

**ELUCIDATING ACTIVATION INDUCED CYTIDINE DEAMINASE  
TARGETING DURING  
IMMUNOGLOBULIN CLASS SWITCH RECOMBINATION**

**By**

**Ahrom Kim**

**A DISSERTATION**

**Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of**

**Microbiology and Molecular Genetics – Doctor of Philosophy**

**2019**

## **ABSTRACT**

### **ELUCIDATING ACTIVATION INDUCED CYTIDINE DEAMINASE TARGETING DURING IMMUNOGLOBULIN CLASS SWITCH RECOMBINATION**

**By**

**Ahrom Kim**

In antigen-stimulated B cells, the immunoglobulin (Ig) heavy chain constant region can be changed through a mechanism called Class Switch Recombination (CSR). This process alters the effector function of the Ig molecule while maintaining its antigen specificity. Defective CSR may result in hyper-IgM syndrome, chromosomal translocations, and autoimmune diseases. Thus, understanding the mechanism of CSR will provide valuable insights into these diseases. This work used a mouse B cell line (CH12F3) to study the mechanism of CSR.

CSR is initiated by a B-cell-specific factor called activation-induced cytidine deaminase (AID), which converts cytosines to uracils in a single stranded DNA at switch (S) regions. In order for AID to access a S region, germline transcription (GLT) through that S region is absolutely required. GLT is driven by a cytokine-dependent promoter upstream of a non-coding exon located upstream of each S region. GLT is also regulated by a large enhancer complex called the 3' regulatory region (3'RR) located at the 3'end of the last C region (C $\alpha$  in mice). One of the functions of 3'RR is to stimulate and enhance GLT. This 28kb region contains four enhancers identified as DNase I Hypersensitive (HS) sites with additional spacer

regions between them. Our results demonstrated that the four HS sites contain all the functional elements necessary for CSR.

Repair of AID-generated uracils in S regions leads to double strand break (DSB) formation. A widely accepted hypothesis in the field is that generation of uracils within close proximity on opposing DNA strands leads to closely generated nicks, resulting in DSBs. Our results demonstrate that AID deaminates evenly across the entire S region and to the same extent between the top and bottom strands. In addition, the AID footprint analyses implicate that DSB formation mostly results from distally spaced nicks.

Several AID mutations that cause hyper-IgM syndrome have been reported to possess active deaminase activity in biochemical assays but are defective for CSR in vivo. Why these mutants cannot support CSR is unknown. We focused on a mutation that results in C-terminal truncation (AID<sup>ΔC</sup>). Our study shows that AID<sup>ΔC</sup> acts as a null allele in the mouse CH12F3 cell line, in contrast to a dominant negative phenotype in human patients. This raised a possibility of species-specific regulation of AID in its stability and/or functionality.

These studies have greatly improved our understanding of AID-targeting and AID activity involving S regions during CSR. Additionally, the cell lines we generated are valuable reagents for future mutagenesis studies of cis-acting DNA elements in 3'RR critical for CSR and AID mutations relevant to human diseases.

## ACKNOWLEDGEMENTS

I first would like to thank my advisor, **Dr. Kefei Yu**, who has fostered and guided me through my doctoral study. He not only gave me the space and resources to pursue my research, but most importantly never hesitated to provide the time and effort to guide me to achieving my goals. He was supportive of everything I thought was necessary for my career.

**Members of the Yu Lab** have been the best people to work with. Li Han and Dr. Shanaz Masani really nurtured me in the Lab. It was my pleasure to work with such an excellent team.

**My committee members**, Dr. Kathy Meek, Dr. Donna Koslowsky, and Dr. Cheryl Rockwell, have been kind and supportive through every step of the way. Dr. Meek's passion for science was always inspiring. Dr. Koslowsky's effort to make time to meet with me not just as a committee member, but also as a graduate director helped me to stay focused. Dr. Rockwell's enthusiasm for my work has always kept me confident.

**Members of the Meek Lab** have been my second home next to Yu lab. Dr. Jessica Neal and Dr. Chris Buehl have been the best officemates one can ever wish for. It was also my joy to see all the Meek Lab undergraduates grow as scientists and move on to their next step in their career.

**Microbiology and molecular genetics department personnel** have helped me hone my skills as a researcher and have made my time here an exciting and fulfilling experience. Dr. Parkin was an amazing mentor and a friend. She is a

true inspiration. Because of her, I found and will continue my passion. I would like to give a huge thanks to Betty Miller and Roseann Bills, who were like the moms of the department. They always made sure I had less thing to worry about.

Last but not least, I would like to thank **my family and friends**. Without any of them, I would not have been able to make it this far. My lovely mom has always been my life support. My true best friend, Rebecca Ra, suffered through many long hours of phone calls. My two beautiful fur babies, Noori and Kami, kept me company through ups and downs, and many late nights of studying and writing. Members of BTOB, Eunkwang Seo, Minhyuk Lee, Changsub Lee, Hyunsik Lee, Peniel Shin, Ilhoon Jung, and Sungjae Yook, gave me times to laugh and smile during toughest times. Many thanks to all my friends who believed in and supported me throughout this process.

