were intercrossed, but viable $Exosc3^{LacZ/LacZ}$ or $Exosc10^{LacZ/LacZ}$ pups were not obtained (Figure 4B). Embryos from $Exosc3^{LacZ/+}$ intercrosses were also genotyped

at 13.5 days of gestation, but once again *Exosc3*$^{LacZ/LacZ}$ embryos failed to be observed. However, this embryo analysis uncovered a discrepancy between the number of uterine decidua and the number of actual embryos. Of the 38 decidua observed, only 30 contained embryos. This difference is in line with the expected Mendelian ratio of *Exosc3*$^{LacZ/LacZ}$ embryos and is likely to account for the *Exosc3*$^{LacZ/LacZ}$ embryos that fail to develop post-implantation.

A




B

| | pups genotyped | WT | HET | KO |
|---|---|---|---|---|
| *Exosc3*$^{LacZ}$ | 106 | 40 | 66 | 0 |
| *Exosc10*$^{LacZ}$ | 98 | 38 | 60 | 0 |

**Figure 4: Evidence for embryonic lethality in *Exosc3* knockout mice**
(A) Schematic of the *Exosc3*$^{LacZ}$ allele (KOMP ID# VG10322). Homologous recombination was used to replace the entire coding region of *Exosc3* (3.8 kb) with the ZEN-Ub1 cassette (LacZ open reading frame, *lox*P flanked human ubiquitin C promoter expressing neomycin phosphotransferase). Red and green boxes indicate *Exosc3* 5' and 3' untranslated regions, respectively.
(B) Embryonic lethality in *Exosc3* and *Exosc10* developmentally deficient mice. Genotypes obtained from 3-week old pups generated from multiple *Exosc3*$^{LacZ/+}$ or *Exosc10*$^{LacZ/+}$ intercrosses.

*Exosc3*$^{LacZ/+}$ heterozygotes were analyzed for potential CSR defects. Wild type and *Exosc3*$^{LacZ/+}$ naive IgM$^{+}$ splenic B cells were isolated and cultured in the presence of LPS with and without IL-4 to induce isotype switching to IgG1 and IgG3, respectively. While *Exosc3*$^{LacZ/+}$ B cells express 50% less *Exosc3* mRNA as compared to wild type B cells (Figure 5A), nearly identical levels of IgG1 and IgG3 conversion was observed between genotypes (Figure 5B). Thus, the absence of *Exosc3* haploinsufficiency in B cells undergoing CSR along with the inability to produce fetal liver chimerism due to embryonic lethality of *Exosc3*$^{LacZ/LacZ}$ embryos prior to E13.5, necessitated the construction of conditional alleles of *Exosc3* and *Exosc10*.

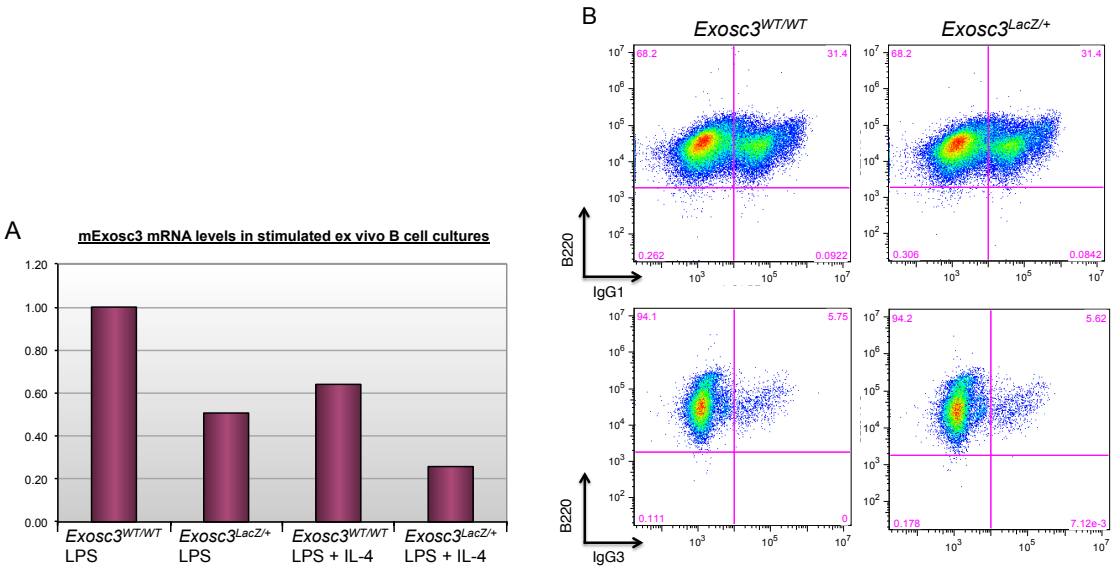



**Figure 5: Absence of *Exosc3* haploinsufficiency in B cells undergoing CSR**
(A) Quantitative RT-PCR analysis of *Exosc3* mRNA expression in cultured B cells undergoing CSR. Expression levels are normalized to beta-actin and plotted relative to LPS treated wild type cells.
(B) Flow cytometric analysis for surface IgG1 and IgG3 on purified B cells stimulated with LPS with and without IL-4, respectively. Upper right quadrant of each plot indicates the percentage of isotype switched B cells.

## 3.3 Construction of *Exosc3* and *Exosc10* conditional alleles

### 3.3.1 *Design strategy for Exosc3 and Exosc10 conditional alleles*

The design of the *Exosc3* and *Exosc10* conditional alleles was primarily influenced by three principles. First, each conditional allele must convert to a null allele upon exposure to Cre recombinase activity. Second, the conditional alleles must exert wild type functionality in the absence of Cre. Third, the accommodation of a reporter system to allow for simple and robust identification of cells bearing conditional alleles pre- and post-Cre activity. These principles are featured in the conditionals by inversion (COIN) allele approach (Economides et al., 2013) and were implemented in the design of the *Exosc3* (Figure 6) and *Exosc10* conditional alleles.

The *Exosc3*$^{COIN}$ allele utilizes a Cre dependent genetic "FLEx switch" (Schnutgen et al., 2003) to inactivate the *Exosc3* open reading frame while simultaneously activating GFP expression from the endogenous locus. This is mediated through the usage of wild

type *lox*P and mutant *lox*2372 Cre recognition sequences while leveraging the activity of Cre recombinase to invert or excise as circular DNA any DNA sequence according to the orientation of the flanking *lox* sites (Abremski et al., 1983). A critical feature of the COIN allele lies in the incompatibility of *lox*P and *lox*2372 sites to serve as Cre substrates (Siegel et al., 2004). *lox*P/*lox*P and *lox*2372/ *lox*2372 pairs can serve as efficient substrates for Cre, while the *lox*P/*lox*2372 pair cannot.

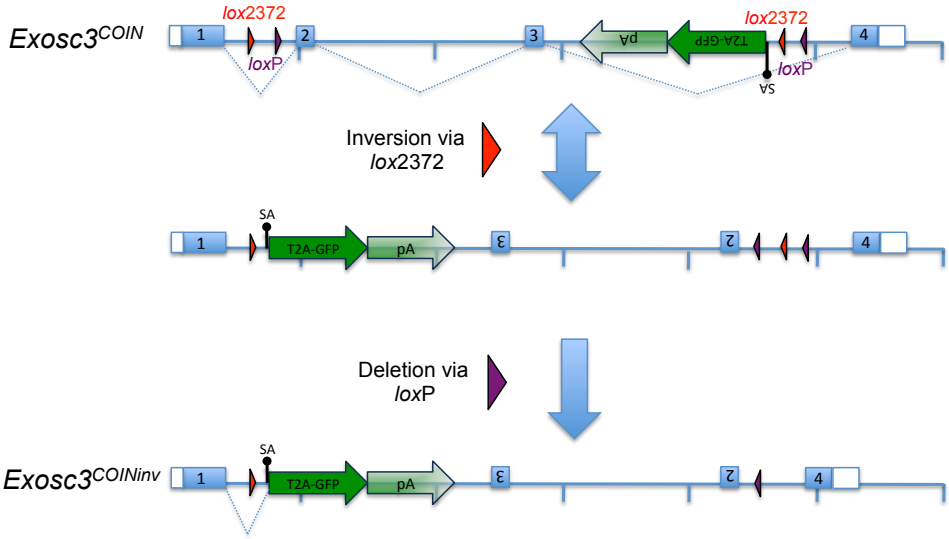



**Figure 6: Schematic of the *Exosc3*$^{COIN}$ allele and conversion to *Exosc3*$^{COINinv}$**
Cre recombinase mediated inversion of the *lox*2372 pair (red triangles) and subsequent deletion via the *lox*P pair (violet triangles). GFP expressing terminal exon represented by green arrows. Exons illustrated as numbered boxes. Splicing pattern indicated by dashed lines. SA, splice acceptor. T2A, ribosomal skipping sequence.

The murine *Exosc3* locus is comprised of 4 exons, all of which are coding, and contains a single consensus coding sequence (CCDS) transcript annotation. To preserve the normal splicing pattern of the *Exosc3* coding transcript in the context of the *Exosc3*$^{COIN}$ allele in the absence of Cre, the upstream and downstream *lox*P/*lox*2372 arrays were placed in the first and third introns, respectively (Figure 6). Immediately upstream of the downstream *lox* array within intron 3 of the *Exosc3*$^{COIN}$ allele lies an inverted GFP expressing terminal exon, which is designed to be silent in the COIN configuration due to its opposite orientation relative to *Exosc3* transcription. The precise

insertion points of the *lox* arrays and GFP terminal exon into the *Exosc3* locus were selected according to weak evolutionary conservation within these introns (Ovcharenko et al., 2004) to minimize the probability of disrupting non-annotated regulatory elements.

Although the *Exosc3* open reading frame is expected to be maintained in the *Exosc3*$^{COIN}$ allele, Cre mediated conversion to *Exosc3*$^{COINinv}$ is engineered to disrupt expression of *Exosc3* through multiple mechanisms. *Exosc3*$^{COIN}$ contains two distinct Cre substrates, namely the pair of inverted *lox*P sites and the pair of inverted *lox*2372 sites. Regardless of which *lox* pair is utilized initially, the final product remains *Exosc3*$^{COINinv}$. Figure 6 illustrates the conversion of *Exosc3*$^{COIN}$ to *Exosc3*$^{COINinv}$ initiating via inversion of the *lox*2372 pair. This results in inversion of the single intervening *lox*P site, such that the *lox*P pair is now in direct repeat orientation. Subsequent deletion between the *lox*P pair results in loss of the intervening *lox*2372 site and locks the allele in the *Exosc3*$^{COINinv}$ configuration. This disrupts *Exosc3* expression through two distinct mechanisms. First, the inversion of *Exosc3* exons 2 and 3 destroys the open reading frame of the *Exosc3* coding transcript. Secondly, inversion of the GFP terminal exon should prematurely terminate transcription arising from the *Exosc3* promoter at the polyadenylation sequence located immediately downstream of GFP. In addition, the inversion mechanism creates an endogenous reporter within the *Exosc3* locus. Transcription of the *Exosc3*$^{COINinv}$ allele results in splicing of exon 1 onto the GFP terminal exon. Fused upstream of the GFP open reading frame lies a splice acceptor sequence followed by a T2A ribosomal skipping sequence (Szymczak et al., 2004). Translation of the *Exosc3*$^{COINinv}$ transcript is predicted to create two distinct protein products; soluble GFP protein and a fusion peptide corresponding to the translation product of *Exosc3* exon 1 and T2A.

The *Exosc10*$^{COIN}$ allele was designed to operate in a nearly identical manner as the *Exosc3*$^{COIN}$ allele with two minor differences. Due to exon phase differences between

*Exosc3* and *Exosc10*, the *Exosc10*$^{COIN}$ allele inverts only exon 2 while still disrupting the open reading frame of the wild type *Exosc10* transcript. Also, the reporter in *Exosc10*$^{COIN}$ was replaced with the far-red fluorescent protein TurboFP636 (Shcherbo et al., 2007).

### 3.3.2 *Exosc3*$^{COINneo}$ *and Exosc10*$^{COINneo}$ *targeting vector construction*

COIN allele targeting vector construction was accomplished through three sequential modifications of *Exosc3* or *Exosc10* containing bacterial artificial chromosomes (BACs) using homologous recombination in *E. coli* (Zhang et al., 1998). A schematic illustrating this process for *Exosc3* is outlined in Figure 7.

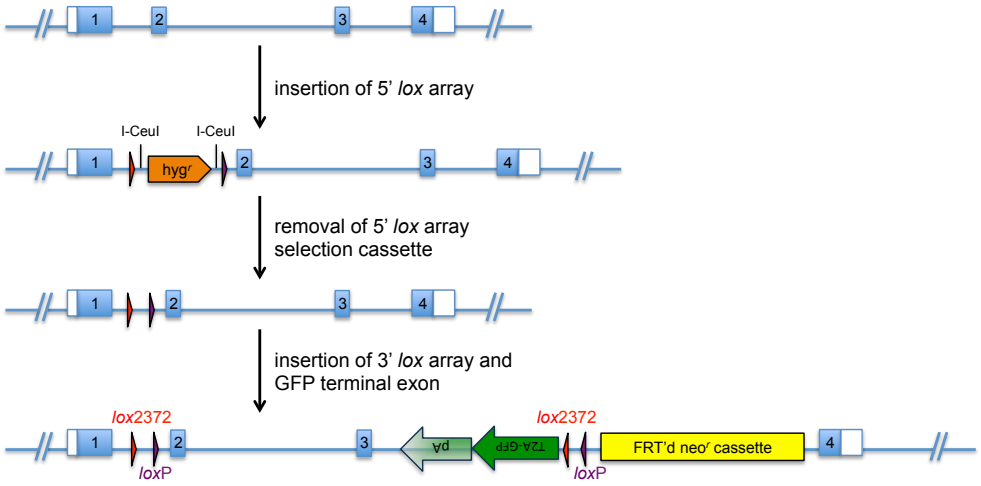



**Figure 7: Flowchart of *Exosc3*$^{COINneo}$ targeting vector construction**
Bacterial homologous recombination was used to insert the 5' *lox* array into intron 1 of *Exosc3*. BAC DNA from correctly modified clones was purified, digested with I-CeuI, and ligated in cis to remove the hyg$^{r}$ selection cassette from the 5' *lox* array. A second homologous recombination step was utilized to insert the inverted GFP terminal exon, 3' *lox* array, and hUbC-neo$^{r}$ selection cassette within intron 3.

BAC clones containing the *Exosc3* and *Exosc10* loci were obtained from a genomic DNA library of the 129Sv mouse strain (Adams et al., 2005). Donor vectors used to introduce the 5' *lox* array were constructed using standard molecular biology techniques. These 5' donor vectors contain a pair of short, ~70 bp sequences that are homologous to regions within introns 1 of *Exosc3* or *Exosc10*. Between the pair of homology sites lie the *lox*2372 and *lox*P elements, which themselves are separated by

an I-CeuI flanked hygromycin selection cassette. Transfer of the modified sequence from the 5' donor vector onto the BAC was performed through homologous recombination. In short, *E. coli* containing *Exosc3* or *Exosc10* BACs were transiently induced to express lambda phage *exo*, *bet*, and *gam* (Red operon) and subsequently transformed with linearized 5' donor vector. 5' *lox* array recombinants were screened by PCR for the presence of correctly modified intron 1 junctions using oligonucleotides located outside of the homology sites together with oligonucleotides specific for the hygromycin selection cassette.

The next step in the vector construction process was removal of the hygromycin selection cassette. BAC DNA from properly targeted 5' *lox* array recombinants was purified and digested with I-CeuI homing endonuclease. Following heat inactivation of I-CeuI, T4 DNA ligase was used to re-circularize the BACs. DH10B *E. coli* cells were transformed with this DNA and plated in the presence or absence of hygromycin. Correctly deleted clones were screened for by PCR across the deletion junction.

The final step in construction of the *Exosc3*$^{COINneo}$ and *Exosc10*$^{COINneo}$ BAC targeting vectors (BACvecs) involved insertion of the inverted terminal exon, 3' *lox* array, and neomycin selection cassette in introns 3 and 2, respectively. Accordingly, 3' donor vectors were constructed containing this modification flanked by *Exosc3* intron 3 or *Exosc10* intron 2 homology sites, in a manner analogous to the 5' donor vectors. Linearized 3' donor vectors were introduced into recombination primed *E. coli* cells containing hyg$^r$ cassette deleted 5' *lox* array modified BAC clones. Recombinants were screened by PCR across both unique junctions introduced during this step using oligonucleotides located outside the homology sites together with oligonucleotides recognizing elements within the modification (terminal exon polyadenylation sequence or neo$^r$ coding sequence).

To further screen the *Exosc3*$^{COINneo}$ and *Exosc10*$^{COINneo}$ BACvecs, both intron modifications were fully sequenced between each pair of homology sites and confirmed for the absence of any mutations. Additionally, each modified BAC was analyzed for the presence of any spurious rearrangements within the large homology arms flanking the *Exosc3* and *Exosc10* loci. Such rearrangements can potentially occur during induction of homologous recombination and can greatly affect subsequent gene targeting steps and locus integrity. Pulsed-field gel electrophoresis of the restriction fragments of the *Exosc3*$^{COINneo}$ and *Exosc10*$^{COINneo}$ BACvecs yielded the expected banding pattern (Figure 8), thus confirming the absence of any gross DNA rearrangements.

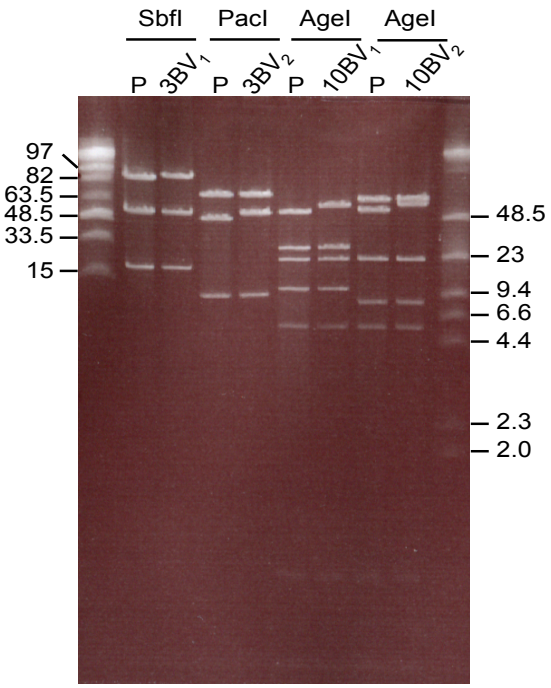



**Figure 8: Analysis of *Exosc3*$^{COINneo}$ and *Exosc10*$^{COINneo}$ BACvec integrity**
Pulsed-field gel electrophoresis of *Exosc3*$^{COINneo}$ and *Exosc10*$^{COINneo}$ BACvec restriction fragments. $3BV_1$, $3BV_2$, $10BV_1$, and $10BV_2$ indicate the COINneo modified *Exosc3* or *Exosc10* containing bMQ BAC library clones 452h18, 386a13, 169f23, and 395a09, respectively. P indicates the corresponding unmodified parental BAC clone. Molecular weight markers are indicated in kilobases. Predicted restriction fragment lengths are as follows (unique fragments highlighted):
SbfI – P: 91*, 57, 23 kb; $3BV_1$: 95*, 57, 23 kb
PacI – P: 72, 52*, 11 kb; $3BV_2$: 72, 56*, 11 kb
AgeI – P: 57*, 33, 27, 13, 5, 1.2, 0.9 kb; $10BV_1$: 62*, 33, 27, 13, 5, 1.2, 0.9 kb
AgeI – P: 67, 57*, 26, 9, 5, 1.2, 0.9 kb; $10BV_2$: 67, 62*, 26, 9, 5, 1.2, 0.9 kb

### 3.3.3 *Gene targeting of Exosc3$^{COINneo}$ in mouse embryonic stem cells*

The BACvec for *Exosc3$^{COINneo}$* was linearized and electroporated into *ROSA26$^{CreERt2/+}$*, 129S6/SvEv x C57BL/6 hybrid ES cells. Following selection with G418 antibiotic, individual colonies of cells were expanded and evaluated for replacement of the wild type allele with the COIN$^{neo}$ allele. BACvecs possess very long homology arms (41 kb and 82 kb for *Exosc3$^{COINneo}$*; 62 kb and 66 kb for *Exosc10$^{COINneo}$*) that are intended to increase the efficiency of gene targeting (Valenzuela et al., 2003). However, such homology arm lengths make it exceedingly difficult to screen for targeted ES cell clones using standard Southern blotting techniques involving restriction enzyme digestion schemes predicated on not cutting within the homology arms. Targeted clones were therefore identified using a quantitative PCR based copy number analysis approach (Frendewey et al., 2010). *Exosc3$^{COINneo/+}$* targeted clones were identified on the basis of differential amplification of wild type allele specific sequences. Specifically, Taqman probes corresponding to ~100 bp regions within *Exosc3* introns 1 and 3 which were disrupted due to the insertion of the *lox* arrays were used to differentiate the wild type from the COIN$^{neo}$ allele. Clone AF8 consistently required ~1 additional PCR cycle relative to the parental ES cell line when amplifying with wild type allele specific probes, whereas a non-targeted locus (*Wnt2b*) was equivalently amplified (Figure 9). Conversely, COIN$^{neo}$ allele specific elements such as GFP and neo$^{r}$ were present at single copy in clone AF8, but were absent in the parental line (not shown). This data collectively implies that a single wild type allele has been replaced with the COIN$^{neo}$ allele at the *Exosc3* locus of clone AF8.

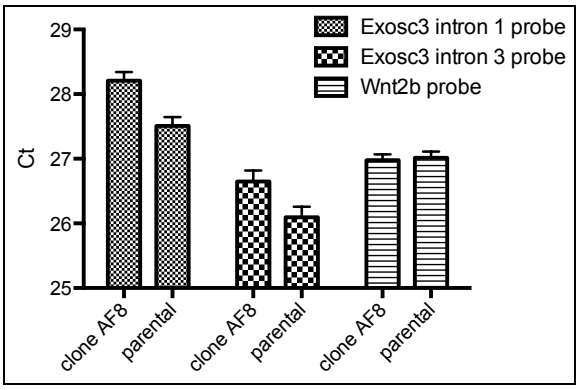



**Figure 9: Confirmation of *Exosc3*$^{COINneo/+}$ targeted ES cell clone AF8**
Quantitative PCR amplification of *Exosc3*$^{WT}$ allele specific elements (intron 1 and 3 probes) in a *Exosc3*$^{COINneo}$ BACvec targeted clone (AF8). The *Wnt2b* probe represents a non-targeted control locus. Genomic DNA was used as template. Ct, cycle threshold.

### 3.3.4 *Gene targeting of Exosc10*$^{COINneo}$ *in mouse embryonic stem cells*

Gene targeting efforts using the *Exosc10*$^{COINneo}$ BACvec failed to produce targeted ES cells, thus prompting an alternative approach. The *Exosc10*$^{COINneo}$ BACvec was used as a template to subclone a ~20 kb fragment containing the entire *Exosc10*$^{COINneo}$ modification with shorter homology arms into a plasmid containing a diphtheria toxin A (DTA) cassette. In comparison to the BACvec, the significantly smaller sized plasmid could be delivered at higher concentrations during ES electroporation. In addition, the DTA cassette provides a strong negative selection pressure against random integration events.

The linearized DTA based *Exosc10*$^{COINneo}$ targeting vector, containing 6.7 kb and 8.2 kb homology arms, was electroporated into *ROSA26*$^{CreERt2/+}$, 129S6/SvEv x C57BL/6 hybrid ES cells. Individual G418 resistant ES cell clones were screened by quantitative PCR using the same approach as previously described for *Exosc3*$^{COINneo}$. Candidate targeted clones were further screened using Southern blotting. Proper targeting of the upstream homology arm was assessed using an external probe located beyond the homology arm boundary (Figure 10A). Targeted *Exosc10*$^{COINneo}$ alleles could be distinguished from *Exosc10*$^{WT}$ alleles with this probe due to length differences of the

HindIII restriction fragments. Each of the candidate clones produced 2 distinct HindIII fragments, corresponding in size to the 16 kb targeted *Exosc10*$^{COINneo}$ allele and the endogenous 12 kb *Exosc10*$^{WT}$ allele (Figure 10B). In contrast, the parental ES cell line yields only the 12 kb *Exosc10*$^{WT}$ HindIII fragment. A similar Southern blot screen involving NsiI digestion and an external downstream probe confirmed that the downstream homology was also properly recombined (not shown). An additional Southern blot was also performed on the HindIII digested clones using a neo$^{r}$ specific probe to determine the number of copies present per genome. Since the targeting vector contains a single HindIII site, the number of restriction fragments detected by the neo$^{r}$ probe corresponds to the number of integration events. Each of the candidate clones contained a single 16 kb neo$^{r}$ specific fragment corresponding to the *Exosc10*$^{COINneo}$ allele (not shown), thus ruling out the presence of random non-targeted integration events.

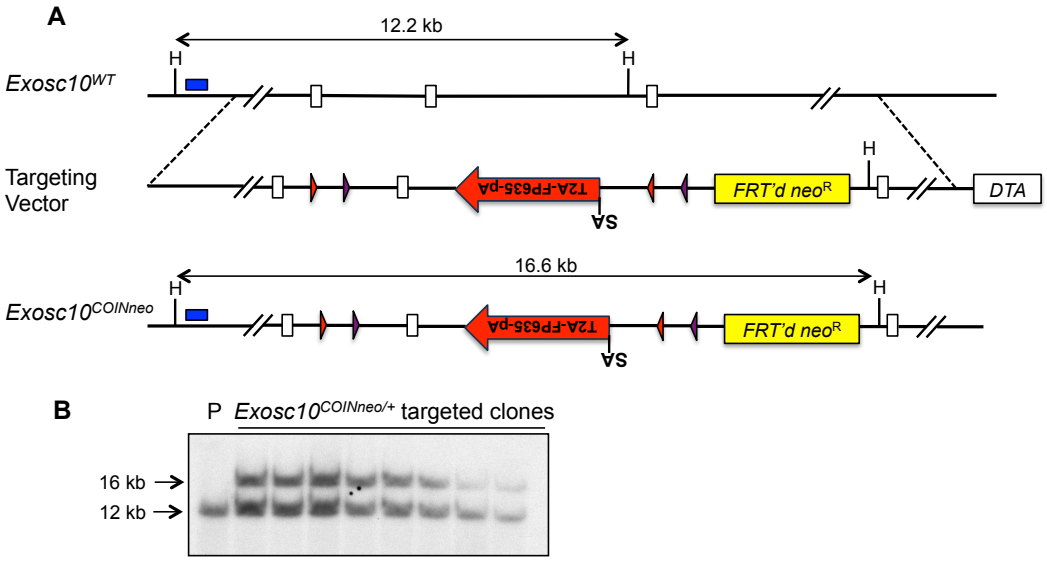



**Figure 10: Southern blot screening of targeted *Exosc10*$^{COINneo/+}$ ES cell clones**
(A) Schematic of the *Exosc10* wild type and COINneo genomic loci. External upstream homology arm Southern blot probe represented by blue box. Dashed lines represent the homology arm boundaries. Exons 1-3 represented by open boxes. Red and violet triangles represent *lox*2372 and *lox*P sites, respectively. Red arrow illustrates the pre-inverted T2A-FP635 terminal exon. SA, splice acceptor site. DTA, diphtheria toxin A cassette. H, HindIII site.
(B) Southern blot screen of HindIII digested *Exosc10*$^{COINneo/+}$ candidate ES cell clones using the external upstream homology arm probe. 16 kb and 12 kb bands correspond to the *Exosc10*$^{COINneo}$ and *Exosc10*$^{WT}$ alleles, respectively. P, parental ES cell clone.

### 3.4 Functional characterization of *Exosc3*$^{COIN}$ allele inversion in B cells

Targeted *Exosc3*$^{COINneo/+}$ *ROSA26*$^{CreERt2/+}$ ES cells were injected into blastocysts and subsequently implanted into pseudopregnant recipients to give rise to chimeric mice. Male chimeras displaying the greatest ES cell contribution on the basis of coat color were crossed with female *ACTB:FLPe* transgenic mice (Rodriguez et al., 2000) resulting in germline deletion of the *FRT* flanked *Exosc3*$^{COINneo}$ neo$^{r}$ cassette. *Exosc3* heterozygous mice were intercrossed to produce wild type, heterozygous, and homozygous *Exosc3* conditional allele genotypes on the *ROSA26*$^{CreERt2/+}$ background. Mendelian ratios of each genotype were observed (not shown).

In order to assess the function of the *Exosc3*$^{COIN}$ allele in live cells, naive splenic B cells were purified from *Exosc3*$^{COIN/+}$ and *Exosc3*$^{COIN/COIN}$ mice using anti-CD43 magnetic bead negative isolation. Cells were treated with 4-hydroxytamoxifen (4-OHT) to activate CreERt2, and stimulated with lipopolysaccharide (LPS) and interleukin 4 (IL-4) to induce CSR to IgG1. CreERt2-mediated inversion of the *Exosc3*$^{COIN}$ allele to *Exosc3*$^{COINinv}$ produces a defined alteration in the genomic locus that can be monitored by Southern blotting due to the repositioning of restriction sites. A strategy was developed to distinguish the WT, COIN, and COINinv alleles of Exosc3 (Figure 11A). Genomic DNA was prepared from B cells at 0, 2, 3, and 4 days following 4-OHT treatment, digested with HindIII enzyme and subjected to Southern blotting analysis. Prior to 4-OHT treatment (day 0), no evidence of CreERt2 activation was detected in either genotype as monitored by the presence of the *Exosc3*$^{COINinv}$ specific fragment (Figure 11B). All subsequent time points following 4-OHT addition (days 2-4) displayed a near complete conversion of the *Exosc3*$^{COIN}$ allele to *Exosc3*$^{COINinv}$ (Figure 11B). As expected, the *Exosc3*$^{WT}$ allele was insensitive to CreERt2 activation and was entirely absent in the *Exosc3*$^{COIN/COIN}$ derived DNA. Furthermore, the absence of unexpected

*Exosc3*-specific fragments provides evidence that the process of COIN allele inversion does not produce faulty chromosomal rearrangements arising from the *Exosc3* locus.

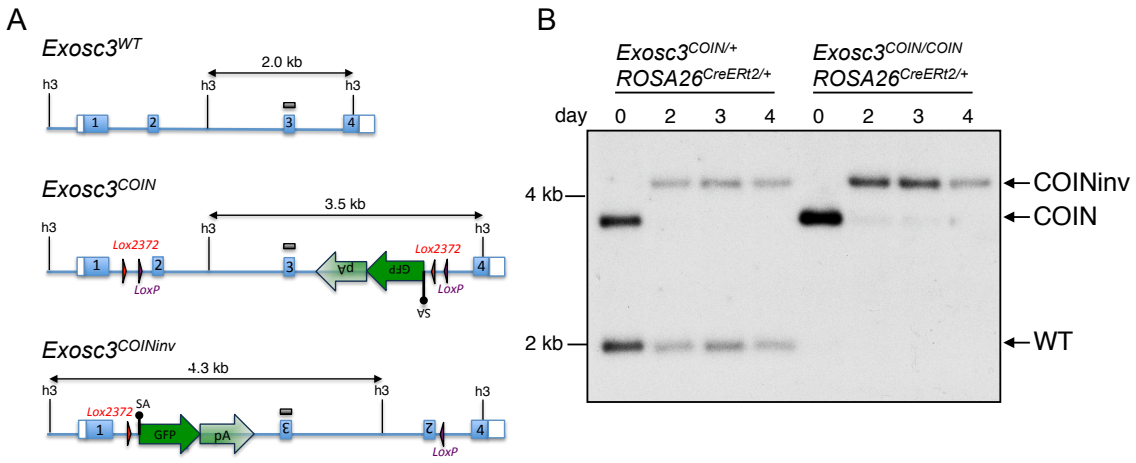



**Figure 11: *Exosc3*$^{COIN}$ allele inversion in B cells**
(A) HindIII restriction maps of the WT, COIN, and COINinv alleles of *Exosc3*. The black shaded box above exon 3 indicates the location of the probe used for Southern blotting in (B).
(B) Southern blot of HindIII digested genomic DNA from naive splenic B cells (day 0) or 4-hydroxytamoxifen (4-OHT) treated (days 2-4), LPS + IL-4 stimulated B cells from *Exosc3*$^{COIN/+}$ and *Exosc3*$^{COIN/COIN}$ mice on *ROSA26*$^{CreERt2/+}$ background. Radiolabeled probe is specific for exon 3 of *Exosc3*.

An important feature of the *Exosc3*$^{COIN}$ allele is induction of the GFP reporter upon inversion. This requires precise matching of the exon phases of *Exosc3* exon 1 and the inverted T2A-GFP terminal exon in order to maintain the translational reading frame required for production of GFP protein. B cells derived from *Exosc3*$^{COIN/+}$ and *Exosc3*$^{COIN/COIN}$ mice carrying the *ROSA26*$^{CreERt2}$ allele were analyzed for GFP fluorescence using flow cytometry. Naive B cells displayed little to no GFP fluorescence (Figure 12), which was consistent with the absence of detectable *Exosc3*$^{COINinv}$ allele as previously shown (Figure 11B). However, when cultured in the presence of 4-OHT, *Exosc3*$^{COIN/+}$ and *Exosc3*$^{COIN/COIN}$ B cells exhibited robust induction of GFP fluorescence (Figure 12).