## TABLE OF CONTENTS

<b>LIST OF TABLES</b>	viii
<b>LIST OF FIGURES</b>	ix
<b>KEY TO ABBREVIATIONS</b>	x
<b>CHAPTER 1: INTRODUCTION</b>	1
1.1 Overview of Class Switch Recombination	2
1.2 Switch Region Sequences	4
1.3 Activation-induced cytidine deaminase	5
1.4 The Role of Transcription in Class Switch Recombination	8
1.5 Generation and Resolution of DNA Double Stranded Breaks	10
1.6 CH12F3 – A Cell-line to Study Class Switch Recombination	12
1.7 Clinical Significance	13
<b>CHAPTER 2: MUTAGENESIS OF 3'RR TO IDENTIFY DNA SEQUENCE MOTIFS CRITICAL FOR AID-TARGETING TO S REGIONS</b>	15
2.1 Introduction	16
2.2 Materials and Methods	18
2.2.1 Reagents, cell culture, and CSR assay	18
2.2.2 Gene targeting	18
2.2.3 Recombinase-mediated cassette exchange	19
2.2.4 RT-PCR	20
2.3 Results	21
2.3.1 Gene targeting of 3'RR	21
2.3.2 Reduced CSR in 3'RR-deleted CH12F3 cells	22
2.3.3 Hyper-CSR upon knock-in of the four HS sites	23
2.4 Discussion	25
<b>CHAPTER 3: ELUCIDATING AID-MEDIATED DEAMINATION DURING CLASS SWITCH RECOMBINATION</b>	28
3.1 Introduction	29
3.2 Materials and Methods	32
3.2.1 Reagents, cell culture, and Class Switch Recombination assay	32
3.2.2 Gene targeting	32
3.2.3 Recombinase-mediated cassette exchange	33
3.2.4 Uracil-DNA glycosylase assay	33
3.2.5 AID footprint analysis	33
3.3 Results	35

3.3.1 Time-lapsed accumulation of AID deamination	35
3.3.2 AID footprints analysis at a single cell level	37
3.3.3 R190X AID in CH12F3	39
3.4 Discussion	43
<b>CHAPTER 4: SUMMARY AND CONCLUDING REMARKS</b>	<b>48</b>
<b>APPENDICES</b>	<b>52</b>
APPENDIX A: Chapter 1 Figures	53
APPENDIX B: Chapter 2 Figures	58
APPENDIX C: Chapter 3 Figures	72
<b>BIBLIOGRAPHY</b>	<b>90</b>

## LIST OF TABLES

<b>Table 1</b> Oligonucleotides used to amplify homology blocks	19
<b>Table 2</b> Oligonucleotides used in RT-PCR	20

## LIST OF FIGURES

<b>Figure 1.</b> Structure of an Immunoglobulin Molecule	53
<b>Figure 2.</b> Overview of Class Switch Recombination	54
<b>Figure 3.</b> Molecular Mechanism of Class Switch Recombination	55
<b>Figure 4.</b> Germline Transcription during Class Switch Recombination	56
<b>Figure 5.</b> Formation of Double Stranded DNA Break during Class Switch Recombination	57
<b>Figure 6.</b> Gene targeting to replace 3'RR with an RMCE cassette	58
<b>Figure 7.</b> Reduced Class Switch Recombination in 3'RR-deleted (from the productive allele) cells	60
<b>Figure 8.</b> Class Switch Recombination in 3'RR-deficient cells after removal of the RMCE cassette	62
<b>Figure 9.</b> Class Switch Recombination in cells with a deletion of 3'RR on both alleles	65
<b>Figure 10.</b> Knock-in of four DNase I HS sites restores Class Switch Recombination	68
<b>Figure 11.</b> Analysis of AID-footprints	72
<b>Figure 12.</b> Occurrence of AID-generated mutation	73
<b>Figure 13.</b> Analysis of AID footprints in pooled populations of S $\alpha$ <sup>core</sup> -DKO cells after 1 or 4 weeks of CIT stimulation	74
<b>Figure 14.</b> AID-footprints in S $\alpha$ region	82
<b>Figure 15.</b> Genome editing at AID locus	83
<b>Figure 16.</b> Analysis of AID footprints in pooled populations of AID <sup>E5/<math>\Delta</math>E5</sup> S $\alpha$ <sup>core</sup> -DKO cells after 1 week of CIT stimulation	86
<b>Figure 17.</b> Model of DSB formation by distally position AID-generated uracils in a collapsed R-loop structure	89

## KEY TO ABBREVIATIONS

Ig	Immunoglobulin
V	Variable
C	Constant
D	Diversity
J	Joining
SHM	Somatic Hypermutation
CSR	Class Switch Recombination
DSB	Double Stranded Break
S	Switch
AID	Activation-Induced Cytidine Deaminase
ssDNA	Single Stranded DNA
NHEJ	Non Homologous End-Joining
BER	Base Excision Repair
MMR	Mismatch Repair
UNG	Uracil DNA Glycosylase
AP	Apurinic/Apyrimidinic
APE	AP Endonuclease
GLT	Germline Transcription
MSH2	MutS Homolog 2
CIT	anti-CD40, IL4, TGF $\beta$

WT	Wild Type
PCR	Polymerase Chain Reaction
C-NHEJ	Classical Non Homologous-End Joining
A-EJ	Alternative End-Joining
a.a.	Amino Acid
hr	Hours
$\mu$ l	Microliter (s)
$\mu$ g	Microgram (s)
$\mu$ M	Micromolar (s)
nM	Nanomolar (s)
ml	Milliliter (s)
kb	Kilobase (s)

## CHAPTER 1: INTRODUCTION

## 1.1 Overview of Class Switch Recombination

The mammalian adaptive immune response requires B cells to produce immunoglobulins (Ig). Ig molecules are present as cell surface B cell receptors (BCR) and secreted molecules, commonly known as antibodies. Igs, or antibodies, can recognize a vast diversity of antigens on foreign pathogens. They are 'Y' shaped molecules that consist of two heavy (50kDa) and two light chains (25kDa), joined together by disulfide bonds (Figure 1). Ig molecules contain an N-terminal variable (V) region involved in antigen binding and a C-terminal constant (C) region that interacts with effector cells and molecules.

The class, or isotype, of an Ig is determined by its heavy chain C region. Mammals have five different Ig isotypes (IgM, IgD, IgG, IgE, and IgA). Each Ig isotype varies in its half-life as well as effector functions, such as neutralization of pathogens, activation of complement, and recruitment of immune effectors to eradicate infections.

Three distinct somatic alterations occur in the B cell genome to generate diverse repertoires of Ig molecules: V(D)J recombination, somatic hypermutation (SHM) and class switch recombination (CSR). The first two processes are involved in generating diversity in antigen binding of Ig. V(D)J recombination is a site-directed recombination event that assembles the V region of an Ig molecule, whereas SHM introduces point mutations into the V region to enable the production of higher affinity antibodies. CSR is a region-specific event that changes the C region of the Ig molecule while maintaining the same V region, allowing the Ig

molecule to recognize the same pathogen but alter its effector function for optimal pathogen clearance.

The mammalian heavy chain germline locus consists of an array of C region genes that represent each isotype ( $\mu$ ,  $\delta$ ,  $\gamma_3$ ,  $\gamma_1$ ,  $\gamma_{2b}$ ,  $\gamma_{2a}$ ,  $\epsilon$  and  $\alpha$  in mice;  $\mu$ ,  $\delta$ ,  $\gamma_3$ ,  $\gamma_1$ ,  $\alpha_1$ ,  $\gamma_2$ ,  $\gamma_4$ ,  $\epsilon$  and  $\alpha_2$  in humans).  $C_\mu$  is the first constant region downstream of the V region. Therefore, naïve B cells primarily express IgM. In antigen-stimulated B cells, the  $C_\mu$  constant region can be replaced by one of the downstream C regions (either  $C_\gamma$ ,  $C_\epsilon$  or  $C_\alpha$  in mammals) through CSR (Figure 2) [1], [2].

B cells initiate CSR by upregulating a B cell-specific enzyme called Activation-Induced Cytidine Deaminase (AID). AID is expressed only when B cells are activated and converts cytosines to uracils in single stranded DNA (ssDNA). In order for AID to access a S region, germline transcription (GLT) through the region is absolutely required. GLT is driven by a cytokine-dependent promoter upstream of a non-coding I exon located upstream of each S region. GLT is also regulated by a large enhancer complex called the 3' regulatory region (3'RR) located at the 3' end of the last C region ( $C_\alpha$  in mice) [3], [4]. Once AID has access to the S region, AID deaminates a fraction of cytosines in that region into uracils, which are then processed by DNA repair factors that ultimately results in DNA double strand breaks (DSBs). Once DSBs are generated in the donor and acceptor S regions, the two S regions can be joined together through classical non-homologous end joining (c-NHEJ) and alternative end joining (A-EJ) pathways. The intervening DNA is deleted, resulting in the V region being juxtaposed to a new C region (Figure 3). As there are no consensus recombination sites in S

regions, CSR is best known as a region-specific recombination. Through this mechanism, only the C region gets swapped; the Ig maintains its antigen specificity with a different Ig effector function. This process is essential for a robust immune response [1], [2], [5].

The process of CSR has several requirements including S regions, GLT, AID, and DNA repair factors, each of which will be discussed below in further detail.

## 1.2 Switch Region Sequences

Switch (S) regions are intronic DNA sequences located upstream of each constant region exon. S regions are highly repetitive regions ranging in size from 1-12 kb [6]. Mammalian switch regions are G-rich in the non-template strand and consist of several tandem repeats which degenerate at the boundaries of the switch regions [2], [6]–[8]. Certain conserved G-rich pentamer motifs such as TGGGG, GGGGT, GGGCT, and GAGCT frequently occur within S regions [6]. The S regions in mice and humans can be classified into two categories with regarding to repeating motifs: The  $S_{\mu}$ ,  $S_{\alpha}$ , and  $S_{\epsilon}$  are primarily composed of pentamers, while the  $S_{\gamma 3}$ ,  $S_{\gamma 1}$ ,  $S_{\gamma 2b}$ ,  $S_{\gamma 2a}$  contain 49-52 base-pair repeats enriched with the pentamers [1], [6].

Large deletions in the  $S_{\mu}$  region have been shown to severely impair CSR to all isotypes, whereas deletion of the  $S_{\gamma 1}$  region only abolishes CSR to IgG1 [9], [10]. Replacing the native  $S_{\gamma 1}$  with an irrelevant intronic sequence cannot rescue switching [11], while inversion of the  $S_{\gamma 1}$  significantly reduced CSR to IgG1 [12]. In contrast to mammalian S regions, amphibian and reptile S regions are A/T rich

[13], [14]. However, *Xenopus* S regions are rich in palindromic repeats and have been shown to partially rescue Sy1 loss in mice [11], indicating that S regions contain common motifs essential for CSR. Unfortunately, due to the length and heterogeneity of the S region sequences and random distribution of CSR breakpoints, it has been difficult to identify features within switch regions that are important for CSR [1], [2], [15], [16].

### **1.3 Activation-induced cytidine deaminase**

Activation-induced cytidine deaminase (AID) is a B cell-specific enzyme that converts cytosines to uracils on a single stranded DNA (ssDNA). It is essential for three important antibody diversification processes: CSR, SHM, and gene conversion (GC). GC is a major mechanism of antibody diversification in birds where the V region is partially replaced by a series of homologous pseudogenes lying upstream on the same chromosome [17]–[19].

AID was discovered as an upregulated factor in the mouse CH12F3 B cell line upon stimulation for CSR [20]. AID belongs to the apolipoprotein B mRNA-editing catalytic polypeptide (APOBEC) family of cytidine deaminases with diverse function in lipid metabolism, inhibition of retrotransposons and retroviruses, and immune diversification [1], [20], [21]. Due to AID's homology with APOBEC1, the only APOBEC protein known to deaminate RNA substrates, AID was initially hypothesized to be an RNA-editing enzyme [1], [22]. However, preponderant evidence indicates that AID acts on DNA, more specifically, ssDNA. AID expressed in *E. coli* promotes C to T transitions in the genomic DNA; an effect greatly

enhanced upon inactivation of uracil repair [1], [23]. Additionally, recombinant AID protein can deaminate ssDNA *in vitro*. Chromatin-immunoprecipitation (ChIP) experiments have demonstrated that AID is associated with S regions in cells undergoing CSR [1], [22]. More importantly, ablation of uracil repair by disrupting both UNG2 and MSH2 totally abolishes CSR and forces all SHM events to C to T transition, unequivocally pointing to a direct role in deamination of DNA cytosines.