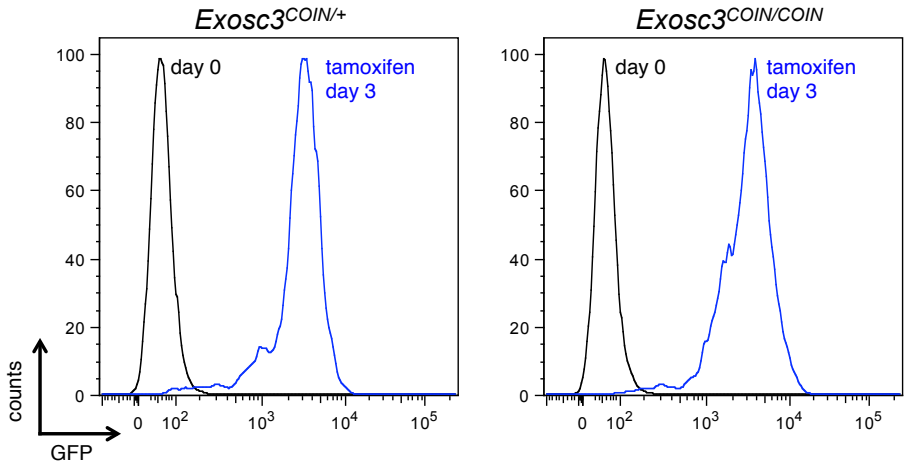



**Figure 12: Induction of GFP fluorescence upon *Exosc3*$^{COIN}$ allele inversion**
Flow cytometry analysis of GFP fluorescence in naive or 4-OHT treated, LPS + IL-4 stimulated B cell cultures. Indicated *Exosc3* genotypes are on *ROSA26*$^{CreERt2/+}$ background.

Most importantly, it was ascertained whether *Exosc3*$^{COINinv}$ behaves as a functional null allele of *Exosc3*. This was determined by measuring the mRNA and protein abundance of *Exosc3* following inversion of the *Exosc3*$^{COIN}$ allele. mRNA levels of *Exosc3* were equivalent between *Exosc3*$^{COIN/+}$ and *Exosc3*$^{COIN/COIN}$ B cells prior to CreERt2 activation (Figure 13A), indicating that the *Exosc3*$^{COIN}$ allele is not hypomorphic. Upon treatment with 4-OHT, *Exosc3* mRNA levels in *Exosc3*$^{COIN/+}$ *ROSA26*$^{CreERt2/+}$ cells are rapidly reduced yet stabilized at 50% of uninduced level (Figure 13A), due to the presence of a single Cre-insensitive *Exosc3*$^{WT}$ allele. In contrast, *Exosc3*$^{COIN/COIN}$ *ROSA26*$^{CreERt2/+}$ cells display a ~95% reduction in *Exosc3* mRNA abundance following 4-OHT treatment (Figure 13A). Whole cell protein extracts were prepared from similarly treated cells and analyzed for *Exosc3* protein expression by immunoblotting. CreERt2 activated *Exosc3*$^{COIN/COIN}$ B cells contained nearly undetectable levels of *Exosc3* protein relative to control *Exosc3*$^{WT/WT}$ cells (Figure 13B).

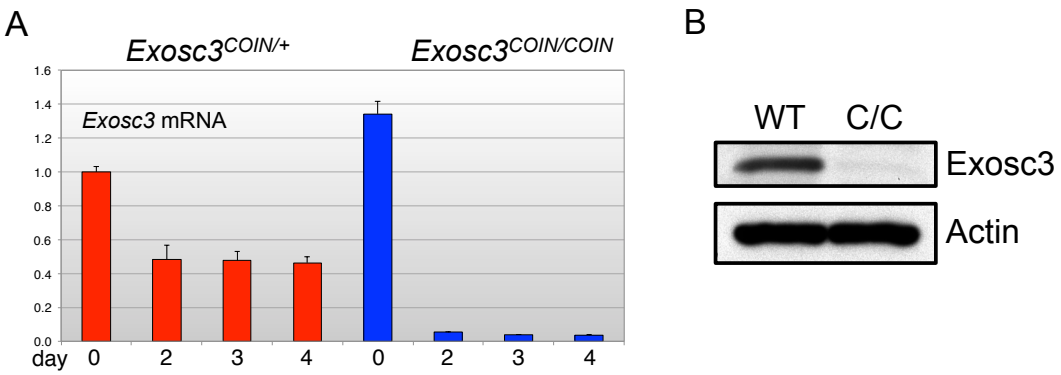



**Figure 13: Depletion of *Exosc3* mRNA and protein upon *Exosc3*$^{COIN}$ inversion**
(A) Quantitative RT-PCR time course analysis of *Exosc3* mRNA expression in naive (day 0) or 4-OHT treated (days 2-4), LPS + IL-4 stimulated B cell cultures. Indicated *Exosc3* genotypes are on *ROSA26*$^{CreERt2/+}$ background. Expression levels are normalized to cyclophilin A (*Ppia*) and plotted relative to naive *Exosc3*$^{COIN/+}$.
(B) Western blot analysis of *Exosc3* protein expression in whole cell extracts from 4-OHT treated, LPS + IL-4 stimulated B cell cultures. Anti-actin used as loading control. WT, *Exosc3*$^{WT/WT}$ *ROSA26*$^{CreERt2/+}$; C/C, *Exosc3*$^{COIN/COIN}$ *ROSA26*$^{CreERt2/+}$.

### 3.5 Class switch recombination in *Exosc3* deficient B cells

Previous findings have implicated the RNA exosome complex as a key component of the CSR machinery in B cells. Using *Exosc3* shRNA knockdown studies in the mouse B cell lymphoma cell line CH12F3, it was observed that RNA exosome is required for efficient CSR and is recruited to transcribed IgH switch sequences through an AID dependent mechanism (Basu et al., 2011). In an effort to confirm these previous results and extend them to primary B cells, CSR efficiency was analyzed in *Exosc3*$^{COIN/COIN}$ B cells following Cre activation. Naive IgM$^{+}$ splenic B cells were purified from *Exosc3*$^{COIN/+}$ and *Exosc3*$^{COIN/COIN}$ mice on the inducible *ROSA26*$^{CreERt2/+}$ background. Cells were cultured in the presence of 4-OHT, and stimulated with LPS + IL-4 to induce isotype switching to IgG1 during the course of Exosc3 depletion. A ~75% reduction in IgG1 conversion was consistently observed in *Exosc3*$^{COIN/COIN}$ B cells compared to control *Exosc3*$^{COIN/+}$ (Figure 14).

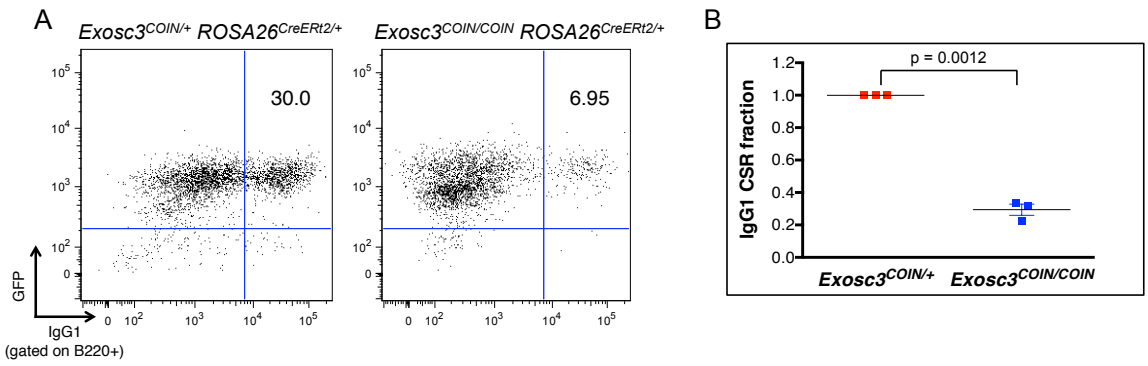



**Figure 14: *Exosc3* depletion leads to class switch recombination impairment**
(A) Representative flow cytometric analysis for surface IgG1 on purified B cells treated with 4-OHT, and stimulated with LPS + IL-4 for 72 hours. Numbers indicate the percentage of $GFP^+$ $B220^+$ B cells having isotype switched to IgG1.
(B) Quantification of IgG1 CSR defect. The efficiency of IgG1 conversion of *Exosc3*$^{COIN/COIN}$ B cells was plotted as the fraction of control *Exosc3*$^{COIN/+}$. Indicated *Exosc3* genotypes are on *ROSA26*$^{CreERt2/+}$ background. Data compiled from three biological replicates. Statistical significance was determined by Student's *t*-test.

To assess whether this observed CSR impairment is due to RNA exosome deficiency and not the result of a deficit in other mediators of CSR, AID expression and IgG1 germline transcription were evaluated in *Exosc3*$^{COIN/COIN}$ B cells. AID is specifically induced in germinal center B cells (Muramatsu et al., 1999), and is the only identified B cell specific factor absolutely required during CSR and SHM (Muramatsu et al., 2000). Likewise, transcription and processing of germline $C_H$ segment transcripts during CSR is involved in promoting generation of ssDNA substrates for AID (Keim et al., 2013). The kinetics of *Aicda* (AID) mRNA accumulation was compared between CSR stimulated *Exosc3*$^{COIN/+}$ and *Exosc3*$^{COIN/COIN}$ B cells using quantitative RT-PCR. *Aicda* induction followed similar kinetics between *Exosc3*$^{COIN/+}$ and *Exosc3*$^{COIN/COIN}$ B cells (Figure 15A). AID protein levels were found to be comparable as well (Figure 15B). In addition, nascent IgG1 germline transcript levels were elevated in *Exosc3*$^{COIN/COIN}$ B cells (Figure 15C). Therefore, differences in AID expression or $C_H$ germline transcription is unlikely to account for the CSR defect observed in *Exosc3* deficient B cells.

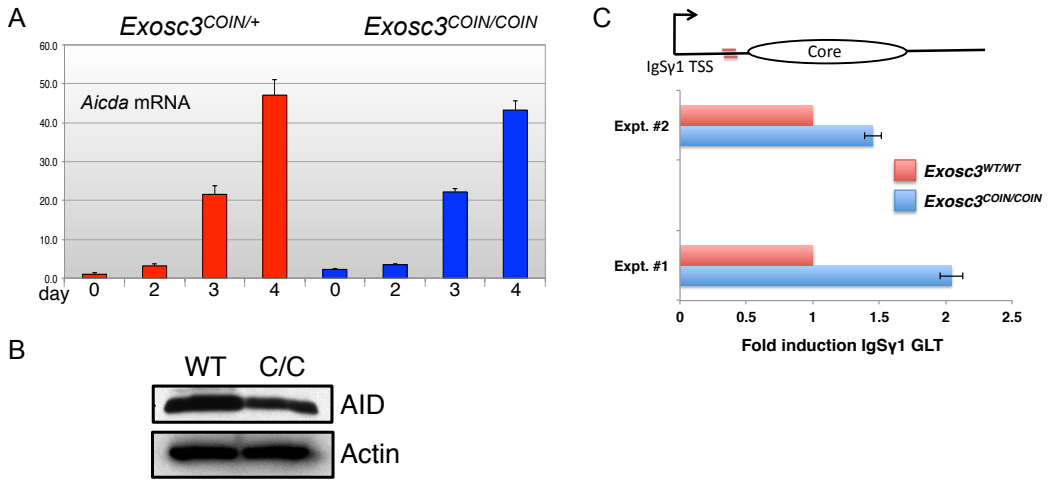



**Figure 15: AID and IgG1 germline transcript expression in CSR stimulated *Exosc3* deficient B cells**
(A) Quantitative RT-PCR time course analysis of *Aicda* mRNA expression in naive (day 0) or 4-OHT treated (days 2-4), LPS + IL-4 stimulated B cell cultures. Indicated *Exosc3* genotypes are on *ROSA26*$^{CreERt2/+}$ background. Expression levels are normalized to cyclophilin A (*Ppia*) and plotted relative to naive *Exosc3*$^{COIN/+}$.
(B) Western blot analysis of AID protein expression in whole cell extracts from 4-OHT treated, LPS + IL-4 stimulated B cell cultures. Anti-actin used as loading control. WT, *Exosc3*$^{WT/WT}$ *ROSA26*$^{CreERt2/+}$; C/C, *Exosc3*$^{COIN/COIN}$ *ROSA26*$^{CreERt2/+}$.
(C) Quantitative RT-PCR analysis of IgSγ1 intron expression. Primers were designed to amplify a region of the IgG1 germline transcript between the IgSγ1 core repeat and the CSR-specific IgG1 non-coding I exon. B cells from two independent pairs of littermate mice of each genotype were treated with 4-OHT and stimulated with LPS + IL-4. Indicated genotypes are on *ROSA26*$^{CreERt2/+}$ background. Expression levels are normalized to beta-actin and plotted relative to *Exosc3*$^{WT/WT}$.

### 3.6 Somatic hypermutation in *Exosc3* deficient B cells

To determine the role of RNA exosome in SHM, *Exosc3*$^{COIN}$ mice were bred onto a *Cd19*$^{Cre/+}$ background. The *Cd19*$^{Cre}$ allele (Rickert et al., 1997) results in Cre recombinase expression specifically in the B cell lineage, beginning early in B cell development in the bone marrow and persisting throughout maturation in the periphery including the germinal center. Peyer's patches from *Exosc3*$^{COIN/+}$ *Cd19*$^{Cre/+}$ and *Exosc3*$^{COIN/COIN}$ *Cd19*$^{Cre/+}$ mice were harvested and germinal center B cells were identified as B220$^{+}$ PNA$^{hi}$ using flow cytometry. Physiologically normal levels of germinal center B cells were present in *Exosc3*$^{COIN/+}$ *Cd19*$^{Cre/+}$ Peyer's patches (Figure 16). In contrast, *Exosc3*$^{COIN/COIN}$ *Cd19*$^{Cre/+}$ germinal center B cell numbers were markedly reduced (Figure 16). This could reflect a developmental arrest in *Exosc3* deficient cells at some stage in the B lineage preceding the germinal center and/or reduced

persistence following *Exosc3* depletion initiating at the earliest stages of development in the bone marrow.

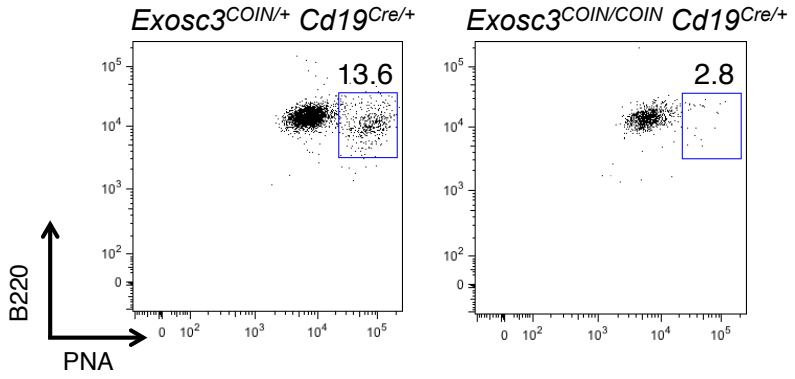



**Figure 16: *Exosc3* depletion in early B cell development leads to loss of germinal center B cells**
Flow cytometric analysis of Peyer's patch derived germinal center B cells from *Exosc3*$^{COIN/+}$ and *Exosc3*$^{COIN/COIN}$ mice on *Cd19*$^{Cre/+}$ background were identified as B220$^{+}$ PNA$^{hi}$ populations. The percentage of germinal center B cells amongst all GFP$^{+}$ B220$^{+}$ cells is indicated.

A more conservative approach to generating *Exosc3* deficient germinal center B cells was subsequently adopted utilizing an *Aicda*$^{Cre}$ allele (Figure 17A). This allele is expected to deliver Cre recombinase expression during later stages of B cell development, specifically during the germinal center reaction. However, there are existing reports of AID expression occurring in male and female germ cells (Morgan et al., 2004; Schreck et al., 2006), which could potentially result in *Exosc3*$^{COIN}$ inversion in a substantial fraction of cells including early B cell progenitors. Since *Exosc3* is expressed in naive splenic B cells, the GFP reporter of *Exosc3*$^{COIN}$ was utilized as a lineage tracer in *Exosc3*$^{COIN/+}$ *Aicda*$^{Cre/+}$ mice to determine whether *Aicda* expression had occurred in precursor cells. GFP fluorescence in B220$^{+}$ B cells as well as other non-B lineage cells of the spleen was rare in *Exosc3*$^{COIN/+}$ *Aicda*$^{Cre/+}$ mice (Figure 17B), suggesting that *Aicda* expression is limited in B cell precursors. However, stimulation of naive B cells with LPS + IL-4, conditions capable of inducing *Aicda* expression (Figure 15A), lead to

GFP production in *Exosc3*$^{COIN/+}$ *Aicda*$^{Cre/+}$ B cells (Figure 17C), confirming that *Aicda*$^{Cre}$ can be properly induced through physiologically relevant stimuli.

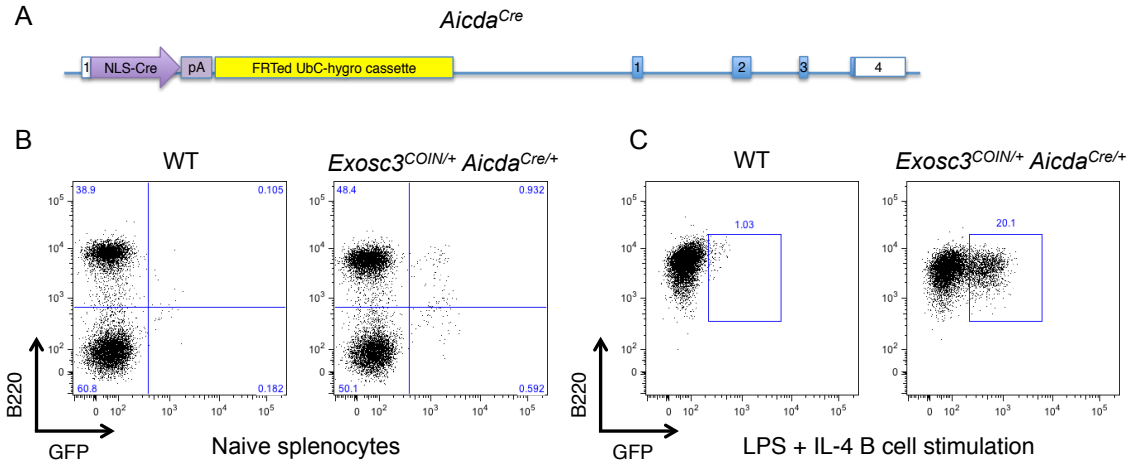
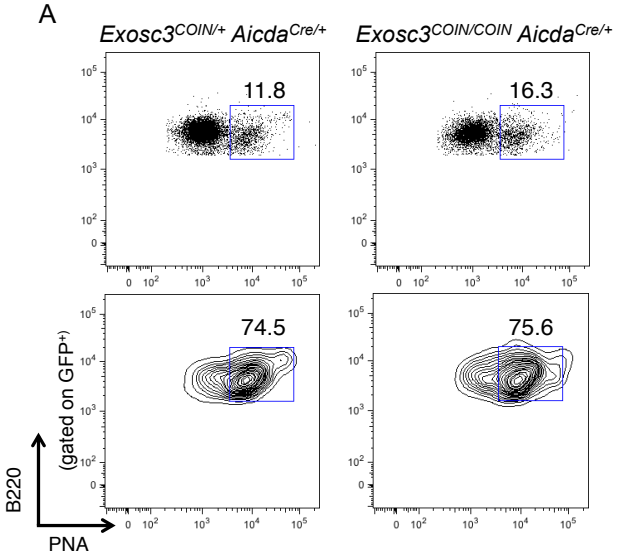



**Figure 17: Characterization of the *Aicda*$^{Cre}$ allele**
(A) Schematic of the targeted *Aicda*$^{Cre}$ allele. An open reading frame comprised of a nuclear localization signal fused to Cre recombinase was used to disrupt the ATG start codon in exon 1 of *Aicda*. Exons are represented as numbered boxes.
(B) Flow cytometric analysis of *Aicda*$^{Cre}$ activity (as determined by GFP fluorescence) in B220$^{+}$ and B220$^{-}$ naive splenocyte populations.
(C) *Aicda*$^{Cre}$ induction in LPS + IL-4 stimulated B cell cultures.

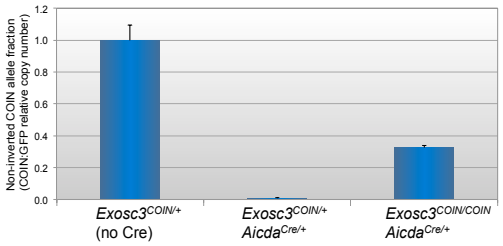
*Exosc3*$^{COIN/+}$ mice were subsequently bred to homozygosity on an *Aicda*$^{Cre/+}$ background and germinal center B cell production was assessed. *Exosc3*$^{COIN/COIN}$ *Aicda*$^{Cre/+}$ mice displayed robust numbers of Peyer's patch derived germinal center B cells, with a moderate increase in relative cell number compared to *Exosc3*$^{COIN/+}$ *Aicda*$^{Cre/+}$ mice (Figure 18A). Furthermore, this increase in germinal center B cells was not due to differential *Aicda*$^{Cre}$ induction as both *Exosc3*$^{COIN/+}$ and *Exosc3*$^{COIN/COIN}$ germinal center B cell fractions displayed very similar GFP fluorescence (Figure 18A). Interestingly, this observed increase in germinal center B cell accumulation phenocopies other AID mutant mouse models exhibiting SHM deficits (McBride et al., 2008; Muramatsu et al., 2000). *Exosc3*$^{COIN/+}$ *Aicda*$^{Cre/+}$ and *Exosc3*$^{COIN/COIN}$ *Aicda*$^{Cre/+}$ germinal center B cells were collected and genomic DNA was prepared. To evaluate *Exosc3*$^{COIN}$ inversion efficiency in these samples, genomic copy number analysis was performed by

comparing the ratio of *Exosc3*$^{COIN}$ : GFP using quantitative PCR. Specifically, the *Exosc3*$^{COIN}$ amplicon was derived from a segment of *Exosc3*$^{COIN}$ that is lost during the process of allele inversion, namely the region between the *lox*2372 and *lox*P sites of the 3' *lox* array (Figure 6). An *Exosc3*$^{COIN}$ allele, prior to inversion would have an *Exosc3*$^{COIN}$ : GFP ratio of 1, whereas an inverted *Exosc3*$^{COINinv}$ allele would have an *Exosc3*$^{COIN}$ : GFP ratio of 0. Using this approach it was determined that ~70% of *Exosc3*$^{COIN}$ alleles were inverted in *Exosc3*$^{COIN/COIN}$ *Aicda*$^{Cre/+}$ germinal center B cells (Figure 18B), indicating the presence of a mixed population of homozygous null Exosc3$^{COINinv/COINinv}$ and heterozygous Exosc3$^{COINinv/COIN}$ genotypes. Allele inversion appeared complete in *Exosc3*$^{COIN/+}$ *Aicda*$^{Cre/+}$ cells (Figure 18B). SHM was evaluated in these same samples by determining the mutation frequency within a defined 645 bp intronic sequence located downstream of the $J_H4$ gene segment of the *IgH* locus (Jolly et al., 1997). Total mutation frequency in *Exosc3*$^{COIN/COIN}$ *Aicda*$^{Cre/+}$ germinal center B cells was reduced to 62% of control *Exosc3*$^{COIN/+}$ *Aicda*$^{Cre/+}$ (p = 7.23E-13) (Figure 18C). This SHM defect was exacerbated at CG base pairs, which are enriched for direct sites of AID action. CG mutation frequency in *Exosc3*$^{COIN/COIN}$ *Aicda*$^{Cre/+}$ cells was reduced to 53% of control *Exosc3*$^{COIN/+}$ *Aicda*$^{Cre/+}$ (p = 4.44E-16) (Figure 18C).

A




B




C

| | Exosc3$^{COIN/+}$ Aicda$^{Cre/+}$ | Exosc3$^{COIN/COIN}$ Aicda$^{Cre/+}$ | p-value |
|---|---|---|---|
| Total clones analyzed | 197 | 203 | |
| Total bases analyzed | 127,065 | 130,935 | |
| Number of mutations | 2,096 | 1,354 | |
| Total mutation frequency | 1.65E-2/bp | 1.03E-2/bp | 7.23E-13 |
| CG mutation frequency | 1.36E-2/CG | 0.72E-2/CG | 4.44E-16 |
| AT mutation frequency | 1.85E-2/AT | 1.26E-2/AT | 1.39E-09 |

**Figure 18: Reduced somatic hypermutation efficiency in *Exosc3* deficient germinal center B cells**
(A) Flow cytometric analysis of Peyer's patch derived germinal center B cells. *Exosc3*$^{COIN/+}$ and *Exosc3*$^{COIN/COIN}$ germinal center B cells on *Aicda*$^{Cre/+}$ background were identified as B220$^{+}$ PNA$^{hi}$ populations. The percentage of germinal center B cells amongst all B220$^{+}$ cells is indicated.
(B) *Exosc3*$^{COIN}$ inversion efficiency in germinal center B cells as determined by *Exosc3*$^{COIN}$ : GFP genomic copy number analysis using quantitative PCR.
(C) Somatic hypermutation analysis of Peyer's patch derived germinal center B cells. Mutation frequencies were determined by sequencing a 645 bp intronic region downstream of the $J_H4$ gene segment of the immunoglobulin heavy chain (IgH) locus. P-values were determined by proportion test.

The proliferative capacity of *Exosc3*$^{COIN/+}$ *Aicda*$^{Cre/+}$ and *Exosc3*$^{COIN/COIN}$ *Aicda*$^{Cre/+}$ B cells was evaluated ex vivo by monitoring the kinetics of GFP induction in these cells and dilution of an intracellular fluorescent dye. When stimulated with LPS + IL-4 these cells produce distinct GFP$^{+}$ and GFP$^{-}$ populations (Figure 19A). *Exosc3*$^{COIN/+}$ *Aicda*$^{Cre/+}$ and *Exosc3*$^{COIN/COIN}$ *Aicda*$^{Cre/+}$ B cell cultures have comparable percentages of GFP$^{+}$ cells after 2.5 days (Figure 19A). When assayed 24 hours later (day 3.5), a comparable increase in GFP$^{+}$ *Exosc3*$^{COIN/+}$ and *Exosc3*$^{COIN/COIN}$ cells was observed (Figure 19B), providing evidence that these cells proliferate at approximately similar rates. Additionally, VPD450 dye dilution assays demonstrated comparable rates of cell division between *Exosc3*$^{COIN/+}$ and *Exosc3*$^{COIN/COIN}$ B cells (Figure 19C), despite a 10-fold reduction in *Exosc3* mRNA levels in *Exosc3*$^{COIN/COIN}$ B cells (Figure 19D).

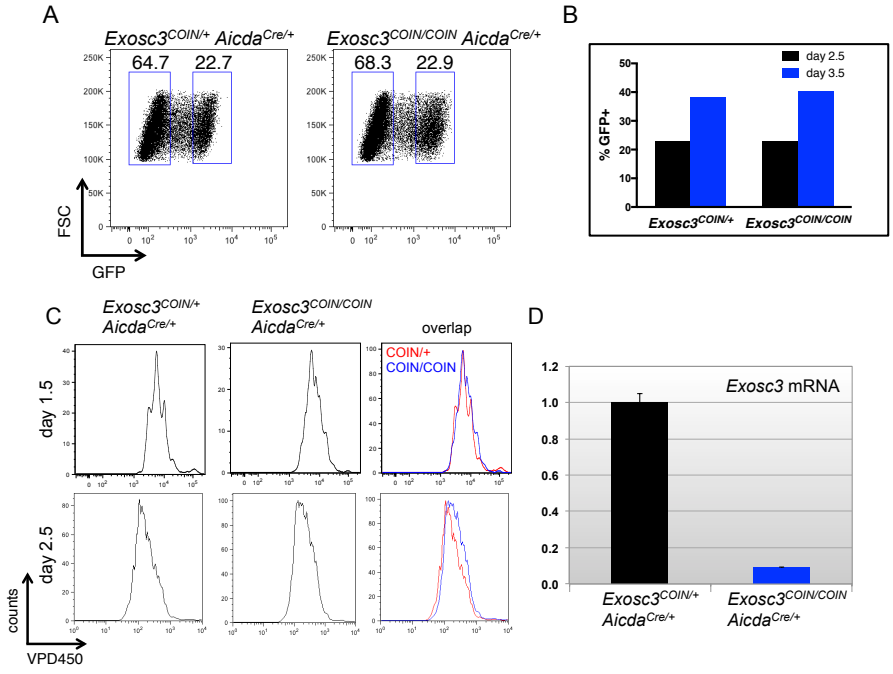


**Figure 19: Ex vivo proliferation analysis of Exosc3 deficient B cells**
(A) FACS analysis indicating the percentage of GFP negative (left gate) and GFP positive (right gate) B cells 2.5 days post LPS stimulation.
(B) Kinetic analysis of GFP positive B cell accumulation at indicated time points post LPS stimulation. Indicated *Exosc3* genotypes are on *Aicda*$^{Cre/+}$ background.
(C) Proliferation analysis determined by VPD450 dye dilution at 1.5 and 2.5 days post LPS stimulation.
(D) Quantitative RT-PCR analysis of *Exosc3* mRNA expression in GFP positive cells at 3.5 days post LPS stimulation. Expression levels are normalized to beta-actin and plotted relative to *Exosc3*$^{COIN/+}$.

It should be noted that while these differences in SHM rates are significant, they likely underrepresent the complete effect of RNA exosome depletion on SHM due to the simultaneous induction of Cre and AID expression in these cells. A temporal window exists between the onset of *Exosc3*$^{COIN}$ inversion and when the effects of RNA exosome deficiency begin to manifest, which is dependent upon *Exosc3* mRNA and protein turnover rates. During this window AID can function in a relatively physiological setting in *Exosc3*$^{COIN/COIN}$ *Aicda*$^{Cre/+}$ cells.

### 3.7 Characterization of the *Exosc3* deficient B cell transcriptome

RNA exosome mediated degradation and/or processing of substrate RNAs is highly efficient within cells. In most cases substrate concentrations in wild type cells are constrained to levels approaching the detections limits of conventional methodologies. Typically RNA exosome substrates are revealed following perturbation of RNA exosome function. Prior analyses of mammalian RNA exosome substrates have relied upon shRNA mediated knockdown of RNA exosome subunits (Flynn et al., 2011; Preker et al., 2008). However, shRNA approaches can be confounded by indeterminate combinations of incomplete knockdown of the intended target gene as well as partial knockdown of unintended genes. Gene targeting approaches implementing well defined genetic changes often circumvent these inherent limitations of shRNA. Therefore, we sought to leverage the *Exosc3*$^{COIN}$ mouse model as a novel tool providing for a more specific and fuller characterization of RNA exosome substrate RNA in mammalian cells. In addition, we hypothesized that a genome-wide transcriptome analysis of *Exosc3* deficient B cells

uncovering RNA exosome substrates, and by extension sites of RNA exosome activity, could provide additional insights into the nature of AID target genes during Ig diversification.

RNA from *Exosc3* deficient B cells reported in the CSR studies above were used for genome-wide transcriptome analyses. Specifically, *Exosc3*$^{COIN/COIN}$ along with control *Exosc3*$^{COIN/+}$ or *Exosc3*$^{WT/WT}$ B cells on a *ROSA26*$^{CreERt2/+}$ background were treated with 4-OHT and stimulated with LPS + IL-4 for 72 hours. Total RNA was extracted from GFP$^{+}$ B cells (for *Exosc3*$^{COIN}$ genotypes), depleted of ribosomal RNA, and processed for next-gen shotgun sequencing (RNA-seq). In short, resulting RNA reads were mapped onto the mouse mm9 reference genome using the TopHat algorithm (Trapnell et al., 2009) and whole transcriptome assembly was performed with Cufflinks (Trapnell et al., 2010). To validate this RNA-seq data pipeline we used the Integrative Genomics Viewer (IGV) browser (Robinson et al., 2011) to visually assess the mapping of RNA reads to the *Exosc3* locus. The *Exosc3*$^{COIN}$ inversion scheme predicts a very distinctive *Exosc3* expression signature whereby allele inversion results in loss of expression corresponding to exons 2-4, while expression of exon 1 remains unperturbed (Figure 6). The predicted pattern of *Exosc3* exon expression was observed in the mapping of *Exosc3*$^{COIN/COIN}$ RNA-seq reads, as visualized through the IGV browser (Figure 20). This mapping data strongly supports the *Exosc3*$^{COIN}$ inversion scheme that was initially designed. In turn, the specific exclusion of *Exosc3*$^{COIN/COIN}$ RNA-seq reads mapping to *Exosc3* exons 2-4 served as a critical internal control towards validating our bioinformatics work scheme.

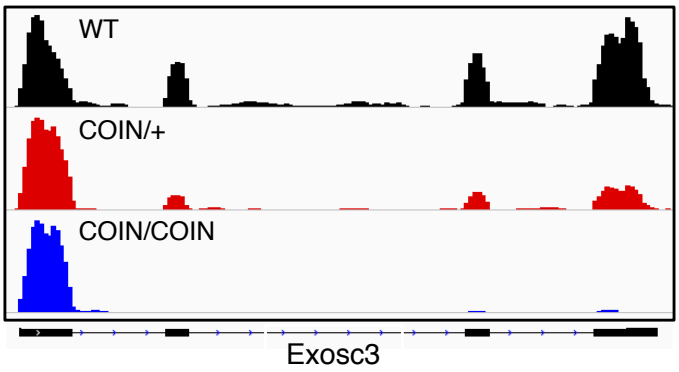



**Figure 20: RNA-seq confirms mechanism of *Exosc3*$^{COIN}$ inversion leading to ablation of *Exosc3* expression**
IGV profile of RNA-seq mapped reads at the *Exosc3* locus from 4-OHT treated, LPS + IL-4 stimulated B cell cultures. *Exosc3* exons 1-4 are indicated from left to right as black rectangles. Indicated *Exosc3* genotypes are on *ROSA26*$^{CreRT2/+}$ background.