AID is not sequence-specific, but it preferentially deaminates sequences that conform to WRC (W=A/T, R=A/G) motifs, thus WRC is known as an AID hotspot [13], [24], [25]. S regions are highly enriched with WRC motifs, mainly in the form of the AGCT motif that is conserved in all S regions [6], [26]. AGCT is a palindromic sequence such that the antiparallel DNA strand is also AGCT, thus two AID hotspots are overlapping each other. AGCT is one of four possible WGCW motifs (W=A/T) that result in overlapping AID hotspots across two DNA strands. It has been hypothesized that AID deamination of cytosines on both DNA strands within the same WGCW motifs would result in closely spaced nicks on both strands, therefore generating a DSB in these particular S regions. Disruption of the WGCW motif impairs CSR, indicating that these motifs are functionally important for efficient CSR [26].

Although the Ig locus is the physiological target of AID, AID has been shown to generate off-target mutations and DSBs in non-Ig genes, such as Myc and Bcl-6 [27]–[30]. Thus, AID expression is likely subjected to regulation on multiple levels to prevent off-target mutations and chromosomal translocations [31]. AID is not expressed in resting naïve B cells, resting memory B cells, or plasma cells, but is

instead expressed mainly in germinal center B cells that can be activated through both T-dependent or T-independent mechanisms. These mechanisms induce AID expression through activation of both the canonical and non-canonical NF- $\kappa$ B pathways [32]. AID can be regulated by both post-transcriptional and post-translational mechanisms. MicroRNAs such as miR-155, miR-181b, and miR-93 can bind to the evolutionarily conserved target sites in the 3'UTR of AID mRNA, thereby, reducing both AID mRNA and AID protein levels [33]–[37]. Additionally, post-translational regulation of AID occurs in multiple ways. AID shuttles between the nucleus and the cytoplasm of B cells that is regulated by a less defined nuclear localization signal and a well-defined C-terminal nuclear export signal [31], [38]. AID's function in CSR and SHM has been reported to be regulated by phosphorylation and dephosphorylation at multiple sites [31], [38].

Over the past ten years, our understanding of the regulation and function of AID has advanced significantly. However, many questions still remain unanswered. Precisely how AID is targeted to V and S regions much more frequently than other genomic loci is still unclear. In addition, it is known that the C-terminus of AID is essential for CSR, but non-essential for SHM [31], [39]. The function of this C-terminal region in CSR remains undetermined. The work described in Chapter 3 focuses on elucidating AID action at a nucleotide level and identifying precise locations and the level of AID deamination activity at S regions during CSR.

## 1.4 The Role of Transcription in Class Switch Recombination

Each C exon is organized as an independent transcriptional unit consisting of a cytokine-dependent promoter, I exon, an intronic S region, and a corresponding C region exon. Each promoter responds to a different combination of cytokines that directs CSR to a specific S region. For example, interleukin-4 (IL4) promotes CSR to IgG1 and IgE, whereas transforming growth factor (TGF)  $\beta$ 1 promotes CSR to IgA [1], [40].

Germline transcription (GLT) through a S region is required for CSR to the corresponding isotype. The primary transcript produced by GLT consists of an I exon, followed by an intronic S region, and C region exons. This primary transcript is spliced, in which the intronic S region is removed, and polyadenylated as a typical mRNA. However, the mature transcript does not encode any protein (Figure 4), hence the name “sterile transcript”.

Transcription through the S region is absolutely required for CSR. Deletion of the germline promoter abolishes CSR to the corresponding isotype, but CSR can be rescued by insertion of a constitutively active heterologous promoter [41]–[48], albeit the normal cytokine regulation of the CSR is lost. In addition, GLT is regulated by a large enhancer region called the 3' regulatory region (3'RR), which is located at the 3' end of the last C region (C $\alpha$  in mice). The 3'RR spans ~28kb and consists of four DNase I hypersensitive (HS) sites (~1kb each). Deletion of individual HS sites has little effect on GLT or CSR [18]. However, deletion of the entire 3'RR inhibits S region transcription and severely impairs CSR to most S regions [49]–[52]. The work described in Chapter 2 focuses on identifying DNA

sequence motifs within 3'RRs that are critical for AID targeting to S regions and CSR.

GLT is thought to transiently separate the two DNA strands to grant access to AID to initiate CSR. It has been well established that GLT through S regions creates R-loops, where the G-rich RNA transcript forms an RNA-DNA hybrid with the C-rich DNA template strand while the non-template DNA strand is single stranded [8], [53]–[55]. The extensive single-stranded regions on the non-template DNA strand was thought to be the binding motif for AID. Elegant mouse studies from the Alt laboratory has found that inversion of the Sy1 region, which disfavors R-loop formation, greatly impairs CSR [12], attesting the important role of R-loops during CSR. However, although the R-loop model explains many features of CSR, certain aspects are not readily explained, mostly on how AID access the template strand and why all switch regions are so repetitive.

In addition to transcription itself, it has been suggested that AID is recruited to the switch regions via interactions with the transcriptional or splicing machinery [1], [43], [56]. AID interacts with RNA polymerase II and its associated proteins, such as Spt5, PAF1, and the FACT histone chaperone complex [56]–[58]. Histone modifications during transcription may also play a role in AID recruitment to S regions through the interaction of AID with the modified histones [59]. Although sterile transcripts do not encode proteins, recent data suggests the intronic S region RNA may have a role in recruiting AID to the S regions [60].

## 1.5 Generation and Resolution of DNA Double Stranded Breaks

CSR occurs through the generation and repair of DNA double strand breaks (DSBs) within the kilobase-long S regions that precede each C region. Once DSBs are generated in the donor and acceptor S regions, the distal DSBs are joined by NHEJ. The intervening DNA is deleted, resulting in the V region juxtaposed to a new C region. The evidence in support of DNA DSBs as essential intermediates has been abundant. Phosphorylated histone H2AX ( $\gamma$ -H2AX), a marker for DNA DSBs has been shown to be induced at Ig loci of B cells undergoing CSR. DSBs can be detected via Ligation-Mediated PCR (LM-PCR) [1], [39]. Disruption of NHEJ factors (e.g. Ku86/70, XRCC4, Lig4) markedly impairs CSR efficiency.