In order to uncover broad, fundamental changes in transcriptional regulation resulting from RNA exosome deficiency, we performed differential expression analysis between *Exosc3*$^{WT/WT}$ and *Exosc3*$^{COIN/COIN}$ transcriptomes according to transcript families (Figure 21). A list of ~25,000 annotated mouse transcripts including mRNAs, lncRNAs, and snoRNAs was assembled from available databases (Guttman et al., 2009; Karolchik et al., 2014; Yoshihama et al., 2013). Differential expression analysis confirmed that global levels of certain ncRNA species were significantly upregulated in the *Exosc3*$^{COIN/COIN}$ transcriptome. snoRNA, which have been previously reported as targets of RNA exosome in yeast (Allmang et al., 1999a), were indeed upregulated in *Exosc3* deficient B cells (p = 8E-10) (Figure 21). Another class of ncRNA upregulated in *Exosc3* deficient B cells, not previously described as RNA exosome targets, was lncRNA (p = 2E-27) (Figure 21). Intriguingly, global mRNA levels were not upregulated in the *Exosc3*$^{COIN/COIN}$ transcriptome (Figure 21). Early reports provided evidence that mature mRNAs can be degraded through the combined actions of decapping coupled 5'-3' Xrn1 mediated degradation (Hsu and Stevens, 1993) in conjunction with RNA exosome mediated 3'-5' degradation (Anderson and Parker, 1998). These two pathways of mRNA degradation are partially redundant (Schaeffer et al., 2010). However, the relative

stabilization of specific mRNAs in mutants of Xrn1 or cytoplasmic RNA exosome function suggest that decapping coupled 5'-3' mediated degradation is the dominant pathway of mRNA turnover (Schaeffer et al., 2010). However, it is possible that RNA exosome function may be limited to specific mRNA types. A subset of mRNAs with particularly short half-lifes such as those possessing 3'UTR AU-rich elements (AREs) may be specifically tagged for rapid RNA exosome mediated degradation (Chen et al., 2001). Consistent with our data indicating a lack mRNA stabilization in the *Exosc3*$^{COIN/COIN}$ transcriptome (Figure 21), recent reports indicate that mammalian DIS3L2, a homolog of DIS3, functions in cytoplasmic 3'-5' degradation of mRNA independently of RNA exosome (Lubas et al., 2013; Malecki et al., 2013).

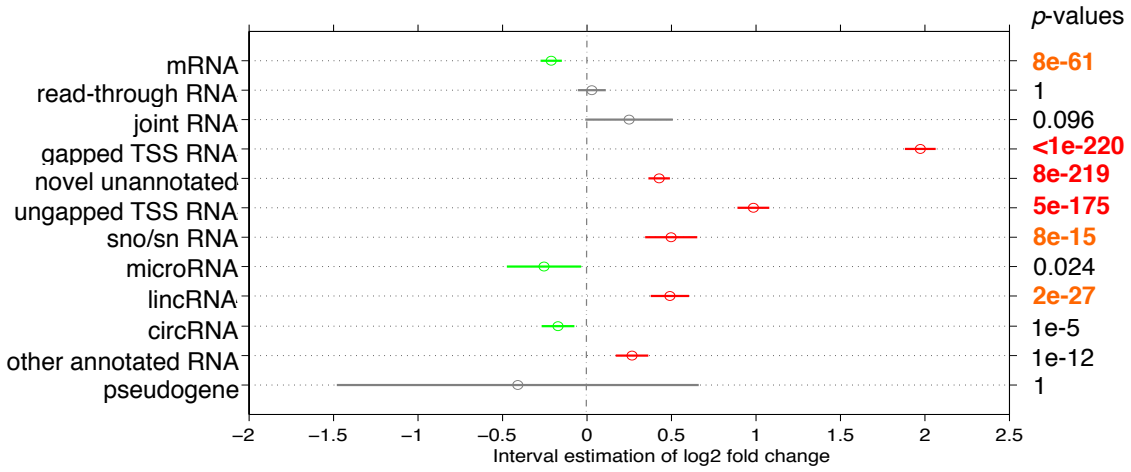



**Figure 21: Differential expression analysis of transcript classes between *Exosc3*$^{COIN/COIN}$ and *Exosc3*$^{WT/WT}$**
The horizontal bar for each transcript type represents the 95% confidence interval (calculated by Student's *t*-distribution after a Bonferroni adjustment) of a base-2 logarithmic fold change between *Exosc3*$^{COIN/COIN}$ and *Exosc3*$^{WT/WT}$ RPKM. Circles indicate mean values. Red and green bars indicate transcript types expressed at significantly higher or lower levels within the *Exosc3*$^{COIN/COIN}$ transcriptome, respectively.

Transcripts mapping to regions upstream of annotated coding gene TSSs constituted the most highly upregulated RNA class observed in the *Exosc3*$^{COIN/COIN}$ B cell transcriptome (Figure 21). These TSS proximal transcripts could be separated into two classes according to their sequence overlap with cognate downstream mRNA. Ungapped TSS-RNAs were defined as TSS upstream transcripts that's assembled RNA-

seq reads were found to overlap with cognate downstream coding transcripts. In contrast, gapped upstream TSS-RNAs do not overlap with cognate downstream coding transcripts and thereby display distinct regions of non-overlap between TSS-RNA and coding transcripts during assembly of mapped RNA-seq reads. We determined that approximately 3,144 genes expressed detectable levels of TSS upstream transcripts in *Exosc3* wild type and *Exosc3* deficient B cells collectively. The bioinformatic analysis revealed 1,127 gapped and 832 ungapped TSS-RNA transcripts displaying $\log_2$ RPKM fold changes greater than 1 in *Exosc3* deficient B cells compared to *Exosc3* wild type B cells. This group of RNA exosome regulated transcripts was collectively labeled as "xTSS-RNA". In *Exosc3* wild type B cells approximately 70% of xTSS-RNA loci display trace levels of expression as indicated by RPKM values less than 1 (Figure 22A). In contrast, upwards of 80% of xTSS-RNA loci displayed RPKM values greater than 1 in *Exosc3* deficient B cells, indicating significant levels of expression (Figure 22A). These data indicate that RNA exosome deficiency leads to a global stabilization of xTSS-RNA. Incidentally, *Hspa9* is illustrative of a locus containing two distinct RNA exosome substrates, namely an upstream xTSS-RNA transcript and an intragenic snoRNA homolog of human *SNORD63* (Figure 22B). In addition, xTSS-RNA and cognate downstream mRNA expression levels correlated poorly (Figure 22C), suggesting that xTSS-RNA and mRNA transcripts may be regulated independently. This result is consistent with the observed lack of mRNA stabilization genome-wide at coding genes in the *Exosc3*$^{COIN/COIN}$ B cell transcriptome (Figure 21).

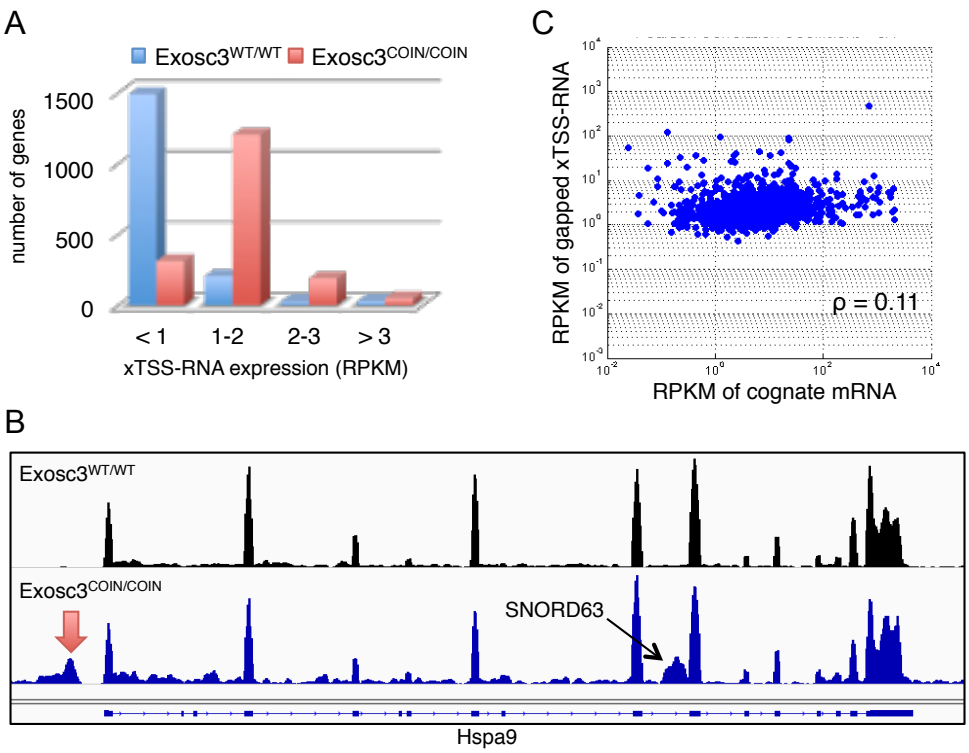



**Figure 22: Stabilization of xTSS-RNA in *Exosc3* deficient B cells**
(A) Distribution of xTSS-RNA RPKM expression levels in *Exosc3*$^{WT/WT}$ and *Exosc3*$^{COIN/COIN}$ B cells.
(B) Representative genomic locus displaying both xTSS-RNA and snoRNA upregulation. IGV profile of RNA-seq mapped reads at the *Hspa9* locus (20 kb window). Red arrow indicates the *Hspa9* xTSS-RNA. SNORD63, a snoRNA expressed within the *Hspa9* locus, is indicated.
(C) Scatter plot of xTSS-RNA and cognate mRNA expression levels in *Exosc3* deficient B cells at individual loci indicating weak correlation. Pearson correlation coefficient is indicated.

Compiling genome wide expression data centered 1 kilobase upstream and downstream to the TSS of coding genes illustrates the steady state accumulation of xTSS-RNA in the TSS upstream region in *Exosc3* deficient B cells (Figure 23A). In such global gene expression analyses, xTSS-RNA as a family of transcripts are nearly undetectable in wild type B cells in the steady state (Figure 23A). Average transcript length for xTSS-RNAs in *Exosc3* deficient B cells was approximately 700 nucleotides (Figure 23B,C). The *Exosc3*$^{COIN/COIN}$ transcriptome was used to guide the annotation of individual xTSS-RNA transcripts since their stabilization in *Exosc3*$^{COIN/COIN}$ permitted a more robust determination of their transcription start and end sites. These annotations were then applied onto the *Exosc3*$^{WT/WT}$ transcriptome. Although expressed at far lower

levels in *Exosc3*$^{WT/WT}$, assembly of RNA-seq reads at xTSS-RNA loci revealed an average xTSS-RNA transcript length of 600 nucleotides in *Exosc3* wild type B cells (Figure 23B). The observed 100 nucleotide difference in xTSS-RNA length in wild type and *Exosc3* deficient cells, while minimal, proved to be statistically significant (p = 2.3E-5) (Figure 23B) and may indicate impaired transcription termination in the absence of RNA exosome coupled xTSS-RNA degradation. The distribution of observed xTSS-RNA lengths in *Exosc3*$^{COIN/COIN}$ was limited to less than 1.5 kb (Figure 23C). Characterization of the gap, or non-overlapping sequence between xTSS-RNA and cognate downstream mRNA, revealed a distribution of gap lengths largely between 100 and 500 bp (Figure 23D). This result strongly suggested that xTSS-RNAs and mRNAs possess distinct TSSs at individual loci.

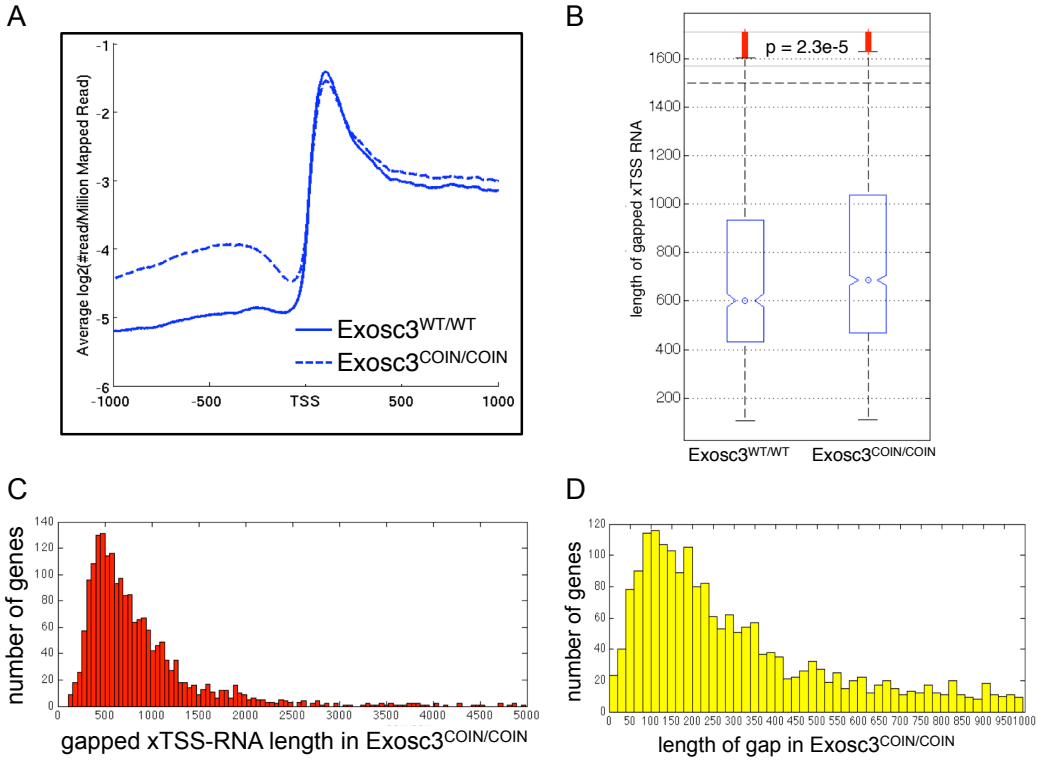



**Figure 23: Characterization of xTSS-RNA distribution around TSSs**
(A) Genome-wide expression level analysis 1 kb upstream and downstream of TSS region for expressed protein coding genes. Coding genes with RPKM > 1 were classified as expressed. Analysis was restricted to coding genes that do not have any known genes within a 4 kb upstream boundary. Indicated genotypes are on *ROSA26*$^{CreERt2/+}$ background.

(B) Boxplot analysis indicating the length of gapped xTSS-RNAs in wild type and *Exosc3* deficient primary B cell cultures. Data represent median values compiled from two biological replicates. Whiskers represent 99% of data values. P-value determined by Wilcoxon rank sum test.
(C) The distribution of observed lengths for all gapped xTSS-RNAs in *Exosc3* deficient B cells. Data compiled from two biological replicates.
(D) The distribution of the lengths of the gap between xTSS-RNA and cognate coding transcript. Data compiled from two biological replicates.

Strand specific RNA sequencing was performed on *Exosc3* wild type and *Exosc3* deficient B cells in order to further characterize xTSS-RNAs for their transcriptional orientation. xTSS-RNA transcription could occur either divergently or in tandem relative to cognate coding transcripts. Genome-wide mapping of stranded RNA-seq reads located within 2 kb upstream and downstream of coding gene TSSs revealed that the great majority of TSS upstream transcripts observed in *Exosc3*$^{WT/WT}$ and *Exosc3*$^{COIN/COIN}$ cells are antisense relative to cognate mRNA transcripts (Figure 24). Therefore, xTSS-RNA and mRNA transcript pairs are largely transcribed in a divergent orientation.

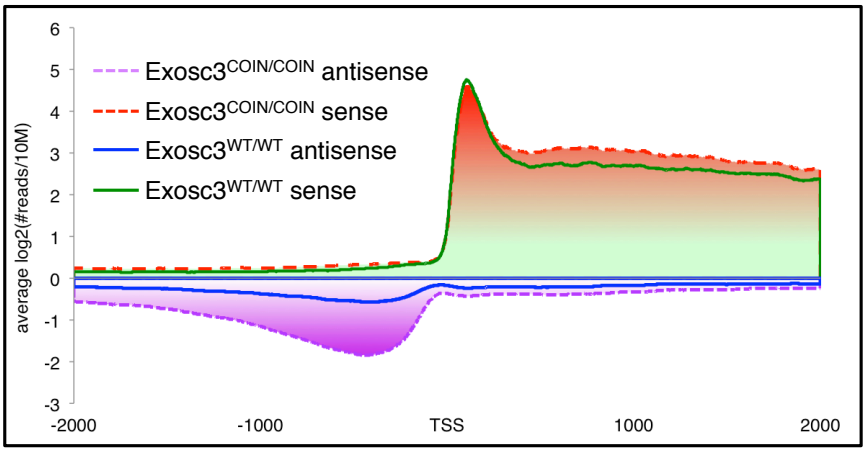



**Figure 24: Antisense nature of xTSS-RNA**
Strand specific distribution of B cell RNA-seq mapped reads 2 kb upstream and downstream of expressed coding gene TSSs. Indicated genotypes are on *ROSA26*$^{CreERt2/+}$ background and cells were treated with 4-OHT and stimulated with LPS and IL-4. Data compiled from two biological replicates.

Transcription profiling reproducibility in *Exosc3*$^{WT/WT}$ and *Exosc3*$^{COIN/COIN}$ B cells was directly examined through statistical means. Independent biological replicates involving B cells derived from separate littermate pairs of *Exosc3*$^{WT/WT}$ and *Exosc3*$^{COIN/COIN}$ mice on a *ROSA26*$^{CreERt2/+}$ background were treated with 4-OHT and stimulated with LPS + IL-4. Live cells were specifically isolated using flow cytometry and

RNA was processed for RNA sequencing as previously described. The two independent sets of gene expression data for both *Exosc3*$^{WT/WT}$ and *Exosc3*$^{COIN/COIN}$ B cells were plotted against one another. This analysis revealed a high degree of statistical correlation between replicate data sets for both *Exosc3*$^{WT/WT}$ and *Exosc3*$^{COIN/COIN}$ B cells (Figure 25), thus highlighting the robustness of the transcriptional phenotype observed in *Exosc3* deficient cells.

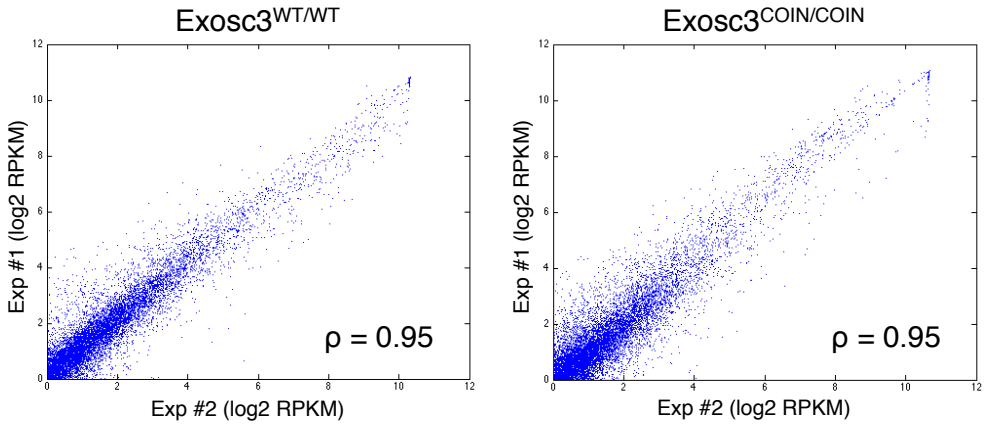



**Figure 25: Replicate analysis of *Exosc3*$^{WT/WT}$ and *Exosc3*$^{COIN/COIN}$ genome-wide transcription profiling** Scatter plots indicating expression levels of individual genes in *Exosc3*$^{WT/WT}$ and *Exosc3*$^{COIN/COIN}$ B cells treated with 4-OHT and stimulated with LPS + IL-4 from two separate littermate pairs. B cells were purified, cultured, FACS sorted, and RNA was purified and sequenced by RNA-seq all independently between the two experiments. Indicated genotypes are on *ROSA26*$^{CreERt2/+}$ background. Pearson correlation is indicated.

### 3.8 RNA exosome substrate ncRNAs mark AID dependent B cell translocation hotspots genome-wide

Although genome-wide transcriptome analyses reveal a stark accumulation of xTSS-RNA transcripts in *Exosc3* deficient B cells, this phenotype is not universally observed across all individual genes. In fact many genes display little to no xTSS-RNA accumulation in *Exosc3*$^{COIN/COIN}$ B cells. Such loci include the genes for beta-actin (*Actb*), interleukin-2 receptor common gamma chain (*Il2rg*), and uracil DNA glycosylase (*Ung*) (Figure 26) to name a few. This data suggests that differential RNA exosome recruitment may occur at various RNAP II dependent promoters. Since both RNA exosome and AID utilize co-transcriptional mechanisms of RNA degradation and DNA deamination,

respectively, we hypothesized whether RNA exosome activity could be detected at AID target genes in the form of xTSS-RNA. Several previously reported AID target genes, including *IgG1*, *Myc*, *Cd83*, and *Pim1* were observed to accumulate significant levels of xTSS-RNA in *Exosc3*$^{COIN/COIN}$ B cells (Figure 27). This indicates that RNA exosome is differentially recruited to the promoter proximal region for a subset of genes, of which includes AID target genes *IgG1*, *Myc*, *Cd83*, and *Pim1*.

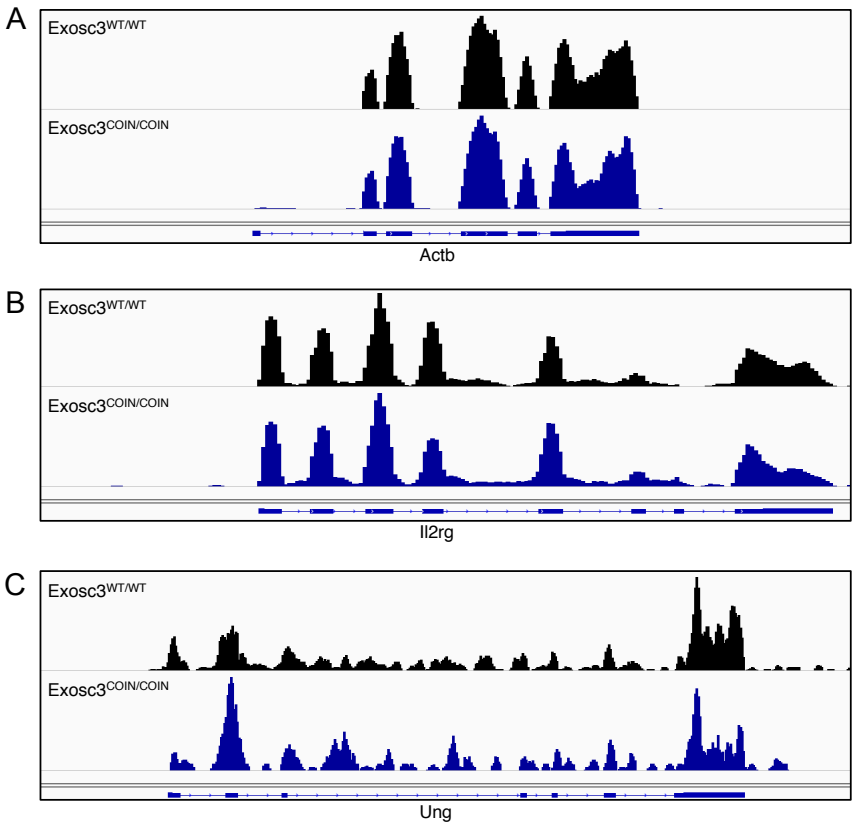



**Figure 26: Absence of xTSS-RNA at three representative loci**
(A-C) IGV profile of RNA-seq mapped reads at the beta actin locus (A) (*Actb*; 7.6 kb window), *Il2rg* locus (B) (5.4 kb window) and *Ung* locus (C) (12 kb window). Direction of transcription at each locus is from left to right. Annotated exons are represented by blue rectangles. Indicated genotypes are on *ROSA26*$^{CreERt2/+}$ background and B cell cultures were treated with 4-OHT and stimulated with LPS + IL-4.

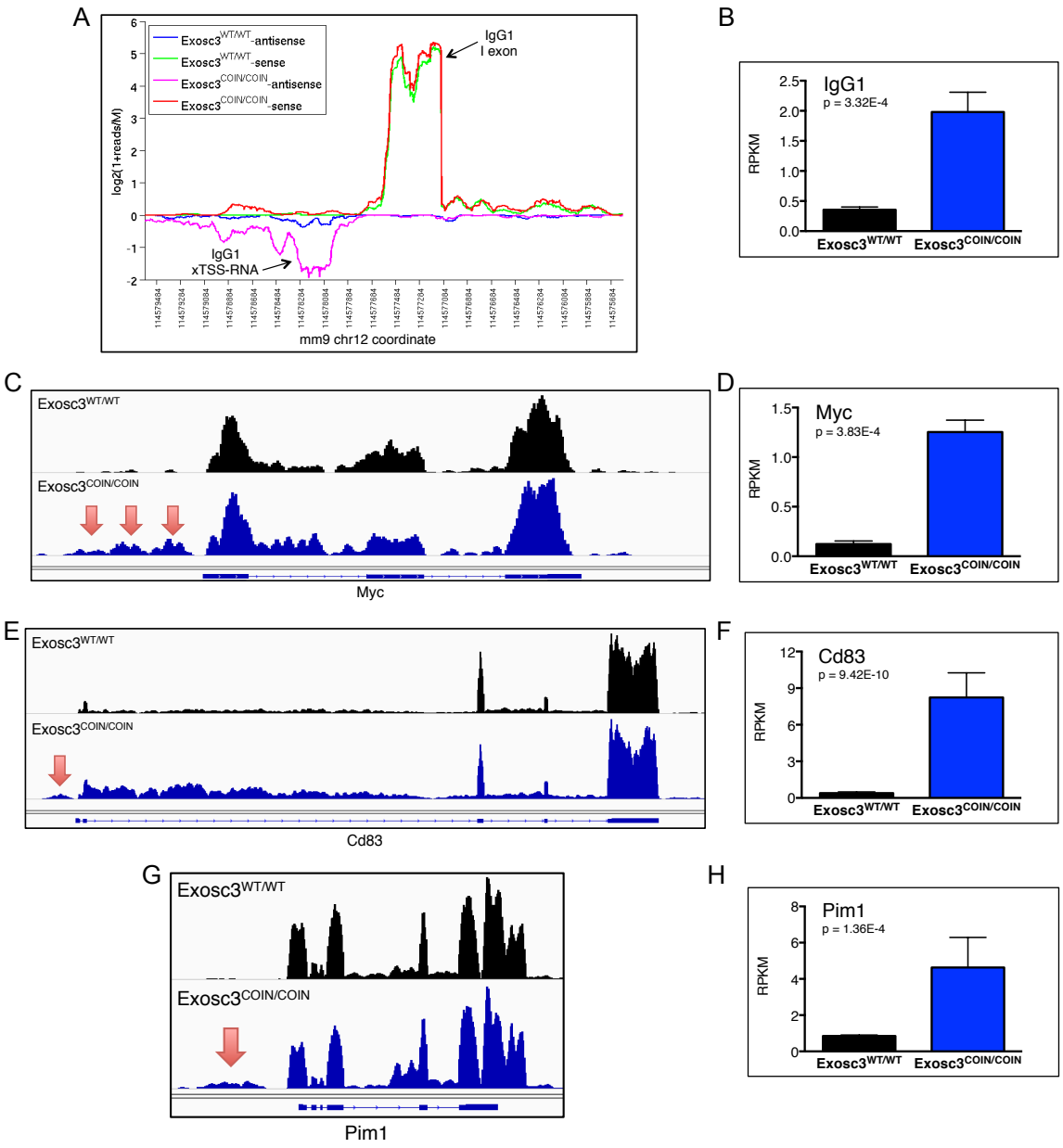



**Figure 27: RNA exosome activity at AID target gene *IgG1*, *Myc*, *Cd83*, and *Pim1* promoter proximal regions**

(A) Strand specific RNA-seq mapped reads for *Exosc3*$^{WT/WT}$ and *Exosc3*$^{COIN/COIN}$ 2 kb upstream and downstream of the CSR-specific *IgG1* germline transcript TSS.
(B) Quantification of *IgG1* xTSS-RNA expression levels from RNA-seq derived RPKM values.
(C) IGV profile of RNA-seq mapped reads at a 9 kb window containing the *Myc* locus. Red arrows highlighting *Myc* xTSS-RNA reads.
(D) Quantification of *Myc* xTSS-RNA expression levels from RNA-seq derived RPKM values.
(E) IGV profile of RNA-seq mapped reads at a 22 kb window containing the *Cd83* locus. Red arrows highlighting *Cd83* xTSS-RNA reads.
(F) Quantification of *Cd83* xTSS-RNA expression levels from RNA-seq derived RPKM values.
(G) IGV profile of RNA-seq mapped reads at a 9 kb window containing the *Pim1* locus. Red arrows highlighting *Pim1* xTSS-RNA reads.
(H) Quantification of *Pim1* xTSS-RNA expression levels from RNA-seq derived RPKM values.
All indicated genotypes are on *ROSA26*$^{CreERt2/+}$ background and purified B cell cultures were treated with 4-OHT and stimulated with LPS + IL-4.

As described earlier, TSSs have been implicated in genome-wide studies of AID dependent chromosomal translocations in mouse B cells (Chiarle et al., 2011; Klein et al., 2011b). These studies revealed that while AID dependent chromosomal translocations occur throughout the genome, their frequency is elevated within TSS proximal regions of active genes. In comparison to silent chromatin, actively transcribed regions of the genome are likely to generate more frequent ssDNA intermediates that may serve as substrates of AID. While the generally more accessible nature of actively transcribed chromatin is likely to contribute somewhat to the enrichment of chromosomal translocations to TSS proximal regions, it is likely that this effect is limited and possibly mediated through stochastic interactions between AID and chromatin. In support of this notion is the observation that many heavily transcribed genes are not involved as partners of AID dependent chromosomal translocations in genome-wide studies (Chiarle et al., 2011; Klein et al., 2011b). Similarly, analysis of the *Exosc3* deficient B cell transcriptome revealed disconnects at many genes between degrees of xTSS-RNA accumulation and downstream mRNA expression levels (Figure 22C). For instance, *IgG1* and *Actb* constitute two of the most strongly transcribed loci in stimulated B cells. However, *IgG1* revealed a strong accumulation of TSS proximal xTSS-RNA (Figure 27A), whereas *Actb* displayed negligible levels of xTSS-RNA expression (Figure 26A). We therefore compared a list of 1,774 genes displaying robust xTSS-RNA expression with 88 genes previously reported to undergo recurrent AID dependent chromosomal translocations in stimulated B cells (Klein et al., 2011b). Despite representing less than 10% of annotated coding genes in the mouse genome, the xTSS-RNA gene list overlapped with nearly half of reported AID dependent translocation hotspots ($p < 0.0001$) (Figure 28A). Specifically, 40 genes were identified as recurrent translocation sites also containing significant xTSS-RNA expression (Figure 28A). In order to further validate this correlation a statistical bootstrapping analysis was performed. Using our B

cell transcriptome data we mathematically modeled 10,000 distinct sets of 88 genes containing similar transcription levels as the 88 translocation hotspot genes set. This analysis revealed that random chance on average would select only 15 genes from the xTSS-RNA gene list (Figure 28B). Therefore, the overlap of 40 genes from the translocation hotspot list within the xTSS-RNA gene list represents a valid enrichment correlation (p = 4.6E-10) (Figure 28B) between xTSS-RNA expression and translocation recurrence.

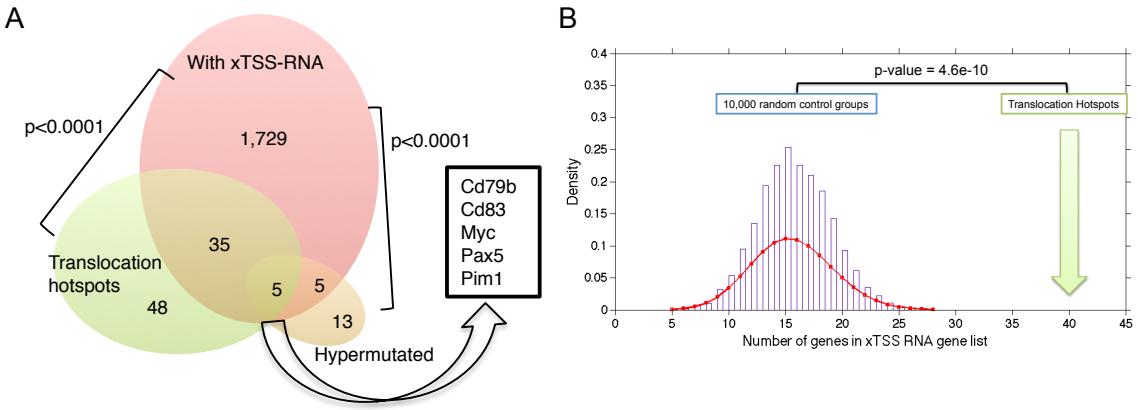



**Figure 28: xTSS-RNA enrichment at sites of recurrent AID dependent chromosomal translocations**
(A) Venn diagram of genes with xTSS-RNA transcripts (pink), genes that undergo recurrent AID dependent chromosomal translocations (green) (Klein et al., 2011b) and genes undergoing somatic hypermutation in the mouse B cell genome (brown) (Liu et al., 2008). A set of 40 genes undergo recurrent AID dependent chromosomal translocations and express xTSS-RNAs, of which 5 (*Myc*, *Cd79b*, *Cd83*, *Pax5*, and *Pim1*) have been shown to be directly hypermutated in mouse B cells. P-value of overlap was determined by Fisher's exact test.
(B) Statistical bootstrapping analysis used to determine the probability of identification of 40 random xTSS-RNA expressing genes solely based on expression level. Ten thousand control gene lists were obtained each containing 88 genes sharing similar expression profiles with the 88 genes contained in the translocation hotspots list. By overlapping random control groups and xTSS-RNA gene list, we simulate the distribution of number of genes contained in the xTSS-RNA gene list. The binomial fitting (red curve) illustrates the distribution of randomly overlapped genes between the control and xTSS-RNA gene sets. P-value was determined by binomial distribution.

Comparing xTSS-RNA expression between *Exosc3*$^{WT/WT}$ and *Exosc3*$^{COIN/COIN}$ B cells specifically for the 40 genes found to contain xTSS-RNA expression and undergo recurrent chromosomal translocations confirmed that these genes accumulate significantly higher levels of xTSS-RNA upon *Exosc3* ablation (Figure 29A). In contrast, mRNA levels at these 40 genes collectively were nearly identical between *Exosc3*$^{WT/WT}$

and *Exosc3*$^{COIN/COIN}$ cells (Figure 29B). This indicates that the observed increase in xTSS-RNA expression at recurrent translocation sites in *Exosc3*$^{COIN/COIN}$ cells is in fact due to loss of RNA exosome function and not simply owing to greater overall transcription at these sites. If RNA exosome activity plays a functional role in the generation of chromosomal translocations, one expectation would be that recurrent translocation sites should be contained amongst the highest xTSS-RNA expressing genes. Indeed recurrent translocation sites were observed to accumulate significantly more xTSS-RNA even in comparison to all other xTSS-RNA expressing genes in *Exosc3*$^{COIN/COIN}$ B cells (Figure 29C).

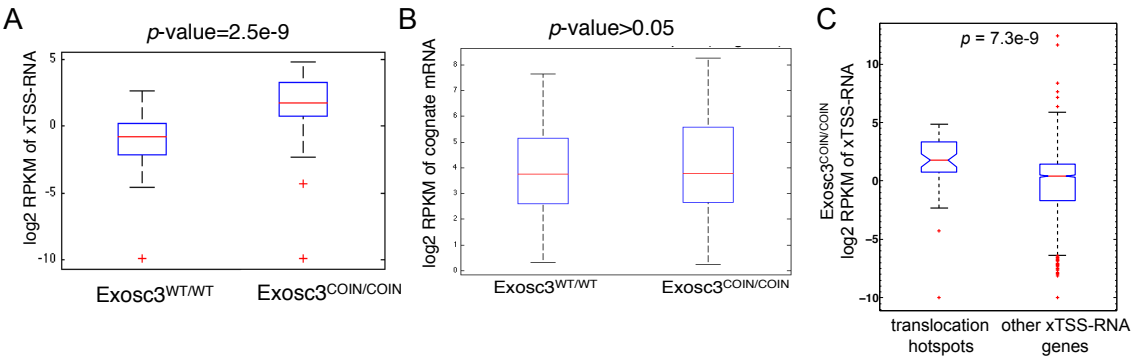



**Figure 29: Recurrent translocations sites display greater TSS proximal RNA exosome activity**
(A) Quantification of xTSS-RNA expression at 40 recurrent translocation sites overlapping with xTSS-RNA gene list from Figure 28A.
(B) Quantification of mRNA expression at 40 recurrent translocation sites overlapping with xTSS-RNA gene list from Figure 28A.
(C) Enrichment of xTSS-RNA expression amongst translocation hotspots in *Exosc3* deficient B cells. Translocation hotspot gene set is comprised of the 40 genes identified in Figure 28A that undergo recurrent AID mediated translocations and display xTSS-RNA accumulation in *Exosc3*$^{COIN/COIN}$ B cells. The "other xTSS-RNA genes" set is comprised of 1,694 genes expressing both xTSS-RNA and cognate mRNA, but have not been reported as recurrent translocation sites in B cells.
Boxplots represent median values compiled from two biological replicates. Whiskers represent 99% of data values. P-values were determined by Wilcoxon rank sum test.

In addition, overlapping the xTSS-RNA gene list with a limited set of 23 genes previously shown to undergo hypermutation in mouse germinal center B cells (Liu et al., 2008) revealed a similar enrichment to that observed with the translocation hotspot list. Specifically, 10 of the 23 reported hypermutation loci (43%) were present in the xTSS-

RNA gene list ($p < 0.0001$) (Figure 28A). Overlapping the xTSS-RNA, translocation hotspots, and hypermutation gene lists revealed a subset of genes consisting of *Myc*, *Pim1*, *Pax5*, *Cd79b*, and *Cd83* (Figure 28A). Interestingly, somatic mutations in each of these genes have been reported in human subjects with diffuse large B-cell lymphoma (Lohr et al., 2012; Pasqualucci et al., 2011). This analysis has revealed a significant enrichment of xTSS-RNA expression at genomic sites undergoing hypermutation and chromosomal translocations, two distinct types of AID induced DNA damage in B cells, suggesting that xTSS-RNA expression and/or RNA exosome function may be intimately involved in the targeting of AID activity.

Translocation breakpoints in AID expressing B cells (Klein et al., 2011b) were mapped onto the mouse genome according to a previously described bioinformatic workflow (Oliveira et al., 2012). This approach provided for a higher resolution analysis by allowing us to visualize sites of *Exosc3* sensitive transcription and translocations for individual genes. One of the most well studied B cell translocation sites owing to its strong oncogenic potential and causative role in Burkitt's lymphoma is *Myc*. As expected, *Myc* is included in the 88 genes reported to undergo recurrent AID dependent chromosomal translocations in B cells (Klein et al., 2011b). In addition, *Myc* was found to express significant levels of xTSS-RNA in *Exosc3* deficient B cells (Figure 27C,D) and as such is included amongst the 40 genes overlapping between the xTSS-RNA and translocation hotspot gene sets depicted in Figure 28A. However, the question remained as to where the chromosomal breakpoints were specifically occurring and whether they corresponded with sites of RNA exosome function. A closer examination of the *Myc* locus revealed two translocation hotspots in B cells. One was located upstream of the *Myc* TSS. The other translocation hotspot was located within the first intron of *Myc*. Strikingly, it was observed that both of these *Myc* translocation hotspots precisely overlapped with areas of *Exosc3* sensitive antisense transcription (Figure 30). The *Myc*

TSS upstream translocation site corresponds with the *Myc* xTSS-RNA transcript while the intron 1 translocation site maps to a genic antisense RNA exosome substrate transcript. This indicates that antisense transcription initiation is likely occurring both upstream of the canonical *Myc* coding transcript TSS and also downstream within the gene body. As these *Myc* antisense transcripts appear less than 1 kb in length and fail to accumulate in wild type B cells (Figure 30), it is likely that they are subjected to early transcription termination coupled to co-transcriptional RNA exosome mediated degradation.

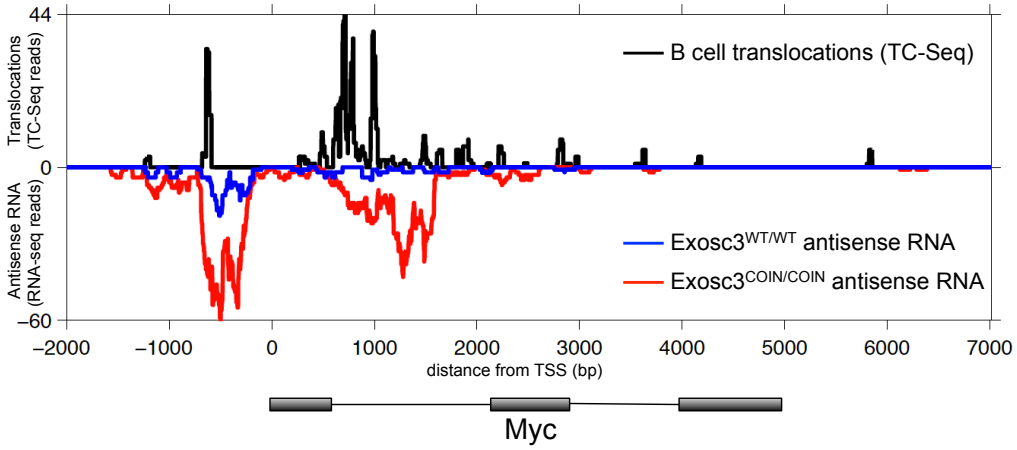



**Figure 30: *Myc* translocation breakpoints at sites of RNA exosome regulated xTSS-RNA and genic antisense transcription**
Mouse B cell translocation counts (Klein et al., 2011b) indicated on the positive y-axis. *Myc* antisense RNA-seq reads derived from *Exosc3*$^{WT/WT}$ or *Exosc3*$^{COIN/COIN}$ B cell transcriptomes indicated on the negative y-axis. The positions of *Myc* exons 1-3 relative to the TSS are represented as grey boxes. Indicated genotypes are on *ROSA26*$^{CreERt2/+}$ background and cells were treated with 4-OHT and stimulated with LPS and IL-4. Data compiled from two biological replicates.

The stabilization of genic antisense RNA exosome substrate transcripts at *Myc* led us to ask if antisense transcription may be preferentially targeted for degradation by RNA exosome. Using strand specific RNA-seq data sets we compared either sense or antisense transcript levels for individual annotated loci between *Exosc3*$^{WT/WT}$ and *Exosc3*$^{COIN/COIN}$. On a genome-wide scale, sense transcript levels are mildly increased between *Exosc3*$^{WT/WT}$ and *Exosc3*$^{COIN/COIN}$ B cells (Figure 31A). In contrast, we observed a much stronger shift towards greater antisense RNA levels at the vast majority of loci in

*Exosc3* deficient cells (Figure 31B). In light of this data, we examined the relationship between antisense transcript levels and chromosomal translocation frequency at translocation hotspot genes. The 88 translocation hotspots were computationally deconstructed into approximately 4 million genomic positions, which were then grouped into 13 tiers corresponding to increasing antisense expression. This approach provided a more accurate analysis of the relationship between antisense transcription and chromosome break frequency by allowing antisense rich and poor regions located within individual genes to be treated independently. Using this method we observed that the probability of detecting a translocation junction (TC-Seq read) was positively correlated with increasing antisense transcript levels (Figure 31C).

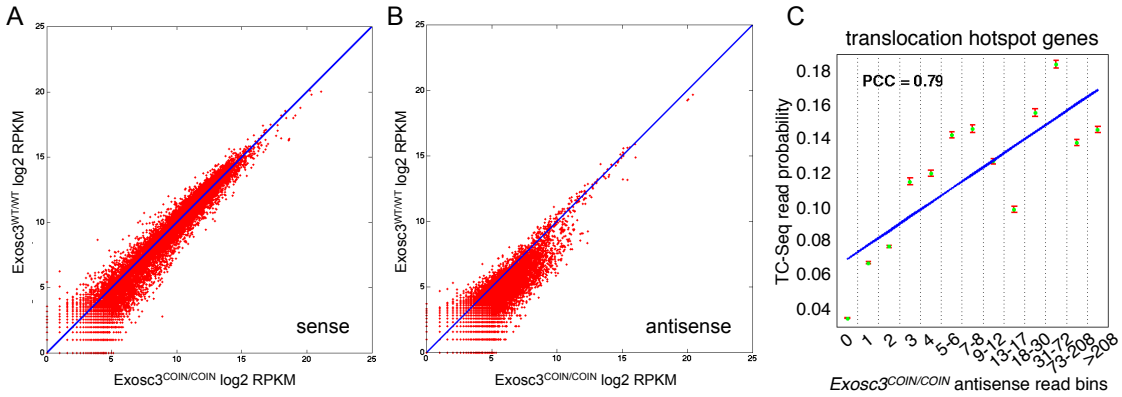



**Figure 31: Genome-wide RNA exosome targeting of antisense transcripts**
(A-B) Scatter plot of sense (A) and antisense (B) transcript expression at annotated loci in *Exosc3*$^{WT/WT}$ and *Exosc3*$^{COIN/COIN}$ B cell transcriptomes. Indicated genotypes are on *ROSA26*$^{CreERt2/+}$ background and cells were treated with 4-OHT and stimulated with LPS and IL-4. Data compiled from two biological replicates.
(C) Correlation between chromosomal breakpoint occurrence and antisense transcript expression levels in *Exosc3* deficient B cells at translocation hotspot genes. Approximately 4 x10$^6$ genomic positions were distributed across 13 bins according to antisense RNA expression levels. A negative binomial distribution was applied to each region to estimate the probability of observing a translocation junction. Error bars indicate the 95% confidence interval of parameter estimation. Blue line represents the robust fit of expected values. Pearson correlation coefficient (PCC) is indicated.

This observation of RNA exosome regulated antisense transcription within the gene body of *Myc* corresponding to a translocation hotspot, prompted the question of whether additional recurrent translocation sites exhibit RNA exosome surveillance. Specifically, we evaluated if any of the 48 translocation hotspots that do not display

significant accumulation of xTSS-RNA upon *Exosc3* ablation (Figure 28A) may in fact be regulated by RNA exosome at sites distal to their TSSs. Strikingly, many translocation hotspot genes devoid of TSS proximal RNA exosome regulation, as defined here by lack of xTSS-RNA accumulation, were indeed observed to contain short antisense transcripts stabilized in *Exosc3*$^{COIN/COIN}$ cells at various sites along the gene body. Included amongst these translocation hotspot genes with primarily distal antisense RNA exosome regulation were *Apobec3*, *Gimap5*, *Ly6a*, and *Ppp3cc* (Figure 32). Thus, cryptic antisense transcription initiation downstream of canonical TSSs may serve as an additional mechanism of RNA exosome mediated recruitment of AID onto transcriptionally active chromatin.

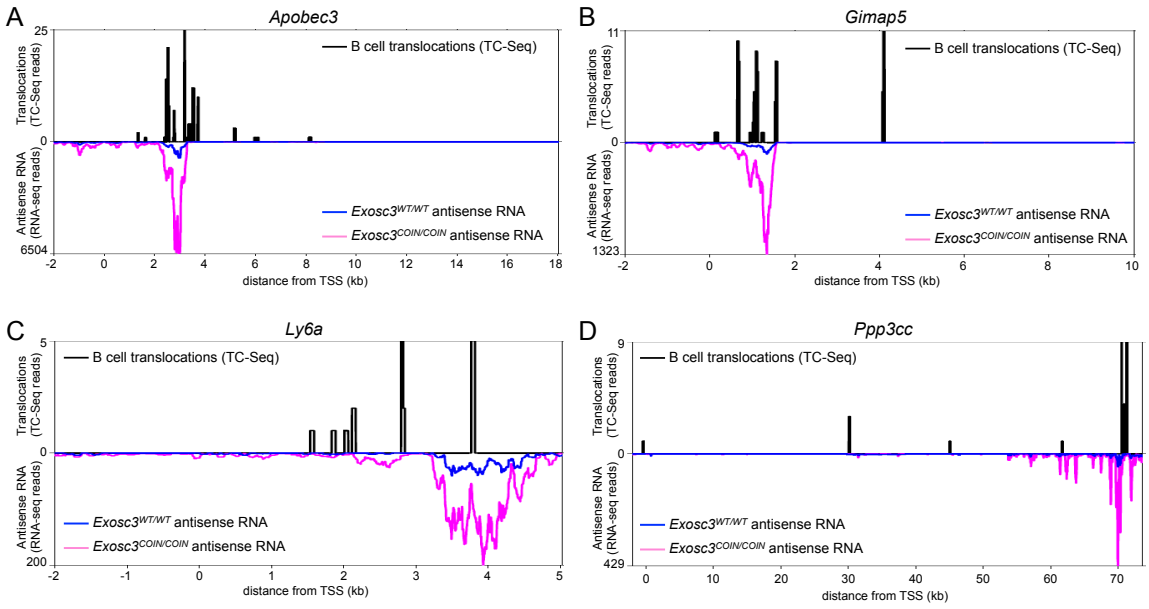



**Figure 32: Distal RNA exosome substrate antisense transcripts at sites of recurrent chromosomal translocations at xTSS-RNA deficient loci**
(A-D) *Apobec3* (A), *Gimap5* (B), *Ly6a* (C), and *Ppp3cc* (D) represent four previously characterized AID dependent translocation hotspot loci in mouse B cells (Klein et al., 2011b), which we observe as devoid of TSS proximal xTSS-RNA transcription but do exhibit distal RNA exosome targeted genic antisense transcription. Mouse B cell translocation counts (Klein et al., 2011b) indicated on the positive y-axis. *Exosc3*$^{WT/WT}$ or *Exosc3*$^{COIN/COIN}$ B cell antisense RNA-seq reads located on the negative y-axis. Indicated genotypes are on *ROSA26*$^{CreERt2/+}$ background and cells were treated with 4-OHT and stimulated with LPS and IL-4. Data compiled from two biological replicates.

In addition to sites undergoing recurrent chromosomal translocations, we looked to see whether the association between antisense transcription and breakpoints could

be extended genome-wide. A collection of 28,947 annotated genomic loci devoid of known overlapping antisense transcripts were evaluated in the *Exosc3*$^{COIN/COIN}$ transcriptome for translocation junctions. Similar to what was observed at translocation hotspots (Figure 31C), the occurrence of chromosomal breakpoints exhibited a positive linear relationship with respect to antisense transcription levels in B cells genome-wide (Figure 33). Furthermore, this analysis revealed a significantly stronger correlation between translocation occurrence and antisense transcription, in comparison to sense transcription (Figure 33). This result is consistent with the observation of greater susceptibility of antisense transcripts towards RNA exosome mediated degradation (Figure 31A,B) and supports a model in which genomic regions exhibiting greater RNA exosome activity are more prone to acquire DNA damage resulting from misappropriated AID activity.

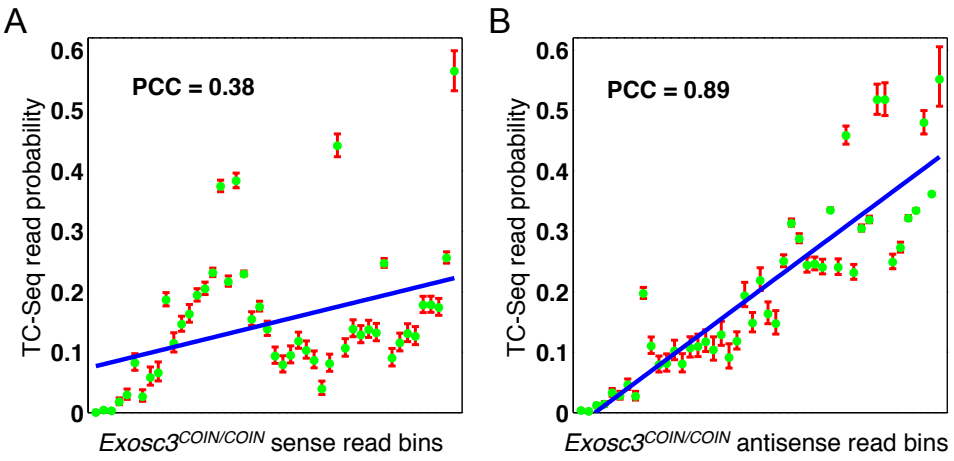



**Figure 33: Genome-wide association between antisense transcription and chromosomal translocation in B cells**
(A-B) Probability of translocation breakpoint occurrence in the B cell genome with respect to sense (A) or antisense (B) transcription levels. Genomic regions comprising 28,947 annotated loci were sorted across 46 expression tiers. A negative binomial distribution was applied to each region to estimate the probability of observing a translocation junction. Blue line represents the robust fit of expected values. Pearson correlation coefficient (PCC) is indicated.

### 3.9 RNA exosome facilitates targeting of AID to divergently transcribed promoters

As shown earlier, the *Exosc3* deficient B cell transcriptome data indicated that while many genes accumulate antisense xTSS-RNA in their promoter upstream region, many other genes exhibited little to no detectable accumulation of xTSS-RNA. This implies that collectively mammalian promoters may exhibit varying degrees of divergent transcription. To examine this further RNA exosome occupancy at divergent and non-divergent gene promoters was assessed using chromatin immunoprecipitation of RRP40 (Exosc3). AID targeted divergent promoters of *Pim1*, *Pax5*, *Myc*, *Cd79b*, and *Cd83* exhibited significant enrichment of RRP40 over background in stimulated wild type B cells (Figure 34A). Consistent with a nearly complete loss of RRP40 protein in $Exosc3^{COIN/COIN}$ cells (Figure 13B), little to no RRP40 occupancy could be observed at these divergent promoters in $Exosc3^{COIN/COIN}$ (Figure 34A). In addition, transcriptionally active non-divergent promoters comprised of *Cmas*, *Atp13a2*, *March2*, and *Fam107b* displayed little RRP40 recruitment in either $Exosc3^{WT/WT}$ or $Exosc3^{COIN/COIN}$ cells (Figure 34A). This data strongly suggests that the accumulation of xTSS-RNA at AID targeted divergently transcribed promoters is directly due to the loss of promoter proximal RNA exosome function in *Exosc3* deficient cells.

Since our data provides evidence of differential RNA exosome recruitment to mammalian promoter regions, we considered the possibility that this may also translate into differences in genomic AID targeting via its association with RNA exosome. Guided by this, as well as previous reports indicating that chromatin associated AID is preferentially localized over active TSSs (Yamane et al., 2011), we explored whether AID might be preferentially targeted to divergently transcribed promoters through an RNA exosome dependent mechanism. Chromatin immunoprecipitation of AID in stimulated $Exosc3^{WT/WT}$ or $Exosc3^{COIN/COIN}$ B cells indicated that AID targeting to

divergent *Pim1*, *Pax5*, *Myc*, *Cd79b*, and *Cd83* promoters was significantly impaired upon *Exosc3* ablation (Figure 34B). These differences in AID targeting efficiency cannot be explained simply through differences in AID expression as stimulated *Exosc3*$^{WT/WT}$ and *Exosc3*$^{COIN/COIN}$ B cells express comparable levels of AID (Figure 15A,B). As previous studies have indicated that AID activity at target genes can be quantitatively influenced by corresponding changes in transcription initiation rates (Bachl et al., 2001; Lee et al., 2001), we compared histone H3 lysine 4 trimethyl (H3K4me3) modification levels at AID targeted divergent promoter regions. H3K4me3 chromatin marks are predominantly localized to promoter regions actively undergoing transcription initiation (Barski et al., 2007; Guenther et al., 2007). Immunoprecipitation of H3K4me3 indicated that this epigenetic marker of transcription initiation was similar in abundance between *Exosc3*$^{WT/WT}$ and *Exosc3*$^{COIN/COIN}$ B cells at divergently transcribed promoters of *Pim1*, *Pax5*, *Myc*, *Cd79b*, and *Cd83* (Figure 34C). As such, impaired AID targeting to divergent promoter regions in *Exosc3* deficient B cells is likely not due to significant changes in transcription initiation. Furthermore, in contrast to divergently transcribed promoter regions, AID occupancy was nearly undetectable at non-divergent promoters of *Cmas*, *Atp13a2*, *March2*, and *Fam107b* in both wild type and *Exosc3* deficient B cells (not shown). Collectively, these data support a model in which RNA exosome is preferentially recruited to divergently transcribed promoter regions relative to non-divergent promoters, and consequently resulting in differential AID targeting efficiencies between such genomic regions.

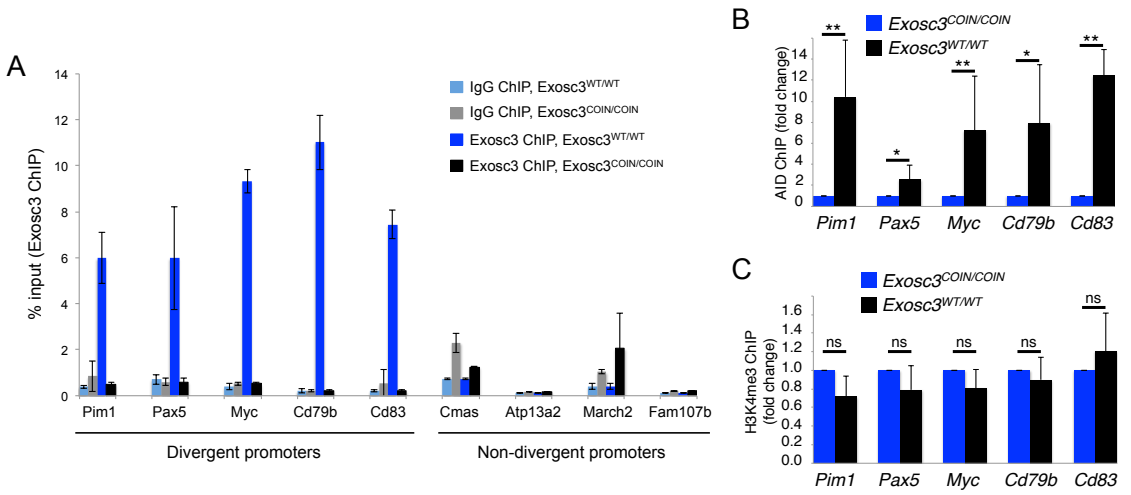



**Figure 34: RNA exosome facilitates recruitment of AID to divergently transcribed promoters**
(A) RRP40 (Exosc3) occupancy at divergent and non-divergent promoters. B cells of indicated genotypes are on *ROSA26*$^{CreERt2/+}$ background, treated with 4-OHT and stimulated with LPS + IL-4. Chromatin immunoprecipitation (ChIP) was performed using anti-RRP40 (Genway) or control rabbit IgG. Quantitative PCR was performed using primers specific for sequences upstream and proximal to the TSS of the indicated genes. Data are plotted as mean values of RRP40 enrichment relative to input from three technical replicates. Error bars represent standard deviation.
(B) *Exosc3* dependent AID targeting to divergent promoters. ChIP was performed using anti-AID or control rabbit IgG. B cell culture and quantitative PCR as described in (A). Data are plotted as fold change of ChIP product normalized to *Exosc3*$^{COIN/COIN}$. $*P < 0.05$, $**P < 0.01$ (log2 transformation of *Z*-test).
(C) H3K4me3 histone marks are maintained in *Exosc3* deficient B cells at sites normally targeted by AID. ChIP was performed with anti-H3K4me3 (Millipore) or control rabbit IgG. B cell culture and quantitative PCR as described in (A). Data are plotted as fold change of ChIP product normalized to *Exosc3*$^{COIN/COIN}$. ns, not significant (*t*-test).

To evaluate whether divergent xTSS-RNA expressing regions contain sequence determinants that facilitate recruitment of AID, we utilized CRISPR/Cas9 gene editing to delete these corresponding regions for AID target genes *Cd83* and *Pim1* (Figure 35A,B). Homozygous deleted *Cd83* and *Pim1* xTSS-RNA clonal CH12F3 B cell lymphoma lines were created (Figure 35C). CH12F3 cells exhibit stimulation dependent AID induction and cytokine directed CSR, illustrating that key AID regulatory mechanisms are recapitulated in these cells. Quantitative RT-PCR revealed that *Pim1* and *Cd83* mRNA levels were slightly reduced by approximately 5% and 30% in *Pim1* and *Cd83* xTSS-RNA region deleted CH12F3 cells, respectively (not shown). However, AID occupancy at *Pim1* and *Cd83* were disproportionately reduced in stimulated CH12F3 cells containing deletion of the corresponding xTSS-RNA region. AID chromatin immunoprecipitation at

*Pim1* and *Cd83* promoter proximal regions was reduced by approximately 85% and 75% in *Pim1* and *Cd83* xTSS-RNA region deleted CH12F3 cells, respectively (Figure 35D,E). SHM was evaluated within the first kilobase downstream of the *Cd83* mRNA TSS in parental and *Cd83* xTSS-RNA region deleted CH12F3 cells. Incidentally, *Cd83* was selected over *Pim1* due to a previous report indicating that AID induced mutations at *Cd83*, unlike *Pim1*, fail to undergo high fidelity DNA repair in germinal center B cells (Liu et al., 2008). As such, we hypothesized that AID mediated low frequency "off-target" mutations introduced at the *Cd83* locus in parental and xTSS-RNA region deleted CH12F3 cells might persist and allow for their quantification. Mirroring the observed reduction in AID occupancy at *Cd83* TSS proximal regions, SHM frequency at *Cd83* was reduced by approximately 75% in stimulated *Cd83* xTSS-RNA region deleted CH12F3 cells relative to parental CH12F3 cells (Figure 35F). These data support the notion that xTSS-RNA expressing regions of divergently transcribed genes contain sequences or motifs that promote AID recruitment. It is possible that the xTSS-RNA region deletions characterized here may have disrupted transcription of the corresponding xTSS-RNA and in essence converted each locus to unidirectional sense transcription. Alternatively, deletion of the xTSS-RNA region may not have disrupted divergent transcription per se, but removed transcribed sequence motifs that promote AID recruitment. This notion will be discussed later in further detail. In either case, xTSS-RNA regions contain sequence determinants that facilitate targeting of AID.

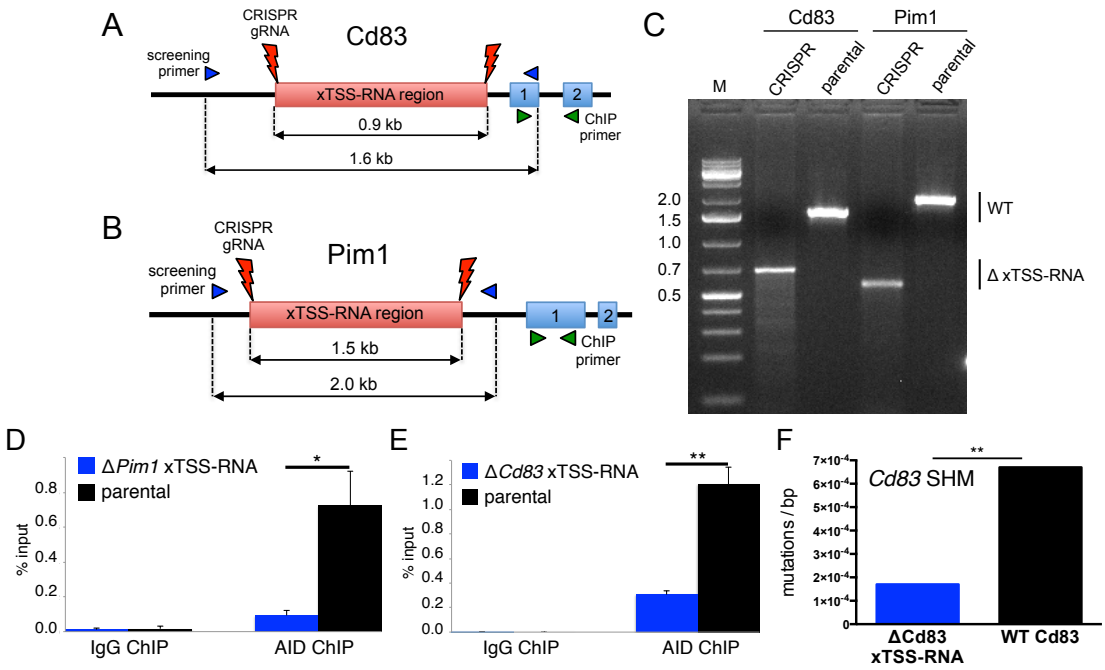



**Figure 35: Deletion of xTSS-RNA expressing region reduces AID recruitment and hypermutation**
(A-B) CRISPR/Cas9-mediated deletion strategy of *Cd83* (A) and *Pim1* (B) xTSS-RNA expressing regions in CH12F3 B cells. Locations of CRISPR/Cas9 guide RNAs (red markings), genotyping primers (blue triangles), ChIP primers (green triangles), and numbered exons (blue boxes) are indicated.
(C) PCR genotyping of Cd83 and Pim1 xTSS-RNA region deleted CH12F3 clones. Predicted wild type and xTSS-RNA deletion amplicons illustrated in (A) and (B). Molecular weight markers represent kilobase units.
(D-E) Deletion of xTSS-RNA region impairs AID targeting to divergently transcribed genes. Anti-AID ChIP was performed on parental or clonal CH12F3 B-cell lymphoma lines containing CRISPR/Cas9 mediated deletions of *Pim1* (D) or *Cd83* (E) xTSS-RNA regions. qPCR was performed using exon 1 specific primers. Mean values from three technical replicates are indicated. *$P < 0.05$, **$P < 0.01$ (*t*-test).
(F) Deletion of *Cd83* xTSS-RNA region impairs SHM. Mutation frequency was determined within a 488 base pair region beginning approximately 150 base pairs downstream of the *Cd83* mRNA TSS. All mutations were derived from unique clonal amplified sequences. Number of sequenced clones for parental and *Cd83* xTSS-RNA region deleted CH12F3 was 69 and 102, respectively. **$P < 0.01$ (proportion test).

## 3.10 Evidence of RNA exosome substrate RNA:DNA hybrids at divergently transcribed AID target sites

One predominant mechanism whereby AID gains access to substrate ssDNA is through the displacement of non-template strand DNA during co-transcriptional RNA:DNA hybrid formation. This mechanism is particularly prevalent within the highly repetitive G-rich switch regions of the *IgH* locus during CSR (Yu et al., 2003). In addition, studies in yeast have shown that the Sen1 helicase component of the RNA exosome cofactor Nrd1-Nab3-Sen1 transcription termination complex plays an important role in resolving co-transcriptional RNA:DNA hybrids (Mischo et al., 2011). In light of these

studies, and our data supporting an RNA exosome dependent effect on AID targeting to divergently transcribed promoter regions, we hypothesized that such regions may be more prone to RNA:DNA hybrid formation.