DSBs occur as a result of the processing of AID-generated uracils in S regions. Uracils can be repaired via two DNA repair pathways: Base Excision Repair (BER) and Mismatch Repair (MMR) (Figure 5) [1], [39].

In BER, uracils are recognized and excised by uracil DNA glycosylase 2 (UNG2), generating abasic sites, or apurinic/apyrimidinic (AP) sites [1], [39]. Mammalian cells express four uracil DNA glycosylases: UNG2, SMUG1, MBD4, and TDG, where UNG2 is the major uracil DNA glycosylase involved in CSR [39]. Studies have shown that B cells from UNG2-deficient mice have impaired CSR [61], [62], whereas SMUG1-deficient cells did not show any defect in CSR [63]. Additionally, SMUG1 can partially compensate for UNG2 in UNG2-deficient cells. However, cells lacking both UNG2 and SMUG1 show 5-fold greater reduction in CSR efficiency compared to cells lacking only UNG2 [64]. AP sites generated by UNG2 are then cleaved by apurinic/apyrimidinic endonuclease 1 (APE1),

introducing nicks to S regions. Normally, the ensuing removal of 5'dRP and re-synthesis by polymerase beta will restore the DNA to its original C:G pair. In CSR, it is suspected that APE1 generated nicks are repaired with greatly reduced efficiency. It is often conceived that two closely spaced nicks on opposite strands in the S region is a main mechanism for DSB formation [1], [39]. Mammalian cells express two AP endonucleases: APE1 and APE2 [39], [65]. APE1 is the main AP endonuclease (accounting for > 90% of cellular APE activity) and is essential for embryonic development in mice and the viability of human cell lines [66]–[68]. On the other hand, APE2 has extremely low endonuclease activity as such that some even question if APE2 should be categorized as an APE [69]. Earlier studies investigating the role of APE1 and APE2 in CSR have shown conflicting results [70], [71]. A later study establish an essential role of APE1, but not APE2, in CSR in CH12F3 cells [65].

In MMR, mutS homolog 2 (MSH2)/MSH6 heterodimer can recognize U:G mismatches and recruit MutL homolog 1 (MLH1), post-mitotic segregation protein (PMS2), as well as Exonuclease1 (EXO1) to AID deamination sites. PMS2 has an endonuclease activity, which may provide an entry point for EXO1. But Exo1 can also load from an existing nick (e.g. generated by APE1). EXO1 can initiate strand excision from a nick in a typical patch repair as in any MMR event. It is conceivable that such resection, upon encountering a nick or another EXO1-dependent resection on the opposite strand, represent another mechanism of DSB formation [39], [63], [72].

The BER pathway is perhaps the major pathway involved in DSB formation during CSR. Studies have shown that UNG2 deficiency results in a severe CSR defect (i.e. 10% of WT level) [61], [62]. MSH2-deficient cells show a modest 2~3-fold reduction in CSR efficiency [63], [72], [73]. When both UNG2 and MSH2 are ablated, CSR is completely abolished, indicating both BER and MMR contribute to the DSB formation [39], [74].

Once DSBs are generated in both donor and acceptor S regions, the intervening DNA is deleted, and the two S regions are joined to complete the recombination. The presence of S region DSBs activates several DNA damage response factors including ATM, 53BP1 and  $\gamma$ -H2AX, which are thought to be involved in recruiting and coordinating DNA repair factors [1], [39], [75], [76]. S region breaks are predominantly joined by the NHEJ pathway, the major DSB repair pathway in all mammalian cells. When NHEJ is compromised, joining can be mediated by less well defined mechanisms collectively called alternative end-joining (A-EJ), resulting in CSR at reduced efficiency. [1]. Cells lacking any of the core c-NHEJ factors such as Ku70, Ku86, DNA Ligase IV, or XRCC4 are severely impaired for CSR (20-50% of wildtype levels) [1], [77], [78]. Although A-EJ is not well understood, some general DNA repair factors including XRCC1, Ligase III, Mre11, Parp1, and CtIP, have been implicated in A-EJ during CSR [80]–[83].

### **1.6 CH12F3 – A Cell-line to Study Class Switch Recombination**

Our lab studies the mechanism of CSR by performing genetic manipulations in a mouse B cell line (CH12F3) capable of robust CSR *in vitro*. CH12F3 cells

originated from murine B cell lymphoma CH12 cells. CH12 cells were derived in B10 H-2<sup>a</sup>H-4<sup>b</sup>/Wts mice when sheep erythrocytes (SRBC) were transferred to study lymphocyte differentiation [84]. A small portion cells derived from CH12 cells are capable of switching to IgA expression upon stimulation [85]. However, these CH12LX cells had two disadvantages for studying the molecular mechanism of CSR: low frequency and background due to spontaneous switching. To overcome these problems, Honjo's group subcloned CH12LX and established CH12F3 [86].

CH12F3 cells switch efficiently from IgM to IgA upon stimulation with cytokines (IL4 and TGF- $\beta$ 1) and CD40 ligand (CD40L) [23]. CSR efficiency can be measured by stimulating cells and stain them for IgA surface expression using a FITC-conjugated anti-mouse IgA antibody. CSR efficiency is determined by the percentage of IgA (switched) cells in the culture. CH12F3 cells have a stable, near-diploid genome. In the particular subclone (CH12F3.2A) used in this study (still called CH12F3), the nonproductive allele has already switched. Therefore, only the productive V<sub>H</sub> allele (harboring functionally rearranged VDJ exon) is in the germline configuration, allowing one-step gene targeting to study *cis*-acting DNA elements at IgH locus important for CSR [16], [23].

## **1.7 Clinical Significance**

CSR is a unique region-specific recombination event that ensures an effective humoral immune system against pathogen infection. Defects in CSR lead to Hyper-IgM syndrome due to failure to make Ig isotypes other than IgM, resulting in chronic respiratory infections. Dysregulated CSR can contribute to autoimmunity

and lymphoid malignancies. For example, human Sporadic Burkitt's lymphomas harbor hallmark c-myc translocations to Ig S regions [19], [87]–[90]. Understanding the mechanism of CSR is highly significant to human health.

**CHAPTER 2: MUTAGENESIS OF 3'RR TO IDENTIFY DNA SEQUENCE  
MOTIFS CRITICAL FOR AID-TARGETING TO S REGIONS**

**The work in this chapter is published in the following research article:**

**Kim, A., Han, L., Santiago, G., Verdun, R., and Yu, K.; (2016); Class-Switch  
Recombination in the Absence of the IgH 3' Regulatory Region; Journal of  
Immunology; 197 (7) 2930-5.**

## 2.1 Introduction

Class switch recombination (CSR) occurs through the generation and repair of DNA double strand breaks (DSBs) within kilobase long switch (S) regions that precede each constant (C) region. Through this mechanism, only the C region is changed, thus antigen specificity is maintained with different immunoglobulin (Ig) effector functions. This process is essential for an effective immune response [1], [2], [93].

CSR is initiated by the enzyme called activation-induced cytidine deaminase (AID), which converts cytosines to uracils within the S region [1], [2]. Another factor that is absolutely required for CSR is germline transcription (GLT) through an S region for the corresponding isotype. It is well established that AID only targets cytosines in these actively transcribed regions. As AID is a single strand-specific deaminase [25], [94]–[97], transcription separates the two DNA strands (at least transiently), potentially providing AID an access to ssDNA. Each S region is preceded by a cytokine-dependent I promoter and a noncoding I exon [1], [2]. At the onset of CSR, I promoters are selectively activated by cytokine signals that drive the synthesis of a primary transcript. The primary transcript produced by GLT consists of an I exon, followed by an intronic S region, and the C region exons. This primary transcript is spliced to form a 'sterile transcript', which does not encode proteins. Transcription of the donor ( $\mu$ ) and acceptor ( $\gamma$ ,  $\epsilon$ ,  $\alpha$ ) S regions is differentially regulated [4], [52], [98]. Although the  $S_\mu$  GLT is constitutively expressed, transcription of acceptor S regions is regulated by cytokine signaling and depends on a large enhancer complex called the 3'

regulatory region (3'RR) located 3' of the last C region (C $\alpha$  in mice) [4], [52], [98]. The 3'RR spans ~28kb, which consist of four DNase I hypersensitive (HS) sites (~1kb each). The four HS sites are in the order of HS3a, HS1,2, HS3b, and HS4, separated by larger spacer regions. Deletion of individual HS sites has minimal effect on GLT or CSR [4]. However, deletion of the entire 3'RR inhibits S region transcription and severely impairs CSR to most downstream isotypes. Each individual HS site is a weak enhancer but together, they form a strong enhancer and a locus control region that confers position-independent B cell lineage-specific expression to associated transgenes [99].

It has been widely accepted that the 3'RR has multiple functions in CSR (e.g., enhancing transcription, recruiting AID, promoting S region synapsis) [4], [93]. However, because the 3'RR is essential for GLT, it has been difficult to distinguish whether 3'RR has a transcription-independent function in AID targeting. The *cis*-elements responsible for the multiple proposed functions of 3'RR have not been well characterized.

Historically, the 3'RR has been difficult to study by gene-targeting strategies because of its size and sequence complexity. In this study, the 3'RR in a mouse B cell line was deleted and replaced with test sequences to explore the role of the 3'RR in AID targeting during CSR.

## **2.2 Materials and Methods**

### **2.2.1 Reagents, cell culture, and CSR assay**

CH12F3 cells were cultured in RPMI 1640 medium supplemented with 10% (v/v) FBS and 50 mM  $\beta$ -mercaptoethanol. For CSR assays, viable CH12F3 cells were seeded at  $5 \times 10^4$  cells/ml in medium containing 1 mg/ml anti-CD40 Ab (eBioscience), 5 ng/ml IL-4 (R&D Systems), and 0.5 ng/ml TGF- $\beta$ 1 (R&D Systems) and grown for 72 hr. Cells were stained with a FITC-conjugated anti-mouse IgA Ab (BD Biosciences) and analyzed on an LSR II flow cytometer (BD Biosciences). CSR efficiency was determined as the percentage of IgA<sup>+</sup> cells. Rat anti-human/mouse AID Ab used for Western blotting was purchased from eBioscience. Anti-AID Ab for chromatin immunoprecipitation (ChIP) analyses was purchased from Active Motif. ChIP assays were performed by a collaborator [100].

### **2.2.2 Gene targeting**

Each homology arm (~2 kb) for gene targeting was PCR amplified from the CH12F3 genomic DNA and cloned into the targeting vector (Figure 6A). A linearized targeting vector was transfected by electroporation with an Amaxa Nucleofector II device, according to the manufacturer's instructions (Kit V, Program K-005). Ten million transfected cells were distributed in five 96-well plates. Puromycin was added to a final concentration of 1 mg/ml at 48 hr post-transfection. Puromycin-resistant colonies were selected after 7–10 d and screened by PCR. Candidate clones were further analyzed by Southern blotting to confirm the desired

genotype. CRISPR/Cas9-mediated genome editing [101] was carried out by co-transfecting 1 mg each of the hCas9 and gRNA vectors into  $2 \times 10^6$  cells. Transfected cells were allowed to recover for 48 hr before seeding into single wells in 96-well plates by limited dilution. Cell clones were selected and analyzed by PCR, followed by Sanger sequencing of the PCR products to confirm the desired indels.