Recently, phosphorylation of histone H3 at serine 10 (H3S10ph) has been associated with R-loop forming regions of chromatin (Castellano-Pozo et al., 2013). Consistent with this observation, other studies have reported that R-loop generating donor and acceptor *IgH* switch regions are specifically marked with H3S10ph during CSR (Li et al., 2013). We therefore examined H3S10ph occupancy as a surrogate for R-loop formation at RNA exosome and AID dually targeted divergent promoters. The H3S10ph chromatin mark was indeed significantly elevated at *Myc*, *Pax5*, *Cd79b*, and *Cd83* promoter regions in wild type stimulated B cells (Figure 36A,B). In contrast, H3S10ph levels in these cells were nearly undetectable at non-divergent promoters of *Cmas*, *Atp13a2*, *March2*, and *Fam107b* (Figure 36A). Interestingly, a significant increase in H3S10ph accumulation was observed at multiple divergently transcribed AID target genes upon *Exosc3* ablation (Figure 36C). These results suggest that actively transcribed divergent promoters may constitute genomic regions that are more prone to R-loop formation. In turn, frequent R-loops at divergent promoters would likely result in greater transcriptional stalling leading to RNA exosome and AID recruitment.

R-loop formation at promoter proximal regions was assessed using a second independent method known as DNA:RNA immunoprecipitation (DRIP). This technique involves a modified nucleic acid immunoprecipitation protocol utilizing the sequence independent RNA:DNA hybrid specific S9.6 monoclonal antibody to enrich for genomic R-loops (Ginno et al., 2012). In this assay we observed a generally greater DRIP efficiency in *Exosc3* deficient B cells at AID targeted divergent promoters compared with non-divergent promoters (Figure 36D). In particular, *Pim1* and *Cd79b* experienced a 5-

fold increase in promoter proximal R-loop formation in response to *Exosc3* ablation (Figure 36D).

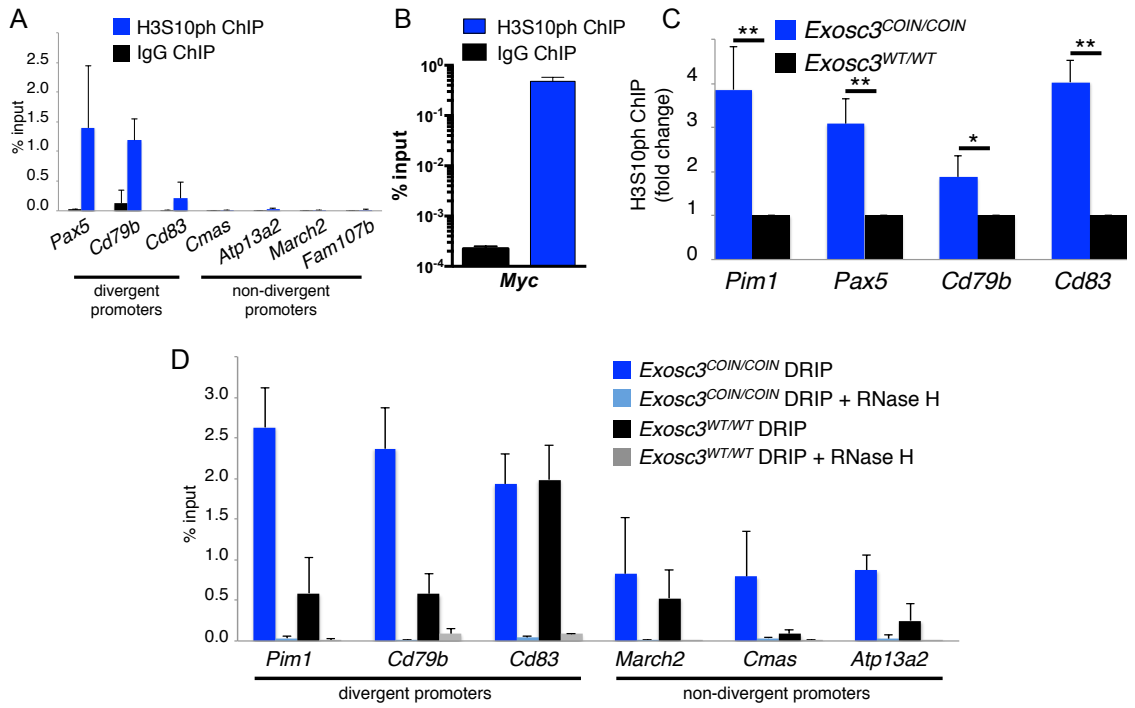



**Figure 36: RNA:DNA hybrids at AID targeted RNA exosome substrate divergent promoters**
(A-B) H3S10ph occupancy at divergent and non-divergent promoters in LPS + IL-4 stimulated wild type B cells. ChIP was performed using anti-H3S10ph (Millipore) or control rabbit IgG. Quantitative PCR was performed using primers specific for sequences upstream and proximal to the TSS of the indicated genes. Data are plotted as mean values of H3S10ph enrichment relative to input from three technical replicates.
(C) H3S10ph accumulation at AID targeted divergent promoters upon *Exosc3* ablation. B cells of indicated genotypes are on *ROSA26*$^{CreERt2/+}$ background, treated with 4-OHT and stimulated with LPS + IL-4. ChIP was performed as described in (A). Data are plotted as mean fold change of *Exosc3*$^{COIN/COIN}$ normalized to *Exosc3*$^{WT/WT}$ from three technical replicates. *$P < 0.05$, **$P < 0.01$ (log2 transformation of Z-test).
(D) DNA:RNA immunoprecipitation (DRIP) of divergent and non-divergent promoter regions using S9.6 monoclonal antibody. B cell culture as described in (C). Quantitative PCR as described in (A). Data are plotted as mean values of DRIP enrichment relative to input from three technical replicates.

### 3.11 Class switch recombination in *Exosc10* deficient B cells

*Exosc10* conditional null mice were generated from *Exosc10*$^{COINneo/+}$ targeted mouse embryonic stem cells using the same approach as described for *Exosc3*. Briefly, *Exosc10*$^{COINneo/+}$ male chimeric mice were derived via blastocyst injection of targeted ES cells. Male chimeric mice displaying nearly complete ES cell contribution on the basis of coat color were crossed with female *ACTB:FLPe* transgenic mice. Resulting pups were genotyped for loss of the neomycin selection cassette and the *ACTB:FLPe* transgene

was outbred during subsequent backcrossings. *Exosc10*$^{COIN/+}$ heterozygotes of the conditional allele were typically crossed with *Exosc10*$^{LacZ/+}$ heterozygous null mice to generate *Exosc10*$^{COIN/LacZ}$ conditional null mice and associated wild type and heterozygotes. In such crosses one parent was usually homozygous for the inducible *ROSA26*$^{CreERt2}$ allele.

Naive splenic B cells were isolated from *Exosc10*$^{COIN/LacZ}$ *ROSA26*$^{CreERt2/+}$ and control mice as described for *Exosc3* conditional mice. B cells were treated with 4-OHT and stimulated with IgG1 CSR conditions using LPS + IL-4. As described earlier, the *Exosc10*$^{COIN}$ allele was engineered to produce an inversion of *Exosc10* exon 2 in the presence of Cre recombinase activity, resulting in the concomitant activation of the red fluorescent FP635 reporter and transcription termination (Figure 37A). Indeed activation of CreERt2 with 4-OHT produced a robust induction of FP635 fluorescence in greater than 90% of *Exosc10*$^{COIN/LacZ}$ *ROSA26*$^{CreERt2/+}$ B cells (Figure 37B). *Exosc10* mRNA levels were measured using a Taqman quantitative PCR gene expression probe specific for the *Exosc10* exons 11-12 junction. As anticipated, *Exosc10* mRNA levels in 4-OHT treated *Exosc10*$^{COIN/LacZ}$ *ROSA26*$^{CreERt2/+}$ B cells were dramatically reduced to less than 5% of control *Exosc10*$^{WT/WT}$ *ROSA26*$^{CreERt2/+}$ B cells (Figure 37C). In addition, Rrp6 (Exosc10) protein abundance was reduced to negligible levels in 4-OHT treated *Exosc10*$^{COIN/LacZ}$ *ROSA26*$^{CreERt2/+}$ B cells (Figure 37D). Collectively, these data help validate the constructed *Exosc10*$^{COIN}$ allele as a functionally conditional allele of *Exosc10* and indicate that 4-OHT treatment of *Exosc10*$^{COIN/LacZ}$ *ROSA26*$^{CreERt2/+}$ B cells constitutes a viable model of *Exosc10* deficiency.

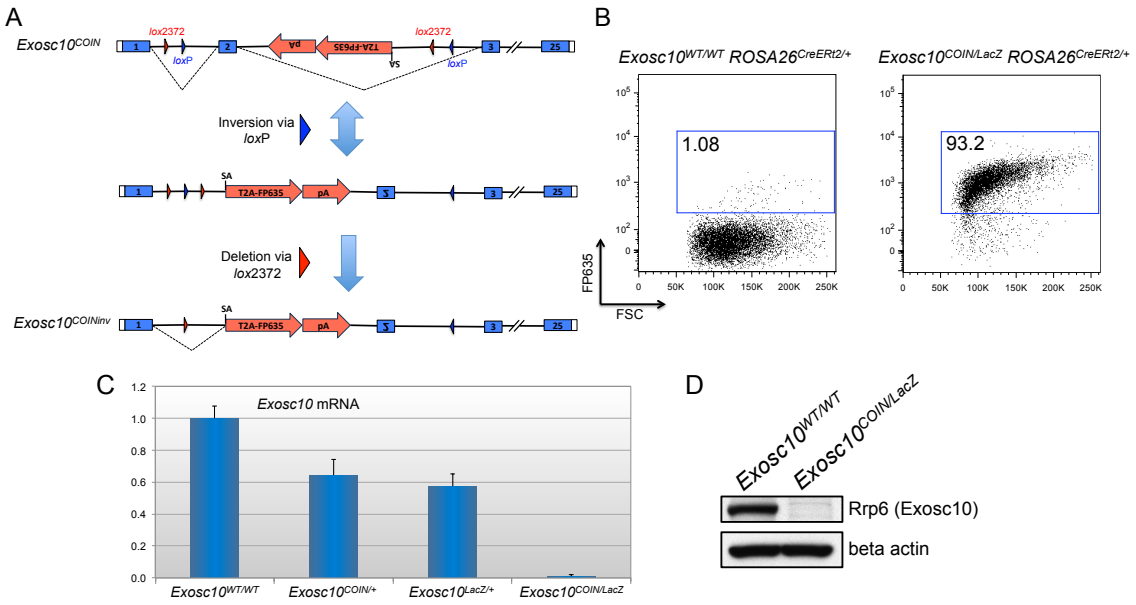



**Figure 37: *Exosc10*$^{COIN}$ allele inversion leads to functional loss of *Exosc10* expression**
(A) Cre mediated inversion scheme of *Exosc10*$^{COIN}$. Schematic illustrates inversion of *Exosc10* exon 2 and activation of the FP635 expressing terminal exon through inversion of the *lox*P pair. A subsequent Cre mediated deletion between the *lox*2372 pair is illustrated. Numbered boxes represent *Exosc10* exons. Red and blue triangles indicate positions of *lox*2372 and *lox*P sites, respectively. Dashed lines indicate *Exosc10* splicing. SA, splice acceptor.
(B) Induction of FP635 fluorescence in *Exosc10* deficient B cells. Cells were cultured in the presence of 4-OHT and stimulated with LPS and IL-4. Number within each plot indicates the percentages of FP635 expressing B cells.
(C) Quantitative RT-PCR analysis of *Exosc10* mRNA levels in ex vivo B cell cultures. Indicated genotypes are on *ROSA26*$^{CreERt2/+}$ background. Cells were treated as described in (B). Expression levels are normalized against beta actin (*Actb*) and plotted relative to *Exosc10*$^{WT/WT}$.
(D) Western blot analysis of Rrp6 (*Exosc10*) protein levels in cultured B cells. Indicated genotypes are on *ROSA26*$^{CreERt2/+}$ background. Cells were treated as described in (B). Anti-actin immunoblot serves as a loading control.

*Exosc10* deficient B cells were evaluated for IgG1 CSR by comparing surface IgG1 expression following LPS and IL-4 stimulation. CSR efficiency was significantly reduced in *Exosc10*$^{COIN/LacZ}$ B cells to less than 50% frequency of wild type control B cells (Figure 38A). As CSR defects can be compounded by improper AID induction and cellular proliferation, these variables were evaluated in our *Exosc10* deficient B cell cultures. Indeed, AID protein levels were comparable between *Exosc10*$^{COIN/LacZ}$ and *Exosc10*$^{WT/WT}$ B cells (Figure 38B). Furthermore, cellular proliferation as determined by dilution of the VPD450 violet fluorescent tracer in *Exosc10*$^{COIN/LacZ}$ B cells was similar to wild type control B cells (Figure 38C). These data indicate that loss of the Rrp6

(*Exosc10*) nuclease subunit of the RNA exosome results in a substantial CSR defect. Interestingly, *Exosc10* ablation consistently produced a milder CSR defect compared to *Exosc3* ablation. Unlike *Exosc3*, which expresses a necessary subunit of the RNA exosome core and is thought to be present in all assembled RNA exosome complexes, the *Exosc10*-encoded Rrp6 subunit is not absolutely required for RNA exosome function (Wasmuth and Lima, 2012). In fact, *Dis3*-encoded Rrp44 nuclease containing RNA exosome complexes can exclude Rrp6 and retain catalytic activity. Therefore, it is possible that the difference in CSR impairment observed between *Exosc3* and *Exosc10* deficient B cells might indicate a potential role for *Dis3* in CSR.

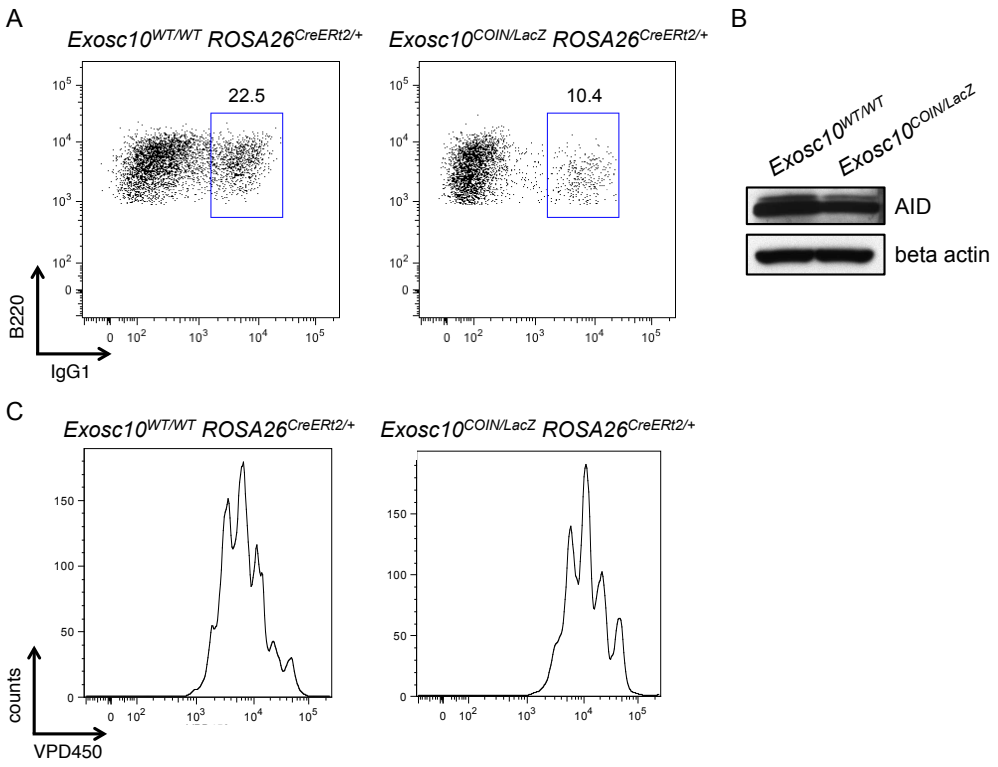



**Figure 38: Impaired CSR in *Exosc10* deficient B cells**
(A) IgG1 CSR frequency in 4-OHT treated, LPS + IL-4 stimulated *Exosc10*$^{WT/WT}$ *ROSA26*$^{CreERt2/+}$ and *Exosc10*$^{COIN/LacZ}$ *ROSA26*$^{CreERt2/+}$ B cell cultures. Numbers within FACS plots indicate the percentage of IgG1 isotype switched B220$^{+}$ B cells. *Exosc10*$^{COIN/LacZ}$ *ROSA26*$^{CreERt2/+}$ plot is gated on FP635$^{+}$ cells.
(B) Western blot analysis of AID protein levels in cultured B cells. Indicated genotypes are on *ROSA26*$^{CreERt2/+}$ background. Cells were treated as described in (A). Anti-actin immunoblot serves as a loading control (reprinted from Figure 37D as anti-actin, -AID, and -Rrp6 immunoblots were performed in parallel using the same lysate samples).
(C) B cell proliferation analysis using VPD450 fluorescent tracer dilution. Plots indicate the degree of VPD450 dilution resulting from cell division following 3 days of 4-OHT treatment. Cells were stimulated as described in (A).

### 3.12 ncRNA discovery in RNA exosome deficient embryonic stem cells

One of the more intriguing aspects of RNA exosome substrate transcripts is their remarkable instability in normal cells. For instance, the upregulation of xTSS-RNA transcripts in *Exosc3* deficient cells often initiates from a nearly undetectable baseline level. We therefore hypothesized that RNA exosome deficiency might serve as an experimental model for ncRNA discovery by virtue of the stabilization of transcripts that would otherwise be rapidly turned over by RNA exosome. For practical purposes such as access to greater cell numbers and avoiding costly and lengthy mouse breeding, *Exosc3* and *Exosc10* conditional null embryonic stem (ES) cells were rederived from isolated blastocysts. $Exosc3^{COIN/COIN}$ and $Exosc10^{COIN/LacZ}$ ES cells containing the $ROSA26^{CreERt2}$ allele were treated with 4-OHT. $GFP^{+}$ *Exosc3* and $FP635^{+}$ *Exosc10* deficient ES cells were purified using flow cytometry. Western blot and qRT-PCR analyses of these samples confirmed a nearly complete loss of *Exosc3* and *Exosc10* derived protein and mRNA (not shown). RNA prepared from these cells was used for strand specific RNA sequencing. As previously performed for *Exosc3* deficient B cells, ES cell RNA-seq reads mapping 2 kilobases upstream or downstream of annotated TSSs were plotted with respect to strand polarity. Similar to what was observed in B cells, $Exosc3^{COIN/COIN}$ ES cells displayed a highly significant accumulation of TSS upstream antisense transcripts ($p = 9.3E-151$) (Figure 39A). These transcripts represent the ES cell counterparts of B cell xTSS-RNA. Interestingly, $Exosc10^{COIN/LacZ}$ ES cells also displayed a statistically significant upregulation of TSS upstream antisense transcripts ($p = 5.6E-21$) (Figure 39B), although their accumulation was far less compared to *Exosc3* deficient ES cells. Again, it is possible that the difference in xTSS-RNA stabilization between *Exosc3* and *Exosc10* deficient ES cells may be due to partial functional redundancy between the two RNA exosome nuclease subunits *Exosc10* and *Dis3*. Although expected, the presence of RNA exosome sensitive xTSS-RNA transcripts in

ES cells confirms that RNA exosome mediated surveillance of promoter proximal transcripts is not restricted to B cells, but more likely represents a general feature of cellular transcription.

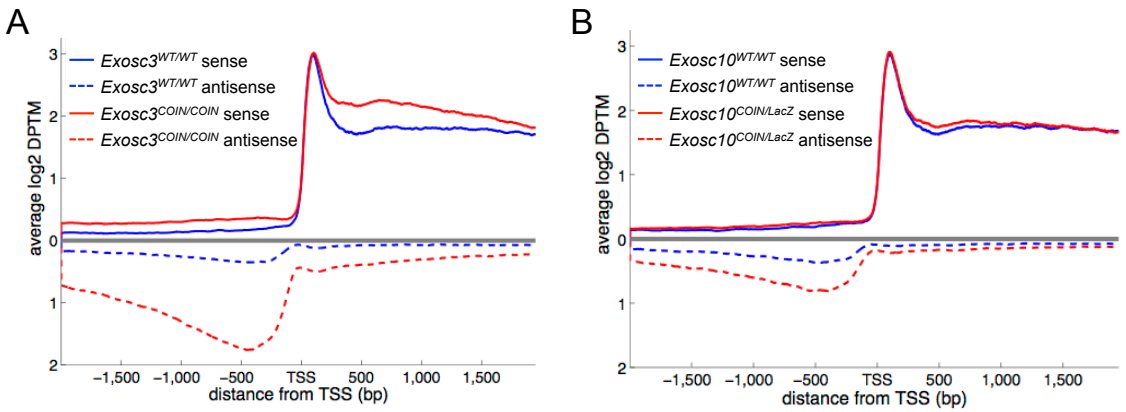



**Figure 39: Accumulation of xTSS-RNA in *Exosc3* and *Exosc10* deficient embryonic stem cells**
(A-B) Distribution of TSS proximal strand specific RNA sequencing reads in *Exosc3* and *Exosc10* deficient ES cells. Indicated genotypes are on *ROSA26*$^{CreERt2/+}$ background and cells were treated with 4-OHT as described. Data was compiled from two replicate experiments.

Transcriptomes of *Exosc3*$^{COIN/COIN}$ and *Exosc10*$^{COIN/LacZ}$ ES cells were computationally reconstructed from RNA-seq data and interrogated for perturbations in ncRNA expression. Known annotated lincRNAs (Guttman et al., 2009) in addition to novel uncharacterized lincRNAs revealed specifically in either *Exosc3* or *Exosc10* deficient transcriptomes were evaluated. In defining novel lincRNA transcripts we incorporated conventional lincRNA characteristics including non-coding potential, multi-exonic splicing pattern, and a minimum transcript length of 200 nucleotides. This analysis yielded a collection of 4,975 ES cell expressed lincRNAs. When compared to wild type ES cell expression, 38% and 22% of these lincRNAs were significantly upregulated in *Exosc3* and *Exosc10* deficient ES cells, respectively (Figure 40). For many of these lincRNAs, expression was often undetectable in wild type cells, a feature reminiscent of RNA exosome substrate xTSS-RNAs. In addition, these *Exosc3* and *Exosc10* dependent changes in lincRNA expression were reproducible (Figure 40A),

suggesting that they represent specific transcriptional events and/or entities and not stochastic transcripts arising from transcriptional noise. Similar to xTSS-RNA, global changes in magnitude of lincRNA expression were generally more pronounced in *Exosc3* deficient cells compared to *Exosc10* deficient cells (Figure 40A). Comparing lincRNA expression changes between *Exosc3*$^{COIN/COIN}$ and *Exosc10*$^{COIN/LacZ}$ ES cells revealed a partially co-linear relationship (0.59 Pearson correlation) (Figure 40B), suggesting that many, but not all, differentially expressed lincRNAs are regulated by RNA exosome complexes containing both Rrp40 and Rrp6. Specifically, 83% of differentially expressed lincRNAs in *Exosc10* deficient ES cells, were also upregulated in *Exosc3* deficient cells (Figure 40C). In contrast, only 48% of *Exosc3* regulated lincRNAs were upregulated in *Exosc10* deficient cells (Figure 40C). Collectively these data support a role for RNA exosome in modulating lincRNA stability. Intriguingly, many *Exosc3* regulated lincRNAs did not present significant perturbations upon *Exosc10* ablation. Several non-mutually exclusive possibilities may explain this data. It is possible that *Exosc3* and *Exosc10* co-regulated lincRNAs may largely be regulated by Rrp6-containing nuclear RNA exosome complexes, whereas *Exosc3*-specific regulated lincRNAs may comprise cytoplasmic RNA exosome substrates. Alternatively, *Exosc3*-specific regulated lincRNAs may be substrates of nuclear and/or cytoplasmic RNA exosome complexes containing the Rrp44 (*Dis3*) nuclease and excluding Rrp6. It is also possible that a role for Rrp6 in regulating *Exosc3*-specific lincRNAs in Rrp6/Rrp44-containing RNA exosome complexes may be hidden due to functional redundancy in resulting Rrrp44-containing RNA exosome complexes.

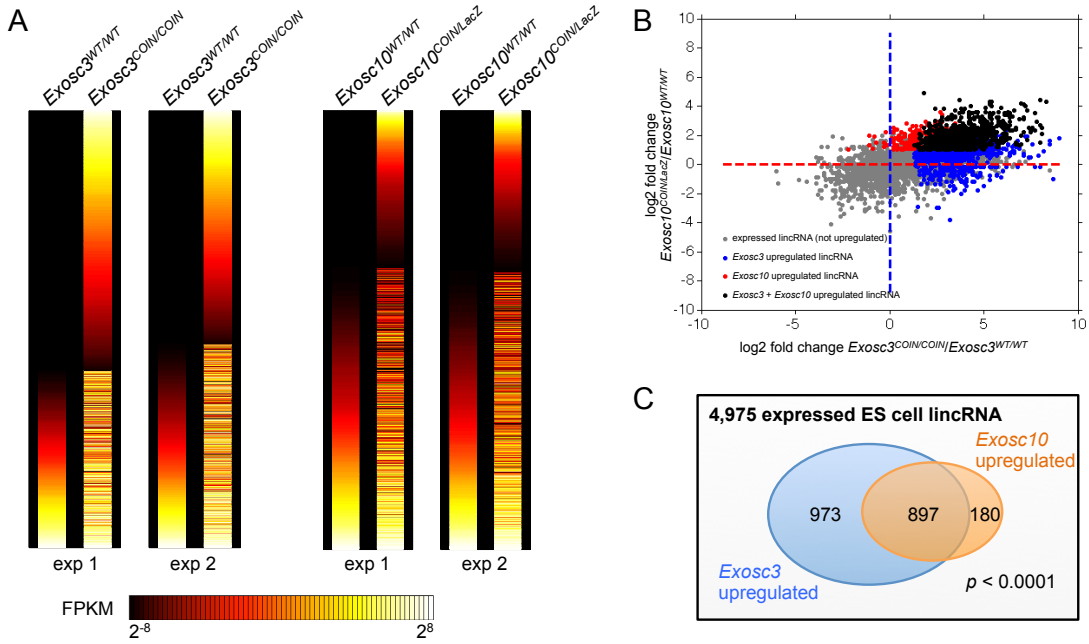



**Figure 40: RNA exosome ablation reveals novel lincRNA in ES cells**
(A) Heat map illustrating lincRNA expression levels in *Exosc3* and *Exosc10* deficient ES cells. Horizontal lines represent individual lincRNA transcripts. Indicated genotypes are on *ROSA26*$^{CreERt2/+}$ background and cells were treated with 4-OHT as described. Replicate experiments are shown.
(B) Scatter plot illustrating *Exosc3* and *Exosc10* dependencies for individual lincRNAs in ES cells. Log2 fold change over wild type control in *Exosc3* and *Exosc10* deficient ES cells plotted on x- and y-axes, respectively.
(C) Venn diagram illustrating the numerical overlap of *Exosc3* and *Exosc10* regulated lincRNAs in ES cells. P-value determined by Fisher's exact test.

Enhancers constitute critical cis-acting chromatin domains that play key roles in transcriptional activation and frequently serve as key determinants of tissue-specific gene expression programs. Often located distal to the promoters they regulate, enhancers are thought to function in close spatial proximity to promoters through long range chromatin looping. Although initially thought to serve only as transcription factor docking sites or scaffolds, it is commonly appreciated now that enhancer function is far more dynamic. For instance, functional enhancer elements are actively transcribed by RNAP II and produce labile enhancer RNAs (eRNAs) (De Santa et al., 2010; Kim et al., 2010). These eRNA transcripts have been proposed as integral components imparting functionality upon enhancers. It one model, eRNAs promote enhancer-promoter interactions through eRNA mediated recruitment of chromatin looping cohesin

complexes and promoter occupied Mediator complex (Lam et al., 2014; Plank and Dean, 2014). Other models have proposed enhancer transcription leads to chromatin remodeling events that promote transcription initiation and/or elongation of target genes (Mousavi et al., 2013).

As eRNAs typically possess short half lives in normal cells, we hypothesized whether their instability was partly due to RNA exosome mediated degradation. In fact, *Exosc3* deficient B cells displayed a 5-fold increase in transcripts mapping to the distal *IgH* 3' regulatory region enhancer, as well as upstream enhancer elements mapping to the *Cd83* locus (not shown). A more complete analysis of enhancer elements in RNA exosome ablated cells was greatly aided by better annotations of these elements in ES cells. Predicted enhancers have been systematically mapped across the genome of mouse ES cells on the basis of H3K4me1, H3K27ac, p300 histone acetyltransferase occupancies, and H3K4me3 exclusion (Shen et al., 2012). eRNA expression levels at these enhancers were compared between wild type and *Exosc3* or *Exosc10* deficient ES cells. We observed significantly higher levels of eRNA expression at 25% of active annotated enhancers in *Exosc3* ablated ES cells relative to wild type control ES cells (Figure 41). In addition, 8% of enhancer elements in ES cells displayed increased eRNA expression upon loss of *Exosc10* (Figure 41). Similar to RNA exosome regulated lincRNAs, many but not all RNA exosome regulated enhancer elements were shared between *Exosc3* and *Exosc10* deficient ES cells (Figure 41B,C). Specifically, eRNA levels at 27% of *Exosc3* dependent enhancers displayed *Exosc10* co-regulation (Figure 41C). In contrast, 90% of *Exosc10* regulated enhancers also displayed increased eRNA levels in *Exosc3* deficient ES cells. Therefore, a significant fraction of enhancer elements in ES cells are transcriptionally regulated by RNA exosome, with Rrp6 function partly restricted within the context of Rrp40-containing RNA exosome core complexes.

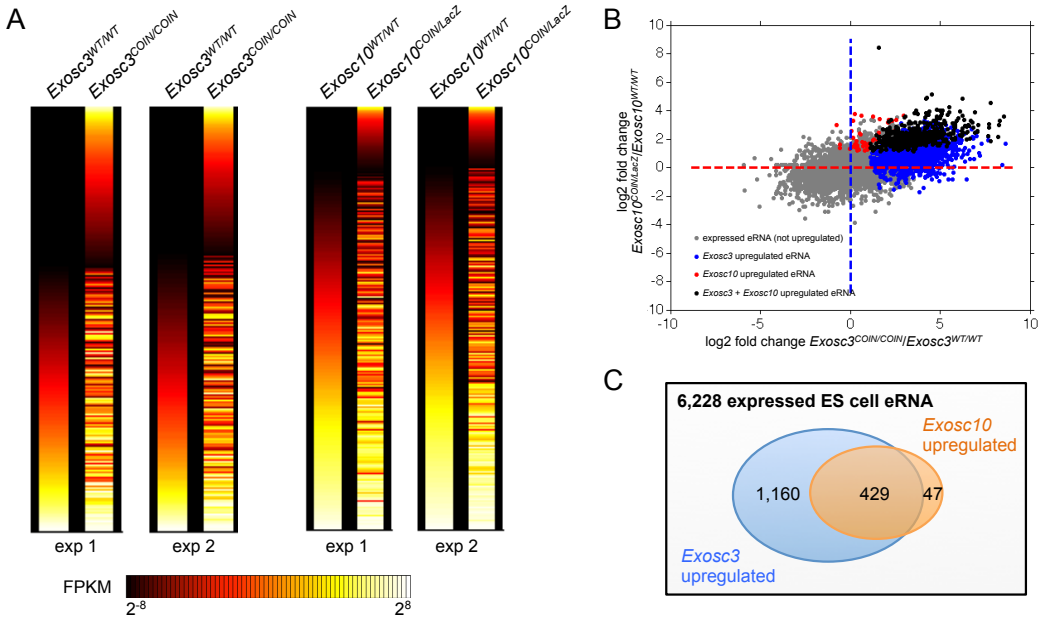



**Figure 41: RNA exosome mediated regulation of transcription at enhancer elements in ES cells**
(A) Heat map illustrating eRNA expression levels in *Exosc3* and *Exosc10* deficient ES cells. Horizontal lines represent individual enhancer elements. Indicated genotypes are on *ROSA26*$^{CreERt2/+}$ background and cells were treated with 4-OHT as described. Replicate experiments are shown.
(B) Scatter plot illustrating *Exosc3* and *Exosc10* co-regulation of eRNA transcription at enhancer elements in ES cells. Log2 fold change over wild type control in *Exosc3* and *Exosc10* deficient ES cells plotted on x- and y-axes, respectively.
(C) Venn diagram illustrating the numerical overlap of *Exosc3* and *Exosc10* regulated enhancers in ES cells.