**Table 1. Oligonucleotides used to amplify homology blocks**

Homology block 1 Fwd	5' CTTTGGCCCAGAAATCACAT 3'
Homology block 1 Rev	5' ACCCCAGCTAGCTCACTTCA 3'
Homology block 2 Fwd	5' CATGGGTTCACGTGAGAATG 3'
Homology block 2 Rev	5' CAGAGGCCATGTACCCTGTT 3'
Homology block 3 Fwd	5' GCATGCAGAGAAGGCATGTA 3'
Homology block 3 Rev	5' TGA CTAGGTT CGCAGGAG 3'
Homology block 4 Fwd	5' TGGGTGGAAGTTGACGTGTA 3'
Homology block 4 Rev	5' AACTCCCTGGGACACAGATG 3'

### 2.2.3 Recombinase-mediated cassette exchange

5ug of exchange vector and 1ug of Cre-expression vector were co-transfected by electroporation. Transfected cells were seeded in 96-well plates at ~100 cells/well and grown for 72 hr before ganciclovir (GANC) was added to a final concentration of 2 mg/ml. GANC-resistant colonies were screened by PCR, followed by Southern blot analyses.

## 2.2.4 RT-PCR

Cells were cultured to a density of  $4 \times 10^5$  cells/ml before cytokines (CIT: anti-CD40, IL4, and TGF- $\beta$ ) were added. 24 hr after the addition of cytokines, total RNA was extracted using TRIzol reagent (Life Technologies), according to the manufacturer's instructions. 1 $\mu$ g of RNA was reverse transcribed into cDNA with random hexamers and Superscript III reverse transcriptase (Life Technologies) in a 20 $\mu$ l reaction. Two microliters of the reverse transcription mixture were used as a template for the PCR reactions.  $\beta$ -actin was used as an internal control.

**Table 2. Oligonucleotides used in RT-PCR**

I $\alpha$ -C $\alpha$ Fwd	5' CCTGGCTGTTCCCCTATGAA 3'
I $\alpha$ -C $\alpha$ Rev	5' GAGCTCGTGGGAGTGTCAGTG 3'
I $\mu$ -C $\mu$ Fwd	5' CTCTGGCCCTGCTTATTGTTG 3'
I $\mu$ -C $\mu$ Rev	5' GAAGACATTTGGGAAGGACTGACT 3'
$\beta$ -actin Fwd	5' TCAGAAGGACTCCTATGTGG 3'
$\beta$ -actin Rev	5' TCTCTTTGATGTCACGCACG 3'

## 2.3 Results

### 2.3.1 Gene targeting of 3'RR

The large size (~28kb) of endogenous 3'RR makes performing mutagenesis directly on the entire 3'RR very difficult. To facilitate mutagenesis of 3'RR, a two-step gene-targeting strategy was carried out to replace 3'RR on the productive IgH allele (rearranged VDJ) in the CH12F3 cell line with an RMCE cassette [26] (Figure 6A). The S $\mu$  and S $\alpha$  of the nonproductive IgH allele in this cell line are already recombined [102]; therefore, the nonproductive allele hybridizes to the 3' probe but not the 5' probe (Figure 6B). First, gene targeting was performed to introduce a LoxP site to the upstream of 3'RR (Figure 6A, step 1). This step deletes HS3a and leaves a LoxP site at the upstream of the remaining enhancers. The second step of gene-targeting inserts an RMCE cassette at the downstream of HS4 (Figure 6A, step 2). This cassette contains puromycin resistance and thymidine kinase genes flanked by one wild-type and one mutant LoxP sites. The 5' LoxP site of the RMCE cassette is identical and in the same orientation as the LoxP site introduced in step 1 (Figure 6A). A subsequent Cre/LoxP deletion removed the entire 28-kb 3'RR, leaving the RMCE cassette that can be used to insert test sequences (Figure 6A). Three independent clones (#3, #6, and #12) were obtained, which are hereafter termed 3'RR-RMCE. For RMCE, a test sequence is cloned into an exchange plasmid flanked by the same pair of heterologous LoxP sites as those on the chromosome (Figure 6C). Co-transfection of the exchange plasmid and a Cre-expressing plasmid results in a swap of the LoxP-flanked sequences between the

chromosome and the exchange plasmid (Figure 6C). A successful exchange event was identified via GANC selection, as cells lose the thymidine kinase gene (Figure 6C), then confirmed by PCR and Southern Blot.

### **2.3.2 Reduced CSR in 3'RR-deleted CH12F3 cells**

Studies performed in mouse models have shown that deletion of 3'RR on a bacterial artificial chromosome–derived transgene [50], [98] and at the endogenous IgH locus [52] abolishes GLT (except for S $\gamma$ 1) and CSR to all acceptor S regions. Unexpectedly, replacing entire 3'RR with an RMCE cassette in the CH12F3 cell line only reduces CSR to IgA. Reduction in CSR is more dramatic at earlier time points, suggesting slower CSR kinetics. However, within 72 hr, CSR levels reached 20–40% of wild-type (WT) controls in the three clones tested (Figure 7). Although CSR efficiency varied among the three clones, all had displayed substantial CSR. This is markedly different from the complete inhibition of CSR in 3'RR-deleted mice [50], [52], [98]. Because S $\alpha$  GLT is dependent on 3'RR in mouse models, a steady-state level of S $\mu$  and S $\alpha$  GLT in 3'RR-deficient cells was measured by semi-quantitative PCR. Consistent with mouse models, S $\mu$  GLT was minimally affected, with or without cytokine stimulation, in WT or 3'RR-deficient CH12F3 cells (Figure 7B). In contrast, although present at low levels in unstimulated WT cells, S $\alpha$  GLT was undetectable in unstimulated 3'RR-deficient cells (Figure 7B). However, upon cytokine stimulation, 3'RR-deficient cells expressed abundant S $\alpha$  GLT that was only slightly lower than that observed in stimulated WT cells (Figure 7B). Among the three clones, clone #12 showed the

lowest level of GLT, which likely accounts for its lower CSR efficiency. These data suggest that 3'RR is not absolutely required for S $\alpha$  GLT in stimulated cells, which is consistent with the observed CSR in 3'RR-deleted cells. AID expression was not significantly changed upon deletion of 3'RR (Figure 7C).