# Chapter 4: Discussion

## 4.1 Transcription stalling as a mechanism of RNA exosome mediated AID targeting

Transcription is a discontinuous process involving frequent pausing and re-initiation of RNAP II as it traverses through the DNA (Churchman and Weissman, 2011). Transcription pausing can be a regulated event such as in the case of promoter proximal pausing, whereby RNAP II having escaped pre-initiation complex formation, frequently synthesizes a ~50 nucleotide long nascent transcript before it undergoes a stable and often prolonged pause on the chromatin (Kwak and Lis, 2013). Pausing of RNAP II can also be the consequence of encountering obstacles in the DNA template. Such impediments can include difficultly transcribed primary sequence elements, repetitive sequences capable of forming secondary structures, high nucleosome density, DNA damaged template, collisions with DNA or RNA polymerases, among others. The key distinction between transcription pausing and arrest relates to the elongation competency of RNAP II (Adelman and Lis, 2012). Paused polymerases persist in the pre-translocated state, enzymatically capable of adding the next nucleotide, with the 3' end of the nascent transcript contained within the active site of the polymerase. It is likely that such complexes may not serve as direct co-transcriptional substrates of RNA exosome as occlusion of the nascent RNA 3' end would prevent RNA exosome access. Indeed promoter proximal paused transcription complexes have been reported to be remarkably stable and resistant to RNA exosome mediated degradation (Henriques et al., 2013). Arrested transcription complexes however contain RNAP II molecules that are incapable of incorporating a nucleotide onto the growing nascent RNA due to a displacement of the RNA 3' end and the polymerase active site. Backtracking of RNA polymerase on DNA leads to transcription arrest. Such arrested complexes can transition back to the paused state through transcription elongation factor TFIIS mediated cleavage of the 3' end extruding portion of the nascent RNA (Adelman et al.,

2005). Nascent transcript sequencing in yeast has uncovered numerous intragenic RNAP II pause sites (Churchman and Weissman, 2011). A majority of these pause sites exhibit a downstream shift by ~18 nucleotides in TFIIS mutants (Churchman and Weissman, 2011), indicating that backtracking may be a relatively frequent event. However, backtracked transcription complexes can also be resolved through RNA exosome recognition of the extruding RNA 3' end resulting in degradation coupled transcription termination (Lemay et al., 2014). An intriguing possibility is that TFIIS and RNA exosome may functionally compete at arrested transcription complexes to promote elongation or degradation, respectively. Consistent with this model, TFIIS has been shown to be absent from AID complexes containing RNA exosome (Sun et al., 2013a). Small RNA deep sequencing has uncovered short RNA fragments of ~18 nucleotides that were proposed to be TFIIS cleavage products of backtracked RNAP II (Taft et al., 2009). If indeed backtracked transcription complexes are targeted by RNA exosome, and/or TFIIS and RNA exosome compete, using small RNA deep sequencing approaches one might expect to observe an accumulation of ~18 nucleotide long TFIIS-generated RNA fragments at genomic sites of RNA exosome substrate accumulation in either *Exosc3* or *Exosc10* deficient cells.

An inevitable product of transcription pausing or arrest is a decrease in transcription elongation rate. Temporary pausing of transcription is thought to facilitate the integration of RNA processing events such as capping, splicing, and polyadenylation (Zhou et al., 2012). Nucleosome remodeling and relief of DNA topological stress may also lag with respect to transcription and demand a pause in the progression of RNA polymerase before elongation can resume. A consequence of decreased transcription elongation is a kinetic shift towards degradation coupled transcription termination. Elongation and degradation coupled transcription termination represent antagonistic processes and in fact compete with one another in vivo. Mutant yeast expressing slowly

elongating RNAP II exhibit premature Sen1 dependent termination, which is coupled to RNA exosome mediated processing (Hazelbaker et al., 2013). Therefore, it is likely that obstacles and/or sequence elements that impede the progression of RNAP II, create a more permissive context for degradation coupled transcription termination.

Actively transcribed B cell *Ig* loci are endowed with multiple properties capable of inducing transcription pausing or arrest. For instance, *IgH* switch regions that precede $C_H$ gene segments and are the sites of AID mediated DNA double strand break formation, are highly enriched for guanosine nucleotides on the non-template DNA strand. Interestingly, recent genome-wide analyses of RNA polymerase pausing in *E. coli* have revealed a G-rich transcription pause consensus sequence (corresponding to the RNA or non-template DNA strand) comprised of $G_{-10}Y_{-1}G_{+1}$ (Larson et al., 2014; Vvedenskaya et al., 2014). Furthermore, *E. coli* RNA polymerase contains a specific structure known as the "core recognition element" that specifically binds with non-template strand guanosine residues in the +1 position, but not with any of the other three nucleotides (Zhang et al., 2012). As many fundamental aspects of transcription are conserved throughout evolution, it is interesting to speculate that similar pause determinants may be involved in eukaryotic transcription as well. If true, *IgH* switch region non-template strand G-richness may provide a direct sequence specific mechanism of favoring transcriptional pause or arrest through DNA-RNAP II interactions, and in turn lead to enhanced transcription termination and recruitment of RNA exosome and AID.

*IgH* switch sequences present obstacles to transcription elongation due to their strong propensity to form R-loops (Yu et al., 2003). As such, RNAP II accumulates within *IgH* switch sequences during germline transcription of $C_H$ genes in CSR, suggesting frequent polymerase pausing or arrest (Rajagopal et al., 2009; Wang et al., 2009b). R-loop formation and persistence can impede transcription elongation through multiple

mechanisms. For instance, an RNA polymerase generating an extended co-transcriptional R-loop in its wake is hindered in its ability to progress along the DNA template (Huertas and Aguilera, 2003). It is possible that the RNA:DNA hybrid trailing the RNA polymerase may create a tethering effect that limits the progression of the transcription complex. In conjunction, R-loop generation by the leading RNA polymerase can induce trailing RNA polymerases to pause or arrest. An elongating transcription complex presumably can collide with an R-loop in its downstream path, resulting in polymerase backtracking or arrest. Electron microscopy studies have reveled extensive RNA polymerase pileups at an R-loop forming locus in vivo (El Hage et al., 2010). Yet another feature of *IgH* switch sequences that may lend them to induce transcription stalling is their potential to create G-quadruplex structures. These are stable planar secondary DNA structures that form on G-rich ssDNA, and have been found to exist in the genomes of mammalian cells (Biffi et al., 2013; Lam et al., 2013). When transcribed in vitro, *IgH* switch sequences can generate G-quadruplex structures (Duquette et al., 2004). Displacement of the G-rich non-template strand during transcription of switch sequences provides an opportunity for G-quadruplex formation. In addition, stabilization of switch sequence R-loops may allow G-quadruplex structures to initiate and persist. Conversely, G-quadruplexes may stabilize the non-template strand in cis, and reduce its base pairing capacity to compete with RNA already hybridized with the template DNA strand. It also remains a possibility that even after resolution of an RNA:DNA hybrid, that G-quadruplexes on the non-template strand may inhibit proper re-annealing of the template and non-template DNA strands of the switch regions. A less than fully duplex DNA template would likely create an obstacle for transcription. A more speculative mechanism of transcription stalling at *Ig* loci may specifically involve AID deaminated cytidine target residues in the DNA. It is conceivable that non-native deoxyuridine serves as a poorer template when utilized by RNAP II. Coevolution between polymerases and

DNA may have optimized the interactions between RNAP II and the four canonical deoxynucleotides. As such, RNAP II may undergo a transcription pause upon encountering deoxyuridine. Recent studies have demonstrated that RNAP II accumulation also occurs at transcribed *Ig* variable regions during hypermutation (Wang et al., 2014). A deoxyuridine based mechanism of transcription stalling may explain RNAP II accumulation at *Ig* variable regions that are largely devoid of R-loop formation. *Ig* loci can provide such an opportunity as deoxyuridine residues sufficiently persist to allow for their direct biochemical detection in AID expressing B cells (Maul et al., 2011). Contrary to other mechanisms of transcription stalling, a deoxyuridine induced stall would serve as an AID dependent downstream event. It may also serve as a parallel mechanism of enhancing mutation clustering in conjunction with error prone DNA repair mechanisms of *Ig* diversification.

Multiple mechanisms therefore combine to hinder the elongation rate of transcription complexes at *Ig* loci. This in turn would create a preponderance of premature transcription termination leading to RNA exosome recruitment for the resolution of arrested transcription complexes (Figure 42). In fact, artificially inducing premature transcriptional arrest at an *Ig* locus through targeted insertion of a transcription termination element leads to enhanced hypermutation in the upstream region proximal to the site of transcription termination (Kodgire et al., 2013). An abundance of distinct transcription obstacles, particularly at *IgH*, may therefore create a synergistic effect with respect to premature transcription termination, which may in turn underlie the “targeting” phenomenon of AID towards *Ig* loci. AID mediated mutations accumulate at *Ig* loci at rates of approximately $10^2$ - $10^3$ fold higher frequency compared to other regions of the genome (Liu et al., 2008). Since AID requires transcription to gain access to substrate ssDNA, it is likely that the high rate of *Ig* transcription in B cells is at least partly responsible for the remarkably efficient targeting of AID to these loci.

However, transcription initiation cannot sufficiently explain the AID targeting phenomenon towards *Ig* loci, as several other equivalently transcribed genes accumulate far less mutations in AID expressing B cells (Chiarle et al., 2011; Klein et al., 2011b). It is more likely that premature transcription termination, rather than transcription initiation, accounts for AID targeting. Several factors involved in premature transcription termination have been found to associate with AID, most likely in the context of a larger transcription complex. RPA binds to ssDNA exposed during transcription and can physically associate with AID and stimulate its activity (Chaudhuri et al., 2004). Elongation factor Spt5 plays a critical role in inducing promoter proximal transcription pausing and associates with pause released transcription elongation complexes, has also been reported to interact with AID (Pavri et al., 2010). As mentioned earlier, RNA exosome and AID biochemically interact as well (Basu et al., 2011). In addition, E3 ubiquitin ligase Nedd4 ubiquitylates RNAP II at sites of DNA damage to promote removal of stalled transcription complexes through proteasomal degradation of RNAP II (Anindya et al., 2007). Nedd4 has recently been demonstrated to promote CSR and is present in a complex containing AID (Sun et al., 2013a). Importantly many of these factors interact with one another, suggesting that they operate within a larger complex to coordinate AID activities. For instance, in addition to AID, Spt5 associates with both RNAP II and RNA exosome (Andrulis et al., 2002; Wada et al., 1998). Furthermore, Nedd4 helps to stabilize the interaction between AID, Spt5, and RNA exosome at transcribed *IgH* switch sequences (Sun et al., 2013a), presumably by facilitating RNA exosome access to nascent transcript 3" ends through RNAP II proteolysis. Collectively, these factors may stabilize AID through multiple interactions at sites of premature transcription termination. In turn, a stable association between AID and termination factors coupled to strong transcriptional stalling at *Ig* loci, may synergize to produce a highly efficient means of directing AID activity in B cells.

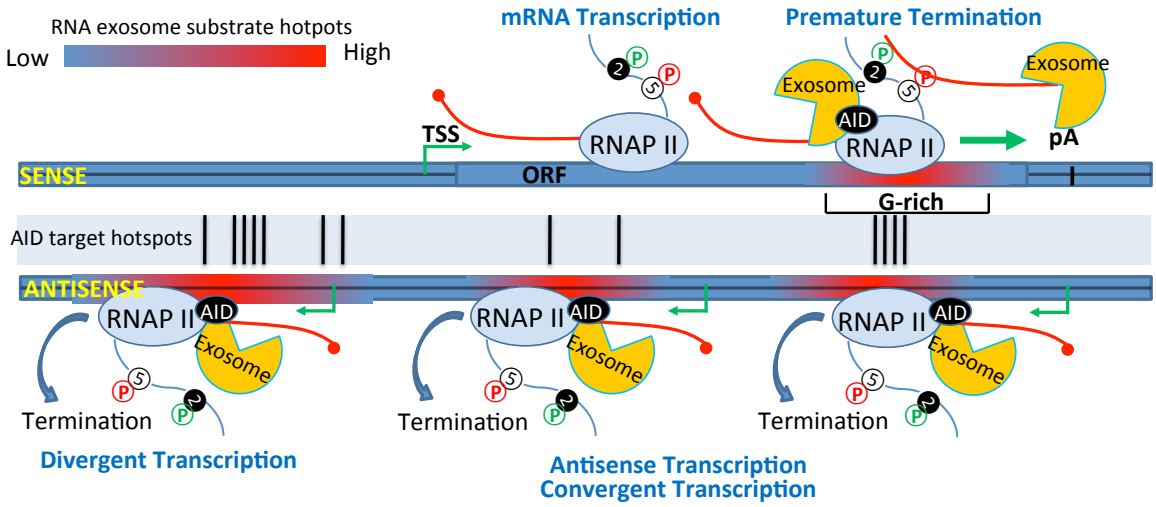



**Figure 42: Model of RNA exosome dependent AID recruitment to sites of premature RNAP II termination**
Sense and antisense post-initiation transcription complexes can undergo transcriptional arrest due to various transcription impediments such as G-rich DNA, R-loops, DNA damage, polymerase collisions, among others. Transcriptional arrest may then result in premature termination leading to RNA exosome and AID recruitment.

Interestingly, AID targeting exhibits a transcriptional context dependency that is tied to transcription initiation proximity. This was elegantly illustrated through the experiments of Peters and Storb where transgenic insertion of a $V_L$ promoter upstream of a $C_L$ gene segment allowed for hypermutation of an *Ig* sequence that would otherwise be spared of AID mediated mutation accumulation (Peters and Storb, 1996). Not only did this experiment disprove the notion that hypermutation of *Ig* variable regions and switch sequences was determined by sequence specific recognition of these loci, it also revealed a relationship between proximity to the site of transcription initiation and targeting of the hypermutation machinery. A clear illustration of this point can also be observed by comparing the related processes of SHM and CSR within the *IgH* locus of normal B cells. Transcription of the protein coding immunoglobulin heavy chain mRNA initiates from the cognate variable region promoter, and proceeds through the assembled VDJ exon, past a several kilobase long switch sequence containing intron, followed by the constant region. During SHM of *IgH* variable regions, AID mediated

mutations are largely targeted to the VDJ exon region while the much further downstream switch sequences are largely spared of mutations. Interestingly, these are the same G-rich, R-loop prone switch sequences that create strong obstacles to transcription elongation during CSR. However, a critical distinction between transcription through switch sequences during SHM and CSR is once again proximity to the site of transcription initiation. During CSR, transcription of switch sequences occurs through the context of $C_H$ germline transcripts that initiate not from the upstream distal $V_H$ promoter, but rather from a proximal $C_H$ promoter. Therefore, transcription through switch sequences can lead to differing outcomes with respect to AID targeting and is dependent upon proximity to the site of transcription initiation. This is likely to reflect a difference in transcription stalling efficiency relative to the distance of the transcription complex from the TSS. Transcriptionally engaged RNAP II undergoes strong transcription stalling in proximity to the promoter (Kwak and Lis, 2013). In contrast, productive elongation is achieved through the association of different complexes with RNAP II. For instance, association of engaged RNAP II with elongation factors such as ELL proteins, super elongation complexes, as well as the PAFc complex (Zhou et al., 2012) may require a temporal window, prior to which the transcription complex may be more prone to undergo transcription arrest. Interestingly, the length of RNA exosome substrate transcripts largely mirrors the restricted spreading of AID initiated mutations to less than 2 kilobases and peaking approximately 500 base pairs from the site of transcription initiation. This may reflect a critical distance where transcription complexes may be more prone to RNA exosome coupled premature termination. Beyond this distance from the TSS, transcription complexes may be less prone to premature termination due to association with elongation factors. Consistent with this model, NNS mediated premature transcription termination can be inhibited by the PAFc elongation complex in yeast (Kim and Levin, 2011).

Antisense transcription arising from intragenic loci may provide yet another mechanism leading to transcription arrest. Multiple genes including *Myc*, *Apobec3*, *Gimap5*, *Ly6a*, and *Ppp3cc* were shown here to express RNA exosome substrate antisense transcripts within gene bodies. Importantly, these sites of antisense intragenic transcription precisely overlapped with breakpoints of AID mediated recurrent chromosomal translocations. Since these genes also express sense mRNA transcripts, there is the potential for convergent transcription between sense and antisense transcription complexes. Therefore, AID dependent intragenic translocations at sites harboring RNA exosome substrate antisense transcripts shown here, could arise from transcriptional arrest and/or interference due to the convergence of head to head RNAP II complexes. However, this would necessitate that sense and antisense intragenic transcription are not mutually exclusive events and can occur on the same template and at the same time. Although this may be exceedingly difficult to observe experimentally in living cells, the possibility clearly exists in principle, and may be approachable in vitro. For instance, one could compare rates of AID association between convergent and unidirectional transcribed chromatin substrates in vitro for evidence of preferential AID recruitment at sites of convergent transcription. One consequence of convergent transcription is head to head collisions between oncoming RNA polymerases, as converging RNA polymerases cannot bypass one another in vitro (Hobson et al., 2012). In vivo, this could result in backtracking arrest of either converging RNAP II complex, leading to RNA exosome coupled premature transcription termination. Alternatively, convergent transcription may also impede transcription elongation through mechanisms not involving actual collisions between RNAP II. As transcription proceeds through DNA, opposite torsional stress accumulates in either direction. Specifically, positive and negative supercoiling of the DNA template occurs ahead of and behind the elongating transcription complex, respectively (Liu and Wang, 1987). During convergent

transcription, positive supercoiling of the DNA ahead of each RNA polymerase would create an area of extensive positive supercoiling lying between the two RNA polymerases. As positive supercoiling of DNA will impede transcription elongation (Gartenberg and Wang, 1992), convergent transcription may lead to transcriptional stalling prior to relaxation of the DNA by DNA gyrase. As discussed earlier, extended transcriptional stalling could ultimately lead to RNA exosome coupled premature transcription termination and AID recruitment to such sites. Consistent with this model, intragenic antisense transcripts accumulate in yeast upon depletion of an RNA exosome coupled transcription termination factor, resulting in transcription interference with convergent sense mRNA transcription complexes (Schulz et al., 2013).

### 4.2 RNA exosome in the maintenance of genomic integrity

Cells continuously operate under selective pressures to maintain the integrity of their genetic material during metabolic processes involving DNA such as transcription and DNA replication. R-loops formed during transcription disrupt the DNA duplex and pose a substantial threat for genomic instability, especially when the processing of such structures is impaired (Aguilera and Garcia-Muse, 2012; Chan et al., 2014; Skourti-Stathaki and Proudfoot, 2014). RNA:DNA hybrids can occur naturally within cells, predominantly during lagging strand DNA synthesis of Okazaki fragments and within the transcription bubble of elongating RNA polymerases. However, analyses performed on various RNA processing mutants have revealed that transient R-loop formation occurs frequently within cells. In fact, nearly all RNA processing pathways are involved to various degrees in the suppression of R-loop formation or persistence and preservation of genomic integrity (Wahba et al., 2011). Many RNA processing pathways have likely coevolved with respect to both RNA metabolic and DNA maintenance functions.

Co-transcriptional R-loops are thought to initiate via mechanisms involving the nascent transcript either invading the DNA duplex or intercepting the template DNA strand prior to its annealing with the non-template strand. Nascent RNA and template DNA exit through different channels within RNAP II (Westover et al., 2004), making it implausible that R-loops result from an extension of the RNA:DNA hybrid present within the transcription bubble. As a result, a temporal opportunity exists for RNA processing factors interacting with the nascent transcript to modulate R-loop formation. Such is the case with THO/TREX, a complex that binds nascent RNA co-transcriptionally to link mRNP formation with nuclear export. Yeast strains carrying deletion mutations in various THO subunits exhibit a transcription dependent hyper-recombination phenotype (Aguilera and Klein, 1990; Piruat and Aguilera, 1998). Importantly, THO mutants accumulate co-transcriptional R-loops and the hyper-recombination phenotype can be suppressed either through cleavage of the nascent transcript or overexpression of RNase H (Huertas and Aguilera, 2003). Similarly, THO depletion in human cells leads to R-loop accumulation and genomic instability (Dominguez-Sanchez et al., 2011). This suggests THO components bind to nascent RNA as it exits RNAP II forming an mRNP complex that inhibits hybridization between the nascent RNA and template DNA strand. Analogously, splicing factors also interact with nascent RNA to inhibit R-loop associated genomic instability. Depletion of splicing factor SRSF1 results in the accumulation of R-loops and DNA double strand breaks, both of which can be suppressed by RNase H overexpression (Li and Manley, 2005). Through multiple mechanisms cells have evolved ways to preserve the integrity of their genome by preventing co-transcriptional R-loop formation by intercepting the nascent transcript and inhibiting its annealing with the template DNA strand.

But how do R-loops create genomic instability in cells? R-loop exposure of the template DNA strand as ssDNA makes it far more accessible to chemical mutagenesis

arising from reactive oxygen species or other environmental insults (Kim and Jinks-Robertson, 2012). As previously mentioned, R-loops can serve as obstacles to transcription elongation, leading to RNA polymerase pileups and collisions (El Hage et al., 2010). Base modification of the exposed ssDNA component of R-loops may need to be bypassed using highly mutagenic translesion DNA polymerases. Replisomes can encounter nicks on the R-loop associated non-template DNA strand leading to the cessation of DNA synthesis and creation of DNA double strand breaks at the site where the DNA template nick overlaps with the site of nucleotide incorporation. However, a major mechanism of R-loop associated DNA damage arises from encounters between the replication and transcription machineries. In fact it appears that eukaryotic DNA replication and transcription have co-evolved to minimize such encounters. For instance, unlike prokaryotes, eukaryotic replisomes and transcription complexes elongate at similar speeds (Helmrich et al., 2013). This minimizes the opportunity for co-directional collisions between DNA and RNA polymerases. In addition, transcription during S-phase occurs largely at sites where replisomes are not actively performing DNA synthesis (Helmrich et al., 2013). However, impaired transcription elongation arising from R-loops increases the likelihood of collisions between replisomes and transcription complexes. These collisions typically result in stalling of the replication fork leading to DNA rearrangements through various mechanisms (Gan et al., 2011; Tuduri et al., 2009; Wellinger et al., 2006). Co-directional collisions between DNA polymerases and R-loop forming backtracked RNA polymerases can create DNA double strand breaks (Dutta et al., 2011). Head to head convergence between DNA polymerases and R-loop arrested transcription complexes can lead to replication fork reversal due to the accumulation of DNA torsional stress (Aguilera and Garcia-Muse, 2012). As the replisome approaches an arrested transcription complex head on, positive supercoiling will continue to build upstream of the replisome. This favors the creation of “chicken-foot” replication fork

structures that involve template switching of the nascent DNA chains. Resolution of such aberrant intermediates in DNA synthesis require recombination mediated repair, which can lead to DNA rearrangements. Furthermore, R-loop dependent collisions involving replisomes and transcription complexes may play an important role in DNA breakage at common fragile sites present in exceptionally long genes. Transcription of human genes exceeding 800 kilobases in length requires more than one cell cycle to complete (Helmrich et al., 2011), thus necessitating an encounter between the replisome and transcription complex. Such collisions result in R-loop formation and genomic instability at common fragile sites (Helmrich et al., 2011). Interestingly, chemical inhibition of topoisomerase 1, which promotes R-loop formation, preferentially reduces expression of long genes due to impaired transcription elongation (King et al., 2013). This suggests that R-loop formation may be more prevalent at long genes, and thereby exhibit more frequent transcription arrest, which in turn may promote the collisions observed between replisomes and transcription complexes at common fragile sites present in long genes.

As evidenced above, R-loops pose a serious risk for development of genomic instability. Cells utilize factors involved in RNP biogenesis, splicing, and DNA torsion relief to prevent R-loop formation. But how do cells resolve R-loops after they have formed, and what role might RNA exosome play in this process? Our findings indicate that RNA exosome is preferentially recruited to divergent promoters enriched in the R-loop associated H3S10ph chromatin mark. In addition, *Exosc3* deficiency lead to an increase in H3S10ph occupancy at RNA exosome targeted promoters. Similar trends were observed using RNA:DNA hybrid specific immunoprecipitation. Collectively, these data suggest that RNA exosome is recruited to R-loop forming loci and suppresses their accumulation, likely through degradation. Multiple lines of evidence support an active role for RNA exosome in the resolution of R-loops. Degradation coupled transcription termination in yeast is thought to initiate through binding of the Nrd1 termination factor

with the carboxyl terminal domain of RNAP II (Steinmetz et al., 2001). As described earlier, Nrd1 is a component of the Nrd1-Nab3-Sen1 (NNS) transcription termination complex. NNS binding to RNAP II is then stabilized through the TRAMP oligoadenylation complex (Grzechnik and Kufel, 2008). This serves to enhance Nrd1 recruitment of RNA exosome to sites of NNS mediated transcription termination (Vasiljeva and Buratowski, 2006). Importantly, the Sen1 subunit of the NNS termination complex possesses RNA:DNA hybrid helicase activity (Kim et al., 1999). Mutation of the Sen1 helicase domain results in R-loop accumulation, increased DNA damage foci, and transcription dependent hyper-recombination (Mischo et al., 2011). Similarly, depletion of senataxin, the mammalian homolog of Sen1, results in transcription termination defects and R-loop accumulation in human cells (Skourti-Stathaki et al., 2011). Furthermore, SUMO conjugated senataxin directly interacts with the Rrp45 core subunit of RNA exosome at sites of transcription dependent DNA damage (Richard et al., 2013). It is becoming increasingly evident that transiently formed co-transcriptional R-loops are quite prevalent and cells utilize premature transcription termination and RNA exosome mediated degradation to suppress their persistence. RNA exosome dependent R-loop resolution would serve as a critical mechanism for preventing genomic instability. One expectation arising from such a model is that RNA exosome deficient cells would display elevated levels of genomic instability.

Chromosomal translocations are a characteristic feature of diffuse large B-cell lymphoma (DLBCL). Karyotype analysis of clinical DLBCL specimens places the incidence of chromosomal abnormalities at nearly 90% of cases (Cigudosa et al., 1999). DLBCL primarily derives from AID expressing germinal center B cells or post-germinal center B cells. As mentioned earlier, studies in mouse B cells have revealed a large number of AID dependent chromosomal translocations (Chiarle et al., 2011; Klein et al., 2011b). Is it possible that the high prevalence of chromosomal translocations in DLBCL

might be due to a combination of R-loops, AID activity, and RNA exosome mediated targeting of AID? Unstable proto-oncogenes mutated in DLBCL are highly enriched for R-loop promoting non-template DNA strand G-rich sequences in comparison to translocation sites present in malignant T cells or AID negative B cells (Duquette et al., 2007). Also, it has been reported that AID expression leads to a decrease in topoisomerase 1 expression in B cells (Kobayashi et al., 2009). Topoisomerase 1 relieves torsional stress derived from negative DNA supercoils to suppress R-loop formation. Therefore, it is intriguing to speculate that AID expressing germinal center B cells may operate in a state more conducive to the formation of genome destabilizing co-transcriptional R-loops. In addition, AID may also play a more active role in the initiation of R-loops. In addition to non-template DNA strand G-richness, a strong determinant of R-loop initiation is a nick in the non-template strand (Roy et al., 2010). As described earlier, deoxyuridine residues resulting from AID mediated deamination of deoxycytidine, are processed by BER and MMR DNA repair pathways involving nicked DNA intermediates. These AID derived nicks may promote R-loop initiation leading to transcriptional arrest and RNA exosome recruitment allowing for the stabilization of AID at these sites. This model is consistent with a previous observation indicating that AID is required for the recruitment of RNA exosome to transcribed *IgH* switch regions in B cells (Basu et al., 2011). Topoisomerase downregulation and AID dependent nicking of the non-template DNA strand may synergize to promote R-loop formation specifically in germinal center B cells. In turn, the greater need to resolve these R-loops would enhance the frequency of RNA exosome recruitment onto chromatin. Coupled with the observation that RNA exosome stabilizes AID occupancy at its target sites, enhanced R-loop formation in AID expressing germinal center B cells may ultimately lead to the DNA rearrangement events that are so frequently observed in DLBCL.

Burkitt's lymphoma is an aggressive form of B-cell non-Hodgkin's lymphoma characterized by rapidly proliferating malignant B cells. Nearly all cases of Burkitt's lymphoma involve chromosomal translocations between *MYC* and one of the immunoglobulin loci (Boxer and Dang, 2001). Approximately 80% of these cases specifically involve translocations between *MYC* and *IgH*. Burkitt's lymphoma translocation breakpoints at *MYC* cluster within 3 regions (Boxer and Dang, 2001). Class I breakpoints map to a region close to the first exon and intron of *MYC*. Class II breakpoints cluster upstream, but proximal, to *MYC* exon 1. Lastly, the class III *MYC* breakpoints map several hundreds of kilobases away from *MYC*. Interestingly, the class I breakpoint overlaps with a transcriptional pause site. Nuclear runoff assays in a promyelocytic leukaemia cell line revealed a transcription elongation block mapping to the boundary between exon 1 and intron 1 at *MYC* (Bentley and Groudine, 1986). This same region is predicted to contain an R-loop forming sequence according to a computational algorithm created to identify potential R-loop sites genome-wide (Wongsurawat et al., 2012). Furthermore, ectopic expression of AID in yeast leads to chromosomal translocations at a transcribed human *MYC* sequence containing the first exon and intron of *MYC* (Ruiz et al., 2011). These heterologous *MYC* translocations were greatly enhanced in a THO deficient background, suggesting that the *MYC* translocations observed arise due to R-loop formation. Our data revealed significant R-loop associated H3S10ph chromatin marks at *Myc*. We also uncovered RNA exosome substrate antisense transcripts precisely overlapping two distinct *Myc* translocation hotspots in B cells. One hotspot mapped immediately upstream of *Myc* exon 1, the other mapped at the boundary region between exon 1 and intron 1. Importantly, these two regions of RNA exosome substrate antisense transcription are the same two regions where *MYC* proximal breakpoints cluster in Burkitt's lymphoma. R-loop mediated premature termination of antisense transcription may facilitate RNA exosome mediated

targeting of AID to these two sites, ultimately resulting in DNA breakage at *MYC*. These breaks could then go on to ligate to DNA breaks frequently arising at *IgH* during immunoglobulin diversification to give rise to the highly oncogenic *MYC/IgH* rearrangements that drive the pathogenesis of Burkitt's lymphoma.

Human mutations in RNA exosome subunit genes have been linked to heritable neurodegenerative disorders. Homozygous recessive mutations in *EXOSC3* are responsible for pontocerebellar hypoplasia type 1 (PCH1) (Wan et al., 2012). Similarly, homozygous recessive mutations in *EXOSC8*, encoding another RNA exosome core subunit, results in cerebellar hypoplasia (Boczonadi et al., 2014). Two distinct heritable neurodegenerative disorders arise from mutations in *SETX*, which encodes the RNA exosome cofactor senataxin. Recessive *SETX* mutations are responsible for ataxia oculomotor apraxia type 2 (AOA2) (Moreira et al., 2004), while dominant *SETX* mutations are tied to a juvenile form of amyotrophic lateral sclerosis (ALS4) (Chen et al., 2004). A mechanistic understanding is presently unclear regarding how these mutations give rise to the clinical features of these disorders. Might an accumulation of genome destabilizing R-loops be partly responsible for these developmental phenotypes? Interestingly, the recently described interaction between senataxin and RNA exosome at sites of transcription dependent DNA damage is abolished in AOA2 specific *SETX* variants (Richard et al., 2013). Furthermore, as all of the neurodegenerative disorders described here are heritable, they involve rare gene variants that circulate in human populations. These recessive variants may predispose otherwise healthy subjects to R-loop associated genomic instability leading to more common malignant transformations. For instance, recurrent somatic mutations in the RNA exosome nuclease subunit gene *DIS3*, have been reported in multiple myeloma (Chapman et al., 2011) and acute myeloid leukemia (Ding et al., 2012).

## 4.3 Mechanisms of RNA exosome targeting to divergently transcribed promoters

Until relatively recently, transcription initiation in eukaryotes was thought to occur unidirectionally. In the case of mRNA, transcription initiation was to proceed in the direction of the translational reading frame. However, the development of nascent RNA labeling and purification schemes coupled with genome-wide deep sequencing has unequivocally demonstrated the bidirectional nature of eukaryotic transcription initiation (Core et al., 2008; Seila et al., 2008). The underlying cause of bidirectional transcription initiation ultimately relates back to eukaryotic promoter structure. Classical TATA box containing promoters are specifically recognized by the TATA-binding protein (TBP) component of the TFIID general transcription factor complex. Typically, approximately 30 base pairs away from the TATA box lies an initiator sequence containing the TSS (Sandelin et al., 2007). These two distinct elements, the TATA box and initiator, create a vector that provides directional information allowing for unidirectional transcription initiation at TATA box containing promoters. However, only approximately 15% of human promoters contain identifiable TATA box elements (Cooper et al., 2006). Instead, the majority of mammalian promoters lacking TATA elements contain CpG dinucleotide rich sequences known as CpG islands (Sandelin et al., 2007). Transcription at CpG island promoters is thought to initiate through CpG-binding transcription factors such as Sp1, which in turn specifically recruit TFIID through direct protein-protein interactions (Hoey et al., 1993). As a result, TFIID binding and ultimately pre-initiation complex formation occurs on either side of CpG-binding transcription factors, leading to divergently directed bidirectional transcription initiation (Wu and Sharp, 2013). Incidentally, CpG island promoters are rare in *Drosophila*, and consequently transcription initiation is largely unidirectional in these species (Core et al., 2012). Transcription factor binding to DNA must overcome competing interactions from histones for the same site. This competition

between transcription factors and histones gives rise to nucleosome depleted regions of chromatin (Bai et al., 2011; Floer et al., 2010). As expected, transcriptionally active mammalian promoters contain pre-initiation complexes at nucleosome depleted TSS regions (Ozsolak et al., 2007). Furthermore, proximal H3K4me3 histone modifications can stimulate pre-initiation complex formation (Lauberth et al., 2013). Consistent with prevalent bidirectional transcription initiation, genome-wide chromatin immunoprecipitation studies have revealed a bivalent symmetrical distribution of H3K4me3 flanking the TSS at mammalian mRNA encoding genes (Barski et al., 2007; Guenther et al., 2007), thus promoting bidirectional pre-initiation complex formation. In fact, two distinct bidirectional transcription complexes can simultaneously occupy a single allele in divergent orientation (Rhee and Pugh, 2012). Despite that, promoter proximal antisense divergent ncRNA such as xTSS-RNA characterized in this work, as well as analogous transcripts observed in different species and by other laboratories (Flynn et al., 2011; Neil et al., 2009; Preker et al., 2008; Xu et al., 2009), are far more unstable compared to their cognate mRNA partners and require inhibition of RNA exosome activity for their reliable detection. Therefore, promoter directionality in eukaryotes is governed not through specific orientation queues during transcription initiation, but rather through differential degradation of divergent sense and antisense transcript pairs.

Recent studies have revealed degradation-coupled early transcription termination as a key determinant underlying promoter directionality in human and mouse. Coding strand motif analyses indicate antisense divergent transcripts are significantly enrichment in promoter proximal polyadenylation signals compared to sense pre-mRNA transcripts (Almada et al., 2013; Ntini et al., 2013). Polyadenylation signals play a critical role in transcription termination of RNAP II transcribed coding genes by specifying the site of endonucleolytic RNA cleavage and polyadenylation (Proudfoot, 2011). Inversely,

5’ splice site sequences recognized by U1 snRNP are strongly enriched within sense pre-mRNA transcripts in comparison with antisense divergent ncRNA transcripts (Almada et al., 2013; Ntini et al., 2013). The relevance of this distribution bias in U1 snRNP binding sites lies in the fact that the U1 snRNP complex can inhibit transcription termination by suppressing polyadenylation signal utilization (Berg et al., 2012; Kaida et al., 2010). Importantly, U1 snRNP mediated suppression of transcription termination occurs within a 1 kilobase distance from the 5’ splice site sequence and operates most effectively proximal to the site of transcription initiation. Together, polyadenylation signal enrichment within antisense divergent transcripts and U1 snRNP binding site enrichment at sense pre-mRNA transcripts work in opposite fashion through transcription termination to impose a strong net transcriptional directionality on an otherwise inherently bidirectional process. It should be noted that additional mechanisms have been shown to promote transcriptional directionality in eukaryotic cells. Gene looping between sites of transcription initiation and termination bring these two regions of chromatin in close spatial proximity. Mutations in Ssu72, a phosphatase associating with sites of transcription initiation and termination, can disrupt gene looping between these sites (Tan-Wong et al., 2012). Surprisingly, disruption of gene looping leads to the accumulation of promoter derived antisense divergent transcripts (Tan-Wong et al., 2012). It is presumed that looping of the termination and initiation sites promotes transcriptional directionality by allowing terminating RNAP II to reinitiate at the promoter. Remodeling of nucleosome positioning can also play a role in transcriptional directionality. In yeast, the Isw2 chromatin remodeling complex directionally repositions nucleosomes towards intergenic noncoding regions and away from coding sequences near sites of transcription initiation (Whitehouse et al., 2007). Loss of Isw2 function results in cryptic transcription arising from noncoding regions adjacent to genic coding sites, presumably due to nucleosome depletion at these regions. More recently, an

unbiased screen for regulators of promoter directionality in yeast has revealed chromatin assembly factor I (CAF-I) as a repressor of promoter derived antisense divergent transcription (Marquardt et al., 2014). CAF-I introduces H3K56ac modified histones onto DNA, a chromatin mark enriched at sites of rapidly exchanged nucleosomes such as nucleosome depleted regions surrounding promoters (Dion et al., 2007; Rufiange et al., 2007). Presumably, CAF-I inhibits antisense divergent transcription by limiting the nucleosome depleted residency time of promoter upstream sequences by depositing H3K56a histones onto these regions, thus preventing DNA access to transcription initiation factors. Therefore, assembly and remodeling of chromatin can also play important roles in the control of promoter directionality.

As discussed above, antisense divergent transcripts are geared to undergo early termination. It remains an open question as to why and how these early terminating transcripts are specified for RNA exosome mediated degradation. Interestingly, RNAP II termination involving polyadenylation signals, such as those present in pre-mRNA, are recognized by cleavage and polyadenylation factors that ultimately promote transcript stabilization. However, these same polyadenylation signals trigger degradation-coupled termination in the context of antisense divergent transcripts such as xTSS-RNA. It has been proposed that proximity to the site of transcription initiation may dictate the choice between RNA polyadenylation versus degradation at polyadenylation signals (Jensen et al., 2013). In yeast, RNA exosome sensitive antisense divergent transcripts are terminated shortly after initiation via the Nrd1-Nab3-Sen1 (NNS) termination complex (Arigo et al., 2006; Thiebaut et al., 2006). The NNS complex physically interacts with RNA exosome and its cofactor adenylation complex TRAMP (Vasiljeva and Buratowski, 2006), coupling early transcription termination with RNA exosome mediated degradation. The TSS position effect on RNA exosome coupled NNS termination involves interactions between the NNS complex and RNAP II C-terminal domain (CTD) heptad repeat.

Specific RNAP II CTD post-translational modifications are generally associated with different stages of RNAP II elongation status. Serine-5 phosphorylation of RNAP II CTD is largely associated with initiating or early elongating RNAP II, whereas serine-2 CTD phosphorylations typically mark RNAP II molecules during productive elongation (Egloff et al., 2012). The Nrd1 subunit of the NNS complex interacts far more strongly with CTD phosphorylations on serine-5 residues compared with serine-2 (Vasiljeva et al., 2008), suggesting that early elongation RNAP II CTD modifications may be implicated in the TSS position effect on degradation coupled early transcription termination. Along with binding to RNAP II CTD, Nrd1 also serves as an RNA binding protein (Creamer et al., 2011; Wlotzka et al., 2011). Interestingly, Nrd1 RNA binding motifs are enriched within TSS associated antisense divergent transcripts compared to partner coding transcripts (Schulz et al., 2013). Consistent with this observation, Nrd1 depletion leads to a significant accumulation of TSS associated antisense divergent transcripts, while sense coding transcripts are hardly affected (Schulz et al., 2013). Collectively, NNS mediated termination, which is tightly coupled with RNA exosome mediated degradation, restricts antisense divergent transcription through both early elongation RNAP II CTD modifications and specific RNA motifs enriched within antisense divergent transcripts. It is also noted that histone modifications also play a role in differentially promoting degradation coupled transcription termination in the antisense direction. H3K79me2 and H3K36me3 modified histones occupy regions of chromatin associated with transcription elongation and are strongly enriched in sense coding regions relative to antisense non-coding regions of divergent transcript pairs (Barski et al., 2007; Seila et al., 2008). H3K36me3 has been found to facilitate recruitment of the histone chaperone FACT complex to promote transcription elongation (Carvalho et al., 2013). Thus, the paucity of H3K36me3 chromatin marks at regions of antisense divergent transcription could result

in greater frequency of RNAP II arrest to further promote degradation coupled early transcriptional arrest over these regions.

Surprisingly, orthologs of NNS termination complex subunits Nrd1 and Nab3 are absent in higher eukaryotes (Jensen et al., 2013), despite being essential for viability in yeast (Conrad et al., 2000; Steinmetz and Brow, 1996). In contrast, the Sen1 helicase subunit of NNS has been retained as human ortholog senataxin (*SETX*) (Moreira et al., 2004). Sen1 and senataxin likely share some of the same functions as deficiency of either factor leads to impaired transcription termination (Mischo et al., 2011; Suraweera et al., 2009). Early studies seeking to uncover interacting proteins of the NNS complex using tagged Nrd1 confirmed the known interaction with Sen1, but also revealed an interaction involving both subunits of the heterodimeric nuclear cap-binding complex (Vasiljeva and Buratowski, 2006). As mentioned earlier, in human cells RNA exosome physically associates with components of the nuclear cap-binding complex (Andersen et al., 2013). Interestingly, depletion of cap-binding complex subunits leads to impaired transcription termination and a synergistic hyper-accumulation of antisense divergent non-coding transcripts in conjunction with RNA exosome depletion (Andersen et al., 2013). It is possible that in mammalian cells senataxin cooperates with the nuclear cap-binding complex to facilitate the recruitment of RNA exosome to arrested antisense divergent transcription complexes.

Multiple reasons may exist for why cells would need to rapidly degrade noncoding transcripts arising from antisense divergent transcription. For instance, what harms might be posed if cells allowed such prevalent transcripts to accumulate under normal situations? One possibility is that antisense divergent transcripts could contain short open reading frames that could give rise to deleterious peptides possessing dominant negative activities or which might be prone to aggregation. If allowed to persist, antisense divergent transcripts could conceivably be processed by cellular RNA

interference machinery into miRNA to deregulate gene expression. Alternatively, antisense divergent ncRNA might function as molecular sponges by competing with mRNA 3' UTRs for miRNA bound RNA-induced silencing complex (RISC) binding. Furthermore, the act of divergent transcription creates a zone of negative DNA supercoiling between the two RNA polymerases. As negative supercoils promote R-loop formation, degradation of associated R-loops may promote transcription and help to maintain genomic stability. Incidentally, this process is likely aberrantly exploited by AID resulting in TSS proximal DNA damage leading to chromosomal translocations in B cells.

Given the risks associated with divergent non-coding transcription, it is interesting to speculate why eukaryotic cells not only tolerate it, but also apparently select for it. Remarkably, 10% of protein coding genes in the human genome are organized in a divergent orientation separated by less than 1 kilobase (Wu and Sharp, 2013). Furthermore, a majority of long non-coding RNA (lncRNA) found in mouse embryonic stem cells are derived from divergently oriented lncRNA-mRNA gene pairs (Sigova et al., 2013). Evidence of positive evolutionary selection for divergent transcription can be observed in species specific lncRNA as well. Comparative genomic studies have revealed over 1,000 divergently transcribed human lncRNAs lack an equivalent mouse ortholog (Gotea et al., 2013). These observations have prompted some to propose that RNA exosome targeted antisense divergent transcripts may serve as evolutionary precursors of lncRNA and mRNA (Wu and Sharp, 2013). In fact, correlative evidence linking transcript stability with sequence content supports this hypothesis. As mentioned above, a consequence of divergent transcription is stabilization of R-loops due to enhanced negative supercoiling of intervening DNA. This would prolong the exposure time of the non-template DNA strand as ssDNA and increases its susceptibility to environmental deamination leading to gains in guanosine

and thymidine residues (Wu and Sharp, 2013). In fact, transcribed regions have produced a mutational strand asymmetry during the course of mammalian evolution resulting in greater guanosine and thymidine accumulation on the non-template DNA strand (Green et al., 2003). Sharp and colleagues have proposed that the aforementioned asymmetric mutational bias may favor the gain of guanosine/thymidine-rich U1 snRNP binding sites and loss of adenosine-rich polyadenylation signals (Wu and Sharp, 2013). This in turn would promote stabilization of RNA exosome targeted antisense divergent transcripts by reducing transcription termination frequency and lengthening the transcription unit. Consistent with this model, RNA exosome substrate antisense divergent ncRNA-mRNA, lncRNA-mRNA, and mRNA-mRNA divergent gene pairs display increasing ratios of U1 snRNP binding sites to polyadenylation signals in the antisense direction, and consequently resulting in increasing transcript stability (Almada et al., 2013). Therefore, in addition to gene duplication followed by sequence divergence, pervasive antisense divergent transcription may create an opportunity for *de novo* gene formation through natural selection.

### 4.4 Concluding remarks

Recent major advances in nucleic acid sequencing technologies have made it increasingly clear that vast portions of eukaryotic genomes are extensively transcribed as noncoding RNA. These transcripts are usually short lived and have long evaded detection. The experiments and analyses performed in this thesis strongly support a role for RNA exosome in the processing and/or degradation of a large class of noncoding transcripts in mammalian cells. Most prominently, RNA exosome substrate noncoding transcripts are abundantly transcribed upstream of many mRNA expressing promoters in an antisense divergent orientation. In addition, a large number of cryptic intragenic initiating antisense transcripts are suppressed through RNA exosome mediated

degradation processes. A consequence of promoter proximal divergent transcription and/or intragenic antisense transcription is the creation of torsional stress or polymerase interference that can promote transient R-loop formation leading to premature transcription termination. RNA exosome is likely recruited to sites of arrested transcription for the purpose of resolving these potentially genome destabilizing structures. Indeed we observed that genomic sites displaying RNA exosome activity displayed higher levels of termination promoting R-loops.

SHM and CSR are two critical B cell pathways involved in the generation of high affinity effector function enabled immunoglobulins. These pathways are initiated via AID mediated transcription coupled deoxycytidine deamination of *Ig* loci. Many studies have revealed extensive transcriptional stalling at *Ig* loci in activated B cells. In the case of *IgH*, structural and sequence determinants such as asymmetric G-richness and repetitiveness can promote R-loop formation that consequently lead to premature transcription termination. Through its interaction with RNA exosome, AID likely gains access to substrate DNA by exploiting cellular mechanisms that resolve arrested transcription complexes through degradation coupled termination.

Although B cells display a remarkable ability to target the mutagenic activity of AID largely to *Ig* loci, the rest of the genome is not entirely spared. For instance, it has long been appreciated that AID is responsible for initiating DNA breaks at *MYC* and *IGH* that ultimately give rise to oncogenic chromosomal translocations driving B cell transformation in Burkitt's lymphoma. In fact, genomic studies of AID occupancy on chromatin have revealed significant AID binding throughout the genome. It has also been demonstrated that AID can introduce extensive amounts of DNA damage covering wide portions of the B cell genome resulting in chromosomal translocations. Findings presented here indicate a significant association between sites of RNA exosome targeted noncoding transcription and recurrent AID dependent chromosomal

translocations. This supports a model whereby RNA exosome mediated RNA processing events recruit AID to arrested noncoding transcription complexes resulting in DNA damage at such sites, thereby creating an important link between transcription coupled RNA processing and genomic integrity in B cells.

# References

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

Abremski, K., Hoess, R., and Sternberg, N. (1983). Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination. Cell *32*, 1301-1311.

Adams, D.J., Quail, M.A., Cox, T., van der Weyden, L., Gorick, B.D., Su, Q., Chan, W.I., Davies, R., Bonfield, J.K., Law, F.*, et al.* (2005). A genome-wide, end-sequenced 129Sv BAC library resource for targeting vector construction. Genomics *86*, 753-758.

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet *13*, 720-731.

Adelman, K., Marr, M.T., Werner, J., Saunders, A., Ni, Z., Andrulis, E.D., and Lis, J.T. (2005). Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. Mol Cell *17*, 103-112.

Aguilera, A., and Garcia-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. Mol Cell *46*, 115-124.

Aguilera, A., and Klein, H.L. (1990). HPR1, a novel yeast gene that prevents intrachromosomal excision recombination, shows carboxy-terminal homology to the Saccharomyces cerevisiae TOP1 gene. Mol Cell Biol *10*, 1439-1451.

Allen, C.D., Ansel, K.M., Low, C., Lesley, R., Tamamura, H., Fujii, N., and Cyster, J.G. (2004). Germinal center dark and light zone organization is mediated by CXCR4 and CXCR5. Nat Immunol *5*, 943-952.

Allen, C.D., Okada, T., Tang, H.L., and Cyster, J.G. (2007). Imaging of germinal center selection events during affinity maturation. Science *315*, 528-531.

Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E., and Tollervey, D. (1999a). Functions of the exosome in rRNA, snoRNA and snRNA synthesis. EMBO J *18*, 5399-5410.

Allmang, C., Petfalski, E., Podtelejnikov, A., Mann, M., Tollervey, D., and Mitchell, P. (1999b). The yeast exosome and human PM-Scl are related complexes of 3' --> 5' exonucleases. Genes Dev *13*, 2148-2158.

Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B., and Sharp, P.A. (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. Nature *499*, 360-363.

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

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.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Andersen, P.R., Domanski, M., Kristiansen, M.S., Storvall, H., Ntini, E., Verheggen, C., Schein, A., Bunkenborg, J., Poser, I., Hallais, M.*, et al.* (2013). The human cap-binding complex is functionally connected to the nuclear RNA exosome. Nat Struct Mol Biol *20*, 1367-1376.

Anderson, J.S., and Parker, R.P. (1998). The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. EMBO J *17*, 1497-1506.

Anderson, L., Henderson, C., and Adachi, Y. (2001). Phosphorylation and rapid relocalization of 53BP1 to nuclear foci upon DNA damage. Mol Cell Biol *21*, 1719-1729.

Andrulis, E.D., Werner, J., Nazarian, A., Erdjument-Bromage, H., Tempst, P., and Lis, J.T. (2002). The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila. Nature *420*, 837-841.

Anindya, R., Aygun, O., and Svejstrup, J.Q. (2007). Damage-induced ubiquitylation of human RNA polymerase II by the ubiquitin ligase Nedd4, but not Cockayne syndrome proteins or BRCA1. Mol Cell *28*, 386-397.

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.

Arigo, J.T., Eyler, D.E., Carroll, K.L., and Corden, J.L. (2006). Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. Mol Cell *23*, 841-851.

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.

Bai, L., Ondracka, A., and Cross, F.R. (2011). Multiple sequence-specific factors generate the nucleosome-depleted region on CLN2 promoter. Mol Cell *42*, 465-476.

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.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Basso, K., and Dalla-Favera, R. (2010). BCL6: master regulator of the germinal center reaction and key oncogene in B cell lymphomagenesis. Adv Immunol *105*, 193-210.

Basso, K., Saito, M., Sumazin, P., Margolin, A.A., Wang, K., Lim, W.K., Kitagawa, Y., Schneider, C., Alvarez, M.J., Califano, A*., et al.* (2010). Integrated biochemical and computational approach identifies BCL6 direct target genes controlling multiple pathways in normal germinal center B cells. Blood *115*, 975-984.

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. Mol Cell *32*, 285-291.

Bentley, D.L., and Groudine, M. (1986). A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL60 cells. Nature *321*, 702-706.

Berg, M.G., Singh, L.N., Younis, I., Liu, Q., Pinto, A.M., Kaida, D., Zhang, Z., Cho, S., Sherrill-Mix, S., Wan, L*., et al.* (2012). U1 snRNP determines mRNA length and regulates isoform expression. Cell *150*, 53-64.

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

Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013). Quantitative visualization of DNA G-quadruplex structures in human cells. Nat Chem *5*, 182-186.

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

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

Bochman, M.L., Paeschke, K., and Zakian, V.A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. Nat Rev Genet *13*, 770-780.

Boczonadi, V., Muller, J.S., Pyle, A., Munkley, J., Dor, T., Quartararo, J., Ferrero, I., Karcagi, V., Giunta, M., Polvikoski, T.*, et al.* (2014). EXOSC8 mutations alter mRNA metabolism and cause hypomyelination with spinal muscular atrophy and cerebellar hypoplasia. Nat Commun *5*, 4287.

Bonneau, F., Basquin, J., Ebert, J., Lorentzen, E., and Conti, E. (2009). The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. Cell *139*, 547-559.

Both, G.W., Taylor, L., Pollard, J.W., and Steele, E.J. (1990). Distribution of mutations around rearranged heavy-chain antibody variable-region genes. Mol Cell Biol *10*, 5187-5196.

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

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

Bousquet-Antonelli, C., Presutti, C., and Tollervey, D. (2000). Identification of a regulated pathway for nuclear pre-mRNA turnover. Cell *102*, 765-775.

Boxer, L.M., and Dang, C.V. (2001). Translocations involving c-myc and c-myc function. Oncogene *20*, 5595-5610.

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.

Briggs, M.W., Burkard, K.T., and Butler, J.S. (1998). Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3' end formation. J Biol Chem *273*, 13255-13263.

Brogden, K.A. (2005). Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nat Rev Microbiol *3*, 238-250.

Burma, S., Chen, B.P., Murphy, M., Kurimasa, A., and Chen, D.J. (2001). ATM phosphorylates histone H2AX in response to DNA double-strand breaks. J Biol Chem *276*, 42462-42467.

Buttner, K., Wenig, K., and Hopfner, K.P. (2005). Structural framework for the mechanism of archaeal exosomes in RNA processing. Mol Cell *20*, 461-471.

Carney, J.P., Maser, R.S., Olivares, H., Davis, E.M., Le Beau, M., Yates, J.R., 3rd, Hays, L., Morgan, W.F., and Petrini, J.H. (1998). The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: linkage of double-strand break repair to the cellular DNA damage response. Cell *93*, 477-486.

Carroll, M.C., and Isenman, D.E. (2012). Regulation of humoral immunity by complement. Immunity *37*, 199-207.

Carvalho, S., Raposo, A.C., Martins, F.B., Grosso, A.R., Sridhara, S.C., Rino, J., Carmo-Fonseca, M., and de Almeida, S.F. (2013). Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription. Nucleic Acids Res *41*, 2881-2893.

Casellas, R., Nussenzweig, A., Wuerffel, R., Pelanda, R., Reichlin, A., Suh, H., Qin, X.F., Besmer, E., Kenter, A., Rajewsky, K.*, et al.* (1998). Ku80 is required for immunoglobulin isotype switching. EMBO J *17*, 2404-2411.

Castellano-Pozo, M., Santos-Pereira, J.M., Rondon, A.G., Barroso, S., Andujar, E., Perez-Alegre, M., Garcia-Muse, T., and Aguilera, A. (2013). R loops are linked to histone h3 s10 phosphorylation and chromatin condensation. Mol Cell *52*, 583-590.

Chahwan, R., Edelmann, W., Scharff, M.D., and Roa, S. (2012). AIDing antibody diversity by error-prone mismatch repair. Semin Immunol *24*, 293-300.

Chan, Y.A., Hieter, P., and Stirling, P.C. (2014). Mechanisms of genome instability induced by RNA-processing defects. Trends Genet *30*, 245-253.

Chapman, M.A., Lawrence, M.S., Keats, J.J., Cibulskis, K., Sougnez, C., Schinzel, A.C., Harview, C.L., Brunet, J.P., Ahmann, G.J., Adli, M.*, et al.* (2011). Initial genome sequencing and analysis of multiple myeloma. Nature *471*, 467-472.

Chaudhuri, J., Basu, U., Zarrin, A., Yan, C., Franco, S., Perlot, T., Vuong, B., Wang, J., Phan, R.T., Datta, A.*, et al.* (2007). Evolution of the immunoglobulin heavy chain class switch recombination mechanism. Adv Immunol *94*, 157-214.

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.

Chen, C.Y., Gherzi, R., Ong, S.E., Chan, E.L., Raijmakers, R., Pruijn, G.J., Stoecklin, G., Moroni, C., Mann, M., and Karin, M. (2001). AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. Cell *107*, 451-464.

Chen, Y.Z., Bennett, C.L., Huynh, H.M., Blair, I.P., Puls, I., Irobi, J., Dierick, I., Abel, A., Kennerson, M.L., Rabin, B.A.*, et al.* (2004). DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). Am J Hum Genet *74*, 1128-1135.

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. Proc Natl Acad Sci U S A *106*, 2717-2722.

Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molinie, B., Ho, Y.J., Myers, D.R., Choi, V.W., Compagno, M., Malkin, D.J., *et al.* (2011). Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. Cell *147*, 107-119.

Chlebowski, A., Lubas, M., Jensen, T.H., and Dziembowski, A. (2013). RNA decay machines: the exosome. Biochim Biophys Acta *1829*, 552-560.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368-373.

Cigudosa, J.C., Parsa, N.Z., Louie, D.C., Filippa, D.A., Jhanwar, S.C., Johansson, B., Mitelman, F., and Chaganti, R.S. (1999). Cytogenetic analysis of 363 consecutively ascertained diffuse large B-cell lymphomas. Genes Chromosomes Cancer *25*, 123-133.

Conrad, N.K., Wilson, S.M., Steinmetz, E.J., Patturajan, M., Brow, D.A., Swanson, M.S., and Corden, J.L. (2000). A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. Genetics *154*, 557-571.

Constantin, N., Dzantiev, L., Kadyrov, F.A., and Modrich, P. (2005). Human mismatch repair: reconstitution of a nick-directed bidirectional reaction. J Biol Chem *280*, 39752-39761.

Cooper, S.J., Trinklein, N.D., Anton, E.D., Nguyen, L., and Myers, R.M. (2006). Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. Genome Res *16*, 1-10.

Core, L.J., Waterfall, J.J., Gilchrist, D.A., Fargo, D.C., Kwak, H., Adelman, K., and Lis, J.T. (2012). Defining the status of RNA polymerase at promoters. Cell Rep *2*, 1025-1035.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845-1848.

Creamer, T.J., Darby, M.M., Jamonnak, N., Schaughency, P., Hao, H., Wheelan, S.J., and Corden, J.L. (2011). Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. PLoS Genet *7*, e1002329.

Cristodero, M., Bottcher, B., Diepholz, M., Scheffzek, K., and Clayton, C. (2008). The Leishmania tarentolae exosome: purification and structural analysis by electron microscopy. Mol Biochem Parasitol *159*, 24-29.

Dalla-Favera, R., Bregni, M., Erikson, J., Patterson, D., Gallo, R.C., and Croce, C.M. (1982). Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. Proc Natl Acad Sci U S A *79*, 7824-7827.

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.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol *8*, e1000384.

De Silva, N.S., Simonetti, G., Heise, N., and Klein, U. (2012). The diverse roles of IRF4 in late germinal center B-cell differentiation. Immunol Rev *247*, 73-92.

de Yebenes, V.G., Belver, L., Pisano, D.G., Gonzalez, 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.

Depoil, D., Zaru, R., Guiraud, M., Chauveau, A., Harriague, J., Bismuth, G., Utzny, C., Muller, S., and Valitutti, S. (2005). Immunological synapses are versatile structures enabling selective T cell polarization. Immunity *22*, 185-194.

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

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

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

Ding, L., Ley, T.J., Larson, D.E., Miller, C.A., Koboldt, D.C., Welch, J.S., Ritchey, J.K., Young, M.A., Lamprecht, T., McLellan, M.D.*, et al.* (2012). Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature *481*, 506-510.

Dion, M.F., Kaplan, T., Kim, M., Buratowski, S., Friedman, N., and Rando, O.J. (2007). Dynamics of replication-independent histone turnover in budding yeast. Science *315*, 1405-1408.

Doma, M.K., and Parker, R. (2006). Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. Nature *440*, 561-564.

Dominguez-Sanchez, M.S., Barroso, S., Gomez-Gonzalez, B., Luna, R., and Aguilera, A. (2011). Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. PLoS Genet *7*, e1002386.

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.

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.

Duquette, M.L., Handa, P., Vincent, J.A., Taylor, A.F., and Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. Genes Dev *18*, 1618-1629.

Duquette, M.L., Huber, M.D., and Maizels, N. (2007). G-rich proto-oncogenes are targeted for genomic instability in B-cell lymphomas. Cancer Res *67*, 2586-2594.

Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M.E., and Nudler, E. (2011). Linking RNA polymerase backtracking to genome instability in E. coli. Cell *146*, 533-543.

Dziembowski, A., Lorentzen, E., Conti, E., and Seraphin, B. (2007). A single subunit, Dis3, is essentially responsible for yeast exosome core activity. Nat Struct Mol Biol *14*, 15-22.

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

Economides, A.N., Frendewey, D., Yang, P., Dominguez, M.G., Dore, A.T., Lobov, I.B., Persaud, T., Rojas, J., McClain, J., Lengyel, P*., et al.* (2013). Conditionals by inversion provide a universal method for the generation of conditional alleles. Proc Natl Acad Sci U S A *110*, E3179-3188.

Egloff, S., Dienstbier, M., and Murphy, S. (2012). Updating the RNA polymerase CTD code: adding gene-specific layers. Trends Genet *28*, 333-341.

El Hage, A., French, S.L., Beyer, A.L., and Tollervey, D. (2010). Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. Genes Dev *24*, 1546-1558.

Floer, M., Wang, X., Prabhu, V., Berrozpe, G., Narayan, S., Spagna, D., Alvarez, D., Kendall, J., Krasnitz, A., Stepansky, A*., et al.* (2010). A RSC/nucleosome complex determines chromatin architecture and facilitates activator binding. Cell *141*, 407-418.

Flynn, R.A., Almada, A.E., Zamudio, J.R., and Sharp, P.A. (2011). Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. Proc Natl Acad Sci U S A *108*, 10460-10465.

Frendewey, D., Chernomorsky, R., Esau, L., Om, J., Xue, Y., Murphy, A.J., Yancopoulos, G.D., and Valenzuela, D.M. (2010). The loss-of-allele assay for ES cell screening and mouse genotyping. Methods Enzymol *476*, 295-307.

Frey, S., Bertocci, B., Delbos, F., Quint, L., Weill, J.C., and Reynaud, C.A. (1998). Mismatch repair deficiency interferes with the accumulation of mutations in chronically stimulated B cells and not with the hypermutation process. Immunity *9*, 127-134.

Frischmeyer, P.A., van Hoof, A., O'Donnell, K., Guerrerio, A.L., Parker, R., and Dietz, H.C. (2002). An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. Science *295*, 2258-2261.

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

Gan, W., Guan, Z., Liu, J., Gui, T., Shen, K., Manley, J.L., and Li, X. (2011). R-loop-mediated genomic instability is caused by impairment of replication fork progression. Genes Dev *25*, 2041-2056.

Gartenberg, M.R., and Wang, J.C. (1992). Positive supercoiling of DNA greatly diminishes mRNA synthesis in yeast. Proc Natl Acad Sci U S A *89*, 11461-11465.

Gazumyan, A., Bothmer, A., Klein, I.A., Nussenzweig, M.C., and McBride, K.M. (2012). Activation-induced cytidine deaminase in antibody diversification and chromosome translocation. Adv Cancer Res *113*, 167-190.

Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chedin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell *45*, 814-825.

Gitlin, A.D., Shulman, Z., and Nussenzweig, M.C. (2014). Clonal selection in the germinal centre by regulated proliferation and hypermutation. Nature *509*, 637-640.

Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. Science *292*, 1876-1882.

Gomez-Gonzalez, B., and Aguilera, A. (2007). Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. Proc Natl Acad Sci U S A *104*, 8409-8414.

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.

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

Gotea, V., Petrykowska, H.M., and Elnitski, L. (2013). Bidirectional promoters as important drivers for the emergence of species-specific transcripts. PLoS One *8*, e57323.

Green, P., Ewing, B., Miller, W., Thomas, P.J., and Green, E.D. (2003). Transcription-associated mutational asymmetry in mammalian evolution. Nat Genet *33*, 514-517.

Greimann, J.C., and Lima, C.D. (2008). Reconstitution of RNA exosomes from human and Saccharomyces cerevisiae cloning, expression, purification, and activity assays. Methods Enzymol *448*, 185-210.

Grzechnik, P., and Kufel, J. (2008). Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast. Mol Cell *32*, 247-258.

Gudipati, R.K., Villa, T., Boulay, J., and Libri, D. (2008). Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. Nat Struct Mol Biol *15*, 786-794.

Gudipati, R.K., Xu, Z., Lebreton, A., Seraphin, B., Steinmetz, L.M., Jacquier, A., and Libri, D. (2012). Extensive degradation of RNA precursors by the exosome in wild-type cells. Mol Cell *48*, 409-421.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.

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

Guo, C., Gerasimova, T., Hao, H., Ivanova, I., Chakraborty, T., Selimyan, R., Oltz, E.M., and Sen, R. (2011a). Two forms of loops generate the chromatin conformation of the immunoglobulin heavy-chain gene locus. Cell *147*, 332-343.

Guo, C., Yoon, H.S., Franklin, A., Jain, S., Ebert, A., Cheng, H.L., Hansen, E., Despo, O., Bossen, C., Vettermann, C*., et al.* (2011b). CTCF-binding elements mediate control of V(D)J recombination. Nature *477*, 424-430.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P*., et al.* (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223-227.

Hakim, O., Resch, W., Yamane, A., Klein, I., Kieffer-Kwon, K.R., Jankovic, M., Oliveira, T., Bothmer, A., Voss, T.C., Ansarah-Sobrinho, C*., et al.* (2012). DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. Nature *484*, 69-74.

Hasler, J., Rada, C., and Neuberger, M.S. (2011). Cytoplasmic activation-induced cytidine deaminase (AID) exists in stoichiometric complex with translation elongation factor 1alpha (eEF1A). Proc Natl Acad Sci U S A *108*, 18366-18371.

Hauser, A.E., Junt, T., Mempel, T.R., Sneddon, M.W., Kleinstein, S.H., Henrickson, S.E., von Andrian, U.H., Shlomchik, M.J., and Haberman, A.M. (2007). Definition of germinal-center B cell migration in vivo reveals predominant intrazonal circulation patterns. Immunity *26*, 655-667.

Hazelbaker, D.Z., Marquardt, S., Wlotzka, W., and Buratowski, S. (2013). Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination. Mol Cell *49*, 55-66.

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.

Helmrich, A., Ballarino, M., Nudler, E., and Tora, L. (2013). Transcription-replication encounters, consequences and genomic instability. Nat Struct Mol Biol *20*, 412-418.

Helmrich, A., Ballarino, M., and Tora, L. (2011). Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. Mol Cell *44*, 966-977.

Henriques, T., Gilchrist, D.A., Nechaev, S., Bern, M., Muse, G.W., Burkholder, A., Fargo, D.C., and Adelman, K. (2013). Stable Pausing by RNA Polymerase II Provides an Opportunity to Target and Integrate Regulatory Signals. Mol Cell *52*, 517-528.

Hirano, M., Das, S., Guo, P., and Cooper, M.D. (2011). The evolution of adaptive immunity in vertebrates. Adv Immunol *109*, 125-157.

Hobson, D.J., Wei, W., Steinmetz, L.M., and Svejstrup, J.Q. (2012). RNA polymerase II collision interrupts convergent transcription. Mol Cell *48*, 365-374.

Hoey, T., Weinzierl, R.O., Gill, G., Chen, J.L., Dynlacht, B.D., and Tjian, R. (1993). Molecular cloning and functional analysis of Drosophila TAF110 reveal properties expected of coactivators. Cell *72*, 247-260.

Hozumi, N., and Tonegawa, S. (1976). Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. Proc Natl Acad Sci U S A *73*, 3628-3632.

Hsu, C.L., and Stevens, A. (1993). Yeast cells lacking 5'-->3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. Mol Cell Biol *13*, 4826-4835.

Huertas, P., and Aguilera, A. (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol Cell *12*, 711-721.

Huong le, T., Kobayashi, M., Nakata, M., Shioi, G., Miyachi, H., Honjo, T., and Nagaoka, H. (2013). In vivo analysis of Aicda gene regulation: a critical balance between upstream enhancers and intronic silencers governs appropriate expression. PLoS One *8*, e61433.

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

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.

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.

Januszyk, K., and Lima, C.D. (2014). The eukaryotic RNA exosome. Curr Opin Struct Biol *24C*, 132-140.

Januszyk, K., Liu, Q., and Lima, C.D. (2011). Activities of human RRP6 and structure of the human RRP6 catalytic domain. RNA *17*, 1566-1577.

Jensen, T.H., Jacquier, A., and Libri, D. (2013). Dealing with pervasive transcription. Mol Cell *52*, 473-484.

Jolly, C.J., Klix, N., and Neuberger, M.S. (1997). Rapid methods for the analysis of immunoglobulin gene hypermutation: application to transgenic and gene targeted mice. Nucleic Acids Res *25*, 1913-1919.

Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., and Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature *468*, 664-668.

Karolchik, D., Barber, G.P., Casper, J., Clawson, H., Cline, M.S., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M.*, et al.* (2014). The UCSC Genome Browser database: 2014 update. Nucleic Acids Res *42*, D764-770.

Kataoka, T., Kawakami, T., Takahashi, N., and Honjo, T. (1980). Rearrangement of immunoglobulin gamma 1-chain gene and mechanism for heavy-chain class switch. Proc Natl Acad Sci U S A *77*, 919-923.

Keim, C., Kazadi, D., Rothschild, G., and Basu, U. (2013). Regulation of AID, the B-cell genome mutator. Genes Dev *27*, 1-17.

Kenter, A.L. (2012). AID targeting is dependent on RNA polymerase II pausing. Semin Immunol *24*, 281-286.

Khamlichi, A.A., Glaudet, F., Oruc, Z., Denis, V., Le Bert, M., and Cogne, M. (2004). Immunoglobulin class-switch recombination in mice devoid of any S mu tandem repeat. Blood *103*, 3828-3836.

Kim, H.D., Choe, J., and Seo, Y.S. (1999). The sen1(+) gene of Schizosaccharomyces pombe, a homologue of budding yeast SEN1, encodes an RNA and DNA helicase. Biochemistry *38*, 14697-14710.

Kim, K.Y., and Levin, D.E. (2011). Mpk1 MAPK association with the Paf1 complex blocks Sen1-mediated premature transcription termination. Cell *144*, 745-756.

Kim, M., Vasiljeva, L., Rando, O.J., Zhelkovsky, A., Moore, C., and Buratowski, S. (2006). Distinct pathways for snoRNA and mRNA termination. Mol Cell *24*, 723-734.

Kim, N., and Jinks-Robertson, S. (2012). Transcription as a source of genome instability. Nat Rev Genet *13*, 204-214.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S.*, et al.* (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182-187.

King, I.F., Yandava, C.N., Mabb, A.M., Hsiao, J.S., Huang, H.S., Pearson, B.L., Calabrese, J.M., Starmer, J., Parker, J.S., Magnuson, T.*, et al.* (2013). Topoisomerases facilitate transcription of long genes linked to autism. Nature *501*, 58-62.

Klein, B.J., Bose, D., Baker, K.J., Yusoff, Z.M., Zhang, X., and Murakami, K.S. (2011a). RNA polymerase and transcription elongation factor Spt4/5 complex structure. Proc Natl Acad Sci U S A *108*, 546-550.

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.* (2011b). Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. Cell *147*, 95-106.

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. Proc Natl Acad Sci U S A *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. J Exp Med *210*, 1481-1492.

Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A., and Darst, S.A. (2000). A structural model of transcription elongation. Science *289*, 619-625.

Kracker, S., Bergmann, Y., Demuth, I., Frappart, P.O., Hildebrand, G., Christine, R., Wang, Z.Q., Sperling, K., Digweed, M., and Radbruch, A. (2005). Nibrin functions in Ig class-switch recombination. Proc Natl Acad Sci U S A *102*, 1584-1589.

Kwak, H., and Lis, J.T. (2013). Control of transcriptional elongation. Annu Rev Genet *47*, 483-508.

LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., and Tollervey, D. (2005). RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. Cell *121*, 713-724.

Lam, E.Y., Beraldi, D., Tannahill, D., and Balasubramanian, S. (2013). G-quadruplex structures are stable and detectable in human genomic DNA. Nat Commun *4*, 1796.

Lam, K.P., Kuhn, R., and Rajewsky, K. (1997). In vivo ablation of surface immunoglobulin on mature B cells by inducible gene targeting results in rapid cell death. Cell *90*, 1073-1083.

Lam, M.T., Li, W., Rosenfeld, M.G., and Glass, C.K. (2014). Enhancer RNAs and regulated transcriptional programs. Trends Biochem Sci *39*, 170-182.

Langerak, P., Nygren, A.O., Krijger, P.H., van den Berk, P.C., and Jacobs, H. (2007). A/T mutagenesis in hypermutated immunoglobulin genes strongly depends on PCNAK164 modification. J Exp Med *204*, 1989-1998.

Larson, M.H., Mooney, R.A., Peters, J.M., Windgassen, T., Nayak, D., Gross, C.A., Block, S.M., Greenleaf, W.J., Landick, R., and Weissman, J.S. (2014). A pause sequence enriched at translation start sites drives transcription dynamics in vivo. Science *344*, 1042-1047.

Lauberth, S.M., Nakayama, T., Wu, X., Ferris, A.L., Tang, Z., Hughes, S.H., and Roeder, R.G. (2013). H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. Cell *152*, 1021-1036.

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.

Lebreton, A., Tomecki, R., Dziembowski, A., and Seraphin, B. (2008). Endonucleolytic RNA cleavage by a eukaryotic exosome. Nature *456*, 993-996.

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, J.H., and Paull, T.T. (2005). ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. Science *308*, 551-554.

Lemay, J.F., Larochelle, M., Marguerat, S., Atkinson, S., Bahler, J., and Bachand, F. (2014). The RNA exosome promotes transcription termination of backtracked RNA polymerase II. Nat Struct Mol Biol *21*, 919-926.

Lennon, G.G., and Perry, R.P. (1985). C mu-containing transcripts initiate heterogeneously within the IgH enhancer region and contain a novel 5'-nontranslatable exon. Nature *318*, 475-478.

Li, G., White, C.A., Lam, T., Pone, E.J., Tran, D.C., Hayama, K.L., Zan, H., Xu, Z., and Casali, P. (2013). Combinatorial H3K9acS10ph histone modification in IgH locus S regions targets 14-3-3 adaptors and AID to specify antibody class-switch DNA recombination. Cell Rep *5*, 702-714.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

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, X., and Manley, J.L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell *122*, 365-378.

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

Liu, J.J., Bratkowski, M.A., Liu, X., Niu, C.Y., Ke, A., and Wang, H.W. (2014). Visualization of distinct substrate-recruitment pathways in the yeast exosome by EM. Nat Struct Mol Biol *21*, 95-102.

Liu, L.F., and Wang, J.C. (1987). Supercoiling of the DNA template during transcription. Proc Natl Acad Sci U S A *84*, 7024-7027.

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.

Liu, Q., Greimann, J.C., and Lima, C.D. (2006). Reconstitution, activities, and structure of the eukaryotic RNA exosome. Cell *127*, 1223-1237.

Lohr, J.G., Stojanov, P., Lawrence, M.S., Auclair, D., Chapuy, B., Sougnez, C., Cruz-Gordillo, P., Knoechel, B., Asmann, Y.W., Slager, S.L.*, et al.* (2012). Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. Proc Natl Acad Sci U S A *109*, 3879-3884.

Lorentzen, E., Basquin, J., Tomecki, R., Dziembowski, A., and Conti, E. (2008). Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family. Mol Cell *29*, 717-728.

Lorentzen, E., and Conti, E. (2005). Structural basis of 3' end RNA recognition and exoribonucleolytic cleavage by an exosome RNase PH core. Mol Cell *20*, 473-481.

Lorentzen, E., Walter, P., Fribourg, S., Evguenieva-Hackenberg, E., Klug, G., and Conti, E. (2005). The archaeal exosome core is a hexameric ring structure with three catalytic subunits. Nat Struct Mol Biol *12*, 575-581.

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

Lubas, M., Christensen, M.S., Kristiansen, M.S., Domanski, M., Falkenby, L.G., Lykke-Andersen, S., Andersen, J.S., Dziembowski, A., and Jensen, T.H. (2011). Interaction profiling identifies the human nuclear exosome targeting complex. Mol Cell *43*, 624-637.

Lubas, M., Damgaard, C.K., Tomecki, R., Cysewski, D., Jensen, T.H., and Dziembowski, A. (2013). Exonuclease hDIS3L2 specifies an exosome-independent 3'-5' degradation pathway of human cytoplasmic mRNA. EMBO J *32*, 1855-1868.

Lumsden, J.M., McCarty, T., Petiniot, L.K., Shen, R., Barlow, C., Wynn, T.A., Morse, H.C., 3rd, 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.

Lykke-Andersen, S., Brodersen, D.E., and Jensen, T.H. (2009). Origins and activities of the eukaryotic exosome. J Cell Sci *122*, 1487-1494.

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

Maki, R., Traunecker, A., Sakano, H., Roeder, W., and Tonegawa, S. (1980). Exon shuffling generates an immunoglobulin heavy chain gene. Proc Natl Acad Sci U S A *77*, 2138-2142.

Makino, D.L., Baumgartner, M., and Conti, E. (2013). Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex. Nature *495*, 70-75.

Malecki, M., Viegas, S.C., Carneiro, T., Golik, P., Dressaire, C., Ferreira, M.G., and Arraiano, C.M. (2013). The exoribonuclease Dis3L2 defines a novel eukaryotic RNA degradation pathway. EMBO J *32*, 1842-1854.

Manis, J.P., Gu, Y., Lansford, R., Sonoda, E., Ferrini, R., Davidson, L., Rajewsky, K., and Alt, F.W. (1998). Ku70 is required for late B cell development and immunoglobulin heavy chain class switching. J Exp Med *187*, 2081-2089.

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

Marquardt, S., Escalante-Chong, R., Pho, N., Wang, J., Churchman, L.S., Springer, M., and Buratowski, S. (2014). A chromatin-based mechanism for limiting divergent noncoding transcription. Cell *157*, 1712-1723.

Martinez-Rucobo, F.W., Sainsbury, S., Cheung, A.C., and Cramer, P. (2011). Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. EMBO J *30*, 1302-1310.

Masani, S., Han, L., and Yu, K. (2013). Apurinic/apyrimidinic endonuclease 1 is the essential nuclease during immunoglobulin class switch recombination. Mol Cell Biol *33*, 1468-1473.

Masuda, K., Ouchida, R., Hikida, M., Kurosaki, T., Yokoi, M., Masutani, C., Seki, M., Wood, R.D., Hanaoka, F., and J, O.W. (2007). DNA polymerases eta and theta function in the same genetic pathway to generate mutations at A/T during somatic hypermutation of Ig genes. J Biol Chem *282*, 17387-17394.

Matsuda, T., Bebenek, K., Masutani, C., Rogozin, I.B., Hanaoka, F., and Kunkel, T.A. (2001). Error rate and specificity of human and murine DNA polymerase eta. J Mol Biol *312*, 335-346.

Matsunaga, T., Park, C.H., Bessho, T., Mu, D., and Sancar, A. (1996). Replication protein A confers structure-specific endonuclease activities to the XPF-ERCC1 and XPG subunits of human DNA repair excision nuclease. J Biol Chem *271*, 11047-11050.

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., 3rd*, et al.* (2011). Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions. Nat Immunol *12*, 70-76.

Max, E.E., Seidman, J.G., and Leder, P. (1979). Sequences of five potential recombination sites encoded close to an immunoglobulin kappa constant region gene. Proc Natl Acad Sci U S A *76*, 3450-3454.

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.

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.

Mischo, H.E., Gomez-Gonzalez, B., Grzechnik, P., Rondon, A.G., Wei, W., Steinmetz, L., Aguilera, A., and Proudfoot, N.J. (2011). Yeast Sen1 helicase protects the genome from transcription-associated instability. Mol Cell *41*, 21-32.

Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., and Tollervey, D. (1997). The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases. Cell *91*, 457-466.

Mitchell, P., Petfalski, E., and Tollervey, D. (1996). The 3' end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism. Genes Dev *10*, 502-513.

Mitchell, P., and Tollervey, D. (2003). An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3'-->5' degradation. Mol Cell *11*, 1405-1413.

Moreira, M.C., Klur, S., Watanabe, M., Nemeth, A.H., Le Ber, I., Moniz, J.C., Tranchant, C., Aubourg, P., Tazir, M., Schols, L*., et al.* (2004). Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. Nat Genet *36*, 225-227.

Morgan, H.D., Dean, W., Coker, H.A., Reik, W., and Petersen-Mahrt, S.K. (2004). Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. J Biol Chem *279*, 52353-52360.

Mousavi, K., Zare, H., Dell'orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol Cell *51*, 606-617.

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.

Mussmann, R., Courtet, M., Schwager, J., and Du Pasquier, L. (1997). Microsites for immunoglobulin switch recombination breakpoints from Xenopus to mammals. Eur J Immunol *27*, 2610-2619.

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.

Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature *457*, 1038-1042.

Nelson, J.R., Lawrence, C.W., and Hinkle, D.C. (1996). Deoxycytidyl transferase activity of yeast REV1 protein. Nature *382*, 729-731.

Ntini, E., Jarvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jorgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P.R.*, et al.* (2013). Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. Nat Struct Mol Biol *20*, 923-928.

Nudler, E., Mustaev, A., Lukhtanov, E., and Goldfarb, A. (1997). The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. Cell *89*, 33-41.

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

Oakley, G.G., and Patrick, S.M. (2010). Replication protein A: directing traffic at the intersection of replication and repair. Front Biosci (Landmark Ed) *15*, 883-900.

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

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 el, 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.

Ovcharenko, I., Nobrega, M.A., Loots, G.G., and Stubbs, L. (2004). ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. Nucleic Acids Res *32*, W280-286.

Ozsolak, F., Song, J.S., Liu, X.S., and Fisher, D.E. (2007). High-throughput mapping of the chromatin structure of human promoters. Nat Biotechnol *25*, 244-248.

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.

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

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

Pasqualucci, L., Trifonov, V., Fabbri, G., Ma, J., Rossi, D., Chiarenza, A., Wells, V.A., Grunn, A., Messina, M., Elliot, O.*, et al.* (2011). Analysis of the coding genome of diffuse large B-cell lymphoma. Nat Genet *43*, 830-837.

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

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.

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Chao, J., Rabadan, R., Economides, A.N., and Basu, U. (2014). Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. Nature *514*, 389-393.

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

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

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

Phan, R.T., and Dalla-Favera, R. (2004). The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells. Nature *432*, 635-639.

Piruat, J.I., and Aguilera, A. (1998). A novel yeast gene, THO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination. EMBO J *17*, 4859-4872.

Plank, J.L., and Dean, A. (2014). Enhancer Function: Mechanistic and Genome-Wide Insights Come Together. Mol Cell *55*, 5-14.

Potter, M., and Wiener, F. (1992). Plasmacytomagenesis in mice: model of neoplastic development dependent upon chromosomal translocations. Carcinogenesis *13*, 1681-1697.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science *322*, 1851-1854.

Proudfoot, N.J. (2011). Ending the message: poly(A) signals then and now. Genes Dev *25*, 1770-1782.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Rada, C., Di Noia, J.M., and Neuberger, M.S. (2004). Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. Mol Cell *16*, 163-171.

Rada, C., Ehrenstein, M.R., Neuberger, M.S., and Milstein, C. (1998). Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. Immunity *9*, 135-141.

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.

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

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

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

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

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

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y.*, et al.* (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell *154*, 1380-1389.

Ranuncolo, S.M., Polo, J.M., Dierov, J., Singer, M., Kuo, T., Greally, J., Green, R., Carroll, M., and Melnick, A. (2007). Bcl-6 mediates the germinal center B cell phenotype and lymphomagenesis through transcriptional repression of the DNA-damage sensor ATR. Nat Immunol *8*, 705-714.

Rappold, I., Iwabuchi, K., Date, T., and Chen, J. (2001). Tumor suppressor p53 binding protein 1 (53BP1) is involved in DNA damage-signaling pathways. J Cell Biol *153*, 613-620.

Reaban, M.E., and Griffin, J.A. (1990). Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. Nature *348*, 342-344.

Reichert, J.M. (2012). Marketed therapeutic antibodies compendium. MAbs *4*, 413-415.

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

Reina-San-Martin, B., Chen, J., Nussenzweig, A., and Nussenzweig, M.C. (2007). Enhanced intra-switch region recombination during immunoglobulin class switch recombination in 53BP1-/- B cells. Eur J Immunol *37*, 235-239.

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.

Reina-San-Martin, B., Nussenzweig, M.C., Nussenzweig, A., and Difilippantonio, S. (2005). Genomic instability, endoreduplication, and diminished Ig class-switch recombination in B cells lacking Nbs1. Proc Natl Acad Sci U S A *102*, 1590-1595.

Renner, D.B., Yamaguchi, Y., Wada, T., Handa, H., and Price, D.H. (2001). A highly purified RNA polymerase II elongation control system. J Biol Chem *276*, 42601-42609.

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.

Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature *483*, 295-301.

Richard, P., Feng, S., and Manley, J.L. (2013). A SUMO-dependent interaction between Senataxin and the exosome, disrupted in the neurodegenerative disease AOA2, targets the exosome to sites of transcription-induced DNA damage. Genes Dev *27*, 2227-2232.

Rickert, R.C., Roes, J., and Rajewsky, K. (1997). B lymphocyte-specific, Cre-mediated mutagenesis in mice. Nucleic Acids Res *25*, 1317-1318.

Ricklin, D., Hajishengallis, G., Yang, K., and Lambris, J.D. (2010). Complement: a key system for immune surveillance and homeostasis. Nat Immunol *11*, 785-797.

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

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

Roberts, R.W., and Crothers, D.M. (1992). Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. Science *258*, 1463-1466.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol *29*, 24-26.

Rodriguez, C.I., Buchholz, F., Galloway, J., Sequerra, R., Kasper, J., Ayala, R., Stewart, A.F., and Dymecki, S.M. (2000). High-efficiency deleter mice show that FLPe is an alternative to Cre-loxP. Nat Genet *25*, 139-140.

Rondon, A.G., Mischo, H.E., Kawauchi, J., and Proudfoot, N.J. (2009). Fail-safe transcriptional termination for protein-coding genes in S. cerevisiae. Mol Cell *36*, 88-98.

Roy, D., Zhang, Z., Lu, Z., Hsieh, C.L., and Lieber, M.R. (2010). Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. Mol Cell Biol *30*, 146-159.

Rufiange, A., Jacques, P.E., Bhat, W., Robert, F., and Nourani, A. (2007). Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1. Mol Cell *27*, 393-405.

Ruiz, J.F., Gomez-Gonzalez, B., and Aguilera, A. (2011). AID induces double-strand breaks at immunoglobulin switch regions and c-MYC causing chromosomal translocations in yeast THO mutants. PLoS Genet *7*, e1002009.

Rush, J.S., Fugmann, S.D., and Schatz, D.G. (2004). Staggered AID-dependent DNA double strand breaks are the predominant DNA lesions targeted to S mu in Ig class switch recombination. Int Immunol *16*, 549-557.

Saito, M., Novak, U., Piovan, E., Basso, K., Sumazin, P., Schneider, C., Crespo, M., Shen, Q., Bhagat, G., Califano, A.*, et al.* (2009). BCL6 suppression of BCL2 via Miz1 and its disruption in diffuse large B cell lymphoma. Proc Natl Acad Sci U S A *106*, 11294-11299.

Sakano, H., Huppi, K., Heinrich, G., and Tonegawa, S. (1979). Sequences at the somatic recombination sites of immunoglobulin light-chain genes. Nature *280*, 288-294.

Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., and Hume, D.A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nat Rev Genet *8*, 424-436.

Schaeffer, D., Clark, A., Klauer, A.A., Tsanova, B., and van Hoof, A. (2010). Functions of the cytoplasmic exosome. Adv Exp Med Biol *702*, 79-90.

Schaeffer, D., Tsanova, B., Barbas, A., Reis, F.P., Dastidar, E.G., Sanchez-Rotunno, M., Arraiano, C.M., and van Hoof, A. (2009). The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. Nat Struct Mol Biol *16*, 56-62.

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

Schneider, C., Kudla, G., Wlotzka, W., Tuck, A., and Tollervey, D. (2012). Transcriptome-wide analysis of exosome targets. Mol Cell *48*, 422-433.

Schnutgen, F., Doerflinger, N., Calleja, C., Wendling, O., Chambon, P., and Ghyselinck, N.B. (2003). A directional strategy for monitoring Cre-mediated recombination at the cellular level in the mouse. Nat Biotechnol *21*, 562-565.

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

Schreck, S., Buettner, M., Kremmer, E., Bogdan, M., Herbst, H., and Niedobitek, G. (2006). Activation-induced cytidine deaminase (AID) is expressed in normal spermatogenesis but only infrequently in testicular germ cell tumours. J Pathol *210*, 26-31.

Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., and Cramer, P. (2013). Transcriptome surveillance by selective termination of noncoding RNA synthesis. Cell *155*, 1075-1087.

Schwickert, T.A., Lindquist, R.L., Shakhar, G., Livshits, G., Skokos, D., Kosco-Vilbois, M.H., Dustin, M.L., and Nussenzweig, M.C. (2007). In vivo imaging of germinal centres reveals a dynamic open structure. Nature *446*, 83-87.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849-1851.

Shaffer, A.L., Yu, X., He, Y., Boldrick, J., Chan, E.P., and Staudt, L.M. (2000). BCL-6 represses genes that function in lymphocyte differentiation, inflammation, and cell cycle control. Immunity *13*, 199-212.

Shcherbo, D., Merzlyak, E.M., Chepurnykh, T.V., Fradkov, A.F., Ermakova, G.V., Solovieva, E.A., Lukyanov, K.A., Bogdanova, E.A., Zaraisky, A.G., Lukyanov, S.*, et al.* (2007). Bright far-red fluorescent protein for whole-body imaging. Nat Methods *4*, 741-746.

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

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V.*, et al.* (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116-120.

Shimizu, A., Takahashi, N., Yaoita, Y., and Honjo, T. (1982). Organization of the constant-region gene family of the mouse immunoglobulin heavy chain. Cell *28*, 499-506.

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.

Siegel, R.W., Velappan, N., Pavlik, P., Chasteen, L., and Bradbury, A. (2004). Recombinatorial cloning using heterologous lox sites. Genome Res *14*, 1119-1129.

Sigova, A.A., Mullen, A.C., Molinie, B., Gupta, S., Orlando, D.A., Guenther, M.G., Almada, A.E., Lin, C., Sharp, P.A., Giallourakis, C.C., *et al.* (2013). Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. Proc Natl Acad Sci U S A *110*, 2876-2881.

Skourti-Stathaki, K., and Proudfoot, N.J. (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes Dev *28*, 1384-1396.

Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell *42*, 794-805.

Spits, H., and Cupedo, T. (2012). Innate lymphoid cells: emerging insights in development, lineage relationships, and function. Annu Rev Immunol *30*, 647-675.

Staals, R.H., Bronkhorst, A.W., Schilders, G., Slomovic, S., Schuster, G., Heck, A.J., Raijmakers, R., and Pruijn, G.J. (2010). Dis3-like 1: a novel exoribonuclease associated with the human exosome. EMBO J *29*, 2358-2367.

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

Stavnezer-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.

Steinmetz, E.J., and Brow, D.A. (1996). Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. Mol Cell Biol *16*, 6993-7003.

Steinmetz, E.J., Conrad, N.K., Brow, D.A., and Corden, J.L. (2001). RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. Nature *413*, 327-331.

Storb, U. (2014). Why Does Somatic Hypermutation by AID Require Transcription of Its Target Genes? Adv Immunol *122*, 253-277.

Sun, J., Keim, C.D., Wang, J., Kazadi, D., Oliver, P.M., Rabadan, R., and Basu, U. (2013a). E3-ubiquitin ligase Nedd4 determines the fate of AID-associated RNA polymerase II in B cells. Genes Dev *27*, 1821-1833.

Sun, J., Rothschild, G., Pefanis, E., and Basu, U. (2013b). Transcriptional stalling in B-lymphocytes: A mechanism for antibody diversification and maintenance of genomic integrity. Transcription *4*, 127-135.

Suraweera, A., Lim, Y., Woods, R., Birrell, G.W., Nasim, T., Becherel, O.J., and Lavin, M.F. (2009). Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. Hum Mol Genet *18*, 3384-3396.

Symmons, M.F., Jones, G.H., and Luisi, B.F. (2000). A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation. Structure *8*, 1215-1226.

Szymczak, A.L., Workman, C.J., Wang, Y., Vignali, K.M., Dilioglou, S., Vanin, E.F., and Vignali, D.A. (2004). Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. Nat Biotechnol *22*, 589-594.

Tafforeau, L., Zorbas, C., Langhendries, J.L., Mullineux, S.T., Stamatopoulou, V., Mullier, R., Wacheul, L., and Lafontaine, D.L. (2013). The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of Pre-rRNA processing factors. Mol Cell *51*, 539-551.

Taft, R.J., Glazov, E.A., Cloonan, N., Simons, C., Stephen, S., Faulkner, G.J., Lassmann, T., Forrest, A.R., Grimmond, S.M., Schroder, K.*, et al.* (2009). Tiny RNAs associated with transcription start sites in animals. Nat Genet *41*, 572-578.

Takahashi, S., Araki, Y., Sakuno, T., and Katada, T. (2003). Interaction between Ski7p and Upf1p is required for nonsense-mediated 3'-to-5' mRNA decay in yeast. EMBO J *22*, 3951-3959.

Tan-Wong, S.M., Zaugg, J.B., Camblong, J., Xu, Z., Zhang, D.W., Mischo, H.E., Ansari, A.Z., Luscombe, N.M., Steinmetz, L.M., and Proudfoot, N.J. (2012). Gene loops enhance transcriptional directionality. Science *338*, 671-675.

Taub, R., Kirsch, I., Morton, C., Lenoir, G., Swan, D., Tronick, S., Aaronson, S., and Leder, P. (1982). Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells. Proc Natl Acad Sci U S A *79*, 7837-7841.

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.

Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., and Libri, D. (2006). Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. Mol Cell *23*, 853-864.

Tomecki, R., Kristiansen, M.S., Lykke-Andersen, S., Chlebowski, A., Larsen, K.M., Szczesny, R.J., Drazkowska, K., Pastula, A., Andersen, J.S., Stepien, P.P.*, et al.* (2010). The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. EMBO J *29*, 2342-2357.

Torchet, C., Bousquet-Antonelli, C., Milligan, L., Thompson, E., Kufel, J., and Tollervey, D. (2002). Processing of 3'-extended read-through transcripts by the exosome can generate functional mRNAs. Mol Cell *9*, 1285-1296.

Tran, T.H., Nakata, M., Suzuki, K., Begum, N.A., Shinkura, R., Fagarasan, S., Honjo, T., and Nagaoka, H. (2010). B cell-specific and stimulation-responsive enhancers derepress Aicda by overcoming the effects of silencers. Nat Immunol *11*, 148-154.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511-515.

Tuduri, S., Crabbe, L., Conti, C., Tourriere, H., Holtgreve-Grez, H., Jauch, A., Pantesco, V., De Vos, J., Thomas, A., Theillet, C.*, et al.* (2009). Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. Nat Cell Biol *11*, 1315-1324.

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-gamma associates with and modulates the abundance of nuclear activation-induced deaminase. J Exp Med *208*, 2385-2391.

Valenzuela, D.M., Murphy, A.J., Frendewey, D., Gale, N.W., Economides, A.N., Auerbach, W., Poueymirou, W.T., Adams, N.C., Rojas, J., Yasenchak, J.*, et al.* (2003). High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. Nat Biotechnol *21*, 652-659.

van Hoof, A., Frischmeyer, P.A., Dietz, H.C., and Parker, R. (2002). Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. Science *295*, 2262-2264.

van Hoof, A., Lennertz, P., and Parker, R. (2000). Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. Mol Cell Biol *20*, 441-452.

Vasiljeva, L., and Buratowski, S. (2006). Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. Mol Cell *21*, 239-248.

Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., and Meinhart, A. (2008). The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. Nat Struct Mol Biol *15*, 795-804.

Victora, G.D., and Nussenzweig, M.C. (2012). Germinal centers. Annu Rev Immunol *30*, 429-457.

Victora, G.D., Schwickert, T.A., Fooksman, D.R., Kamphorst, A.O., Meyer-Hermann, M., Dustin, M.L., and Nussenzweig, M.C. (2010). Germinal center dynamics revealed by multiphoton microscopy with a photoactivatable fluorescent reporter. Cell *143*, 592-605.

Vivier, E., Tomasello, E., Baratin, M., Walzer, T., and Ugolini, S. (2008). Functions of natural killer cells. Nat Immunol *9*, 503-510.

Vuong, B.Q., Herrick-Reynolds, K., Vaidyanathan, B., Pucella, J.N., Ucher, A.J., Donghia, N.M., Gu, X., Nicolas, L., Nowak, U., Rahman, N.*, et al.* (2013). A DNA break- and phosphorylation-dependent positive feedback loop promotes immunoglobulin class-switch recombination. Nat Immunol *14*, 1183-1189.

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.

Vvedenskaya, I.O., Vahedian-Movahed, H., Bird, J.G., Knoblauch, J.G., Goldman, S.R., Zhang, Y., Ebright, R.H., and Nickels, B.E. (2014). Transcription. Interactions between RNA polymerase and the "core recognition element" counteract pausing. Science *344*, 1285-1289.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F.*, et al.* (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes Dev *12*, 343-356.

Wahba, L., Amon, J.D., Koshland, D., and Vuica-Ross, M. (2011). RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. Mol Cell *44*, 978-988.

Walker, J.R., Corpina, R.A., and Goldberg, J. (2001). Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. Nature *412*, 607-614.

Wan, J., Yourshaw, M., Mamsa, H., Rudnik-Schoneborn, S., Menezes, M.P., Hong, J.E., Leong, D.W., Senderek, J., Salman, M.S., Chitayat, D.*, et al.* (2012). Mutations in the RNA exosome component gene EXOSC3 cause pontocerebellar hypoplasia and spinal motor neuron degeneration. Nat Genet *44*, 704-708.

Wang, D., and Landick, R. (1997). Nuclease cleavage of the upstream half of the nontemplate strand DNA in an Escherichia coli transcription elongation complex causes upstream translocation and transcriptional arrest. J Biol Chem *272*, 5989-5994.

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

Wang, L., Wuerffel, R., Feldman, S., Khamlichi, A.A., and Kenter, A.L. (2009b). S region sequence, RNA polymerase II, and histone modifications create chromatin accessibility during class switch recombination. J Exp Med *206*, 1817-1830.

Wang, X., Fan, M., Kalis, S., Wei, L., and Scharff, M.D. (2014). A source of the single-stranded DNA substrate for activation-induced deaminase during somatic hypermutation. Nat Commun *5*, 4137.

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

Wasmuth, E.V., Januszyk, K., and Lima, C.D. (2014). Structure of an Rrp6-RNA exosome complex bound to poly(A) RNA. Nature *511*, 435-439.

Wasmuth, E.V., and Lima, C.D. (2012). Exo- and endoribonucleolytic activities of yeast cytoplasmic and nuclear RNA exosomes are dependent on the noncatalytic core and central channel. Mol Cell *48*, 133-144.

Wellinger, R.E., Prado, F., and Aguilera, A. (2006). Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. Mol Cell Biol *26*, 3327-3334.

Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004). Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. Science *303*, 1014-1016.

Whitehouse, I., Rando, O.J., Delrow, J., and Tsukiyama, T. (2007). Chromatin remodelling at promoters suppresses antisense transcription. Nature *450*, 1031-1035.

Wiesendanger, M., Kneitz, B., Edelmann, W., and Scharff, M.D. (2000). Somatic hypermutation in MutS homologue (MSH)3-, MSH6-, and MSH3/MSH6-deficient mice reveals a role for the MSH2-MSH6 heterodimer in modulating the base substitution pattern. J Exp Med *191*, 579-584.

Wilson, T.M., Vaisman, A., Martomo, S.A., Sullivan, P., Lan, L., Hanaoka, F., Yasui, A., Woodgate, R., and Gearhart, P.J. (2005). MSH2-MSH6 stimulates DNA polymerase eta, suggesting a role for A:T mutations in antibody genes. J Exp Med *201*, 637-645.

Wlotzka, W., Kudla, G., Granneman, S., and Tollervey, D. (2011). The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. EMBO J *30*, 1790-1803.

Wongsurawat, T., Jenjaroenpun, P., Kwoh, C.K., and Kuznetsov, V. (2012). Quantitative model of R-loop forming structures reveals a novel level of RNA-DNA interactome complexity. Nucleic Acids Res *40*, e16.

Wu, X., and Sharp, P.A. (2013). Divergent transcription: a driving force for new gene origination? Cell *155*, 990-996.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.C., Dufour, M.E., Boulay, J., Regnault, B., Devaux, F., Namane, A., Seraphin, B.*, et al.* (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell *121*, 725-737.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Munster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature *457*, 1033-1037.

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.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell *97*, 41-51.

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.

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

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

Yancopoulos, G.D., DePinho, R.A., Zimmerman, K.A., Lutzker, S.G., Rosenberg, N., and Alt, F.W. (1986). Secondary genomic rearrangement events in pre-B cells: VHDJH replacement by a LINE-1 sequence and directed class switching. EMBO J *5*, 3259-3266.

Yoshihama, M., Nakao, A., and Kenmochi, N. (2013). snOPY: a small nucleolar RNA orthological gene database. BMC Res Notes *6*, 426.

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., Du Pasquier, 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.

Zhang, Y., Buchholz, F., Muyrers, J.P., and Stewart, A.F. (1998). A new logic for DNA engineering using recombination in Escherichia coli. Nat Genet *20*, 123-128.

Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M.X., Arnold, E., and Ebright, R.H. (2012). Structural basis of transcription initiation. Science *338*, 1076-1080.

Zhou, Q., Li, T., and Price, D.H. (2012). RNA polymerase II elongation control. Annu Rev Biochem *81*, 119-143.