In order to rule out an artifact caused by the RMCE cassette in these 3'RR-deleted cells, the cassette was removed via exchange with an empty exchange vector (Figure 8A, 8B). As expected, removal of the RMCE cassette did not alter CSR efficiency (Figure 8C). To determine whether the 3'RR on the nonproductive allele can promote CSR on the productive allele in *trans*, a large deletion on the nonproductive allele (from the 5' of J<sub>H</sub>1 to the 3' of HS4) was performed by CRISPR/Cas9-mediated genome editing (Figure 9A, 9B); this resulted in a cell line lacking the 3'RR on both alleles. This deletion had no effect on CSR in the 3'RR-RMCE clones (Figure 9C), indicating that CSR in these cells is not mediated in *trans* by the 3'RR on the nonproductive allele. This experiment also eliminated the possibility that the observed CSR was actually *trans* CSR between two interallelic S regions (Figure 9A, 9B).

### **2.3.3 Hyper-CSR upon knock-in of the four HS sites**

To determine which elements within the 3'RR are required to facilitate WT CSR, a fragment (~2.5 kb) containing only the four HS sites was inserted into the RMCE cassette (Figure 10A, 10B). The DNA fragment was made based on studies showing that deletion of the four HS sites on a bacterial artificial chromosome transgene inhibited GLT and CSR to the same degree as deletion of the entire

3'RR [103]. Although the spacer regions were not deemed dispensable, it appears that the primary regulatory elements are located within these enhancers. Moreover, a DNA fragment containing the four HS sites adjacent to one another was used previously to drive B cell-specific gene expression in transgenic mice, suggesting this fragment can function as a locus control region [99]. Introducing just the four HS sites resulted in a hyper-CSR phenotype in many of the daughter clones (Figure 10C). This was observed with all three 3'RR-RMCE clones (Figure 10D). These data strongly suggest that the four HS sites are the primary drivers of the CSR-enhancing functions of 3'RR. As a control, a size-matched irrelevant DNA fragment from mouse Lig4 intron 1 was tested in parallel and failed to restore CSR in all of the three 3'RR-RMCE clones (Figure 10D). Two of the HS knock-in subclones (#3.5 and #3.3) that displayed the largest range in CSR efficiency were selected for GLT analyses (Figure 10C). Although only clone #3.5 had detectable GLT without stimulation, both clones expressed comparable levels of GLT after stimulation (Figure 10E). Thus, the differences in CSR efficiency after introducing the four HS sites cannot be directly explained by different levels of GLT. As expected, knock-in of the four HS sites did not alter AID expression (Figure 10F). AID recruitment to S regions was examined by ChIP analyses. AID occupancies at S $\mu$  and S $\alpha$  were not affected by the deletion of 3'RR or by knock-in of the four HS sites (Figure 10G). The slightly reduced ChIP signal for S $\alpha$  in clone #12 may reflect the relatively low GLT in this clone. These data indicate that 3'RR is not required for recruitment of AID to S regions in CH12F3 cells.

## 2.4 Discussion

This study shows that deletion of the 3'RR in the CH12F3 cell line only modestly reduces CSR to IgA, which is different from mouse models. It has been suggested that the 3'RR may have distinct functions at different stages of B cell development [104]. Thus, the difference between mouse models and the cell line could be explained by the fact that the 3'RR is absent throughout B cell development in 3'RR-deleted mice, whereas in the CH12F3 cell line, the loss of the 3'RR occurs at the brink of CSR. It is possible that 3'RR may have a critical role in establishing a local chromatin structure that is conducive to S region transcription prior to the onset of CSR. Once this permissive environment is established, the CSR reaction itself, including AID recruitment, is no longer highly dependent on the 3'RR. In this regard, CH12F3 may represent a developmental stage that is slightly beyond that of naïve splenic B cells. Although naïve splenic B cells require the 3'RR for activating germline transcription, CH12F3 cells seem to be already primed for S $\alpha$  transcription, even in the absence of the 3'RR. A noticeable difference between CH12F3 and naïve B cells is that GLT of acceptor S region is undetectable in naïve B cells prior to cytokine stimulation, whereas S $\alpha$  GLT is detectable in unstimulated CH12F3 cells, implicating a different epigenetic status of the IgH locus between these two cell types.

Transcription is absolutely required for CSR to occur. Thus, transcription control elements are speculated to have an important role in CSR as a whole, but also in the targeting of AID to the S region. The primary transcriptional control elements in CSR are GLT promoters and the 3'RR. Studies have shown the GLT

promoters can be replaced with a heterologous promoter [43]–[48], leaving the 3'RR to be the primary element that defines AID-targeting specificity. In fact, the 3'RR contains many transcription factor binding sites, including E2A sites that have been implicated in recruiting AID [105]–[107]. However, direct analysis of the role of the 3'RR in AID targeting has been difficult because of its essential role in supporting germline transcription in mouse models. The significant level of GLT and CSR in this 3'RR-deleted cell line strongly suggests that AID targeting to IgH can occur without the 3'RR.

Understanding the components that specify AID targeting of the IgH locus remains unrealized. One possibility is that the region being transcribed (i.e., S region) plays a more prominent role in AID targeting during CSR than previously thought. S region transcription can induce a secondary DNA structure known as an R-loop, which contains an extensive stretch of ssDNA that may facilitate AID binding [12], [53]. Alternatively, AID may bind to the G-quartet motifs on the GLT and be targeted back to S regions via an RNA-guided mechanism [108]. The V region is not G rich and does not form an R-loop; thus, AID targeting to V regions during SHM may still require a specific set of transcription factors. It was reported that SHM is greatly reduced in 3'RR-deleted mice [109]; however, there is also a marked reduction in V region transcription [110]. Therefore, the role of the 3'RR in targeting AID to V<sub>H</sub> during SHM remains unknown.

A recent study showed an important role of the palindromic structure of the 3'RR in AID targeting to the V<sub>H</sub> region and SHM [111]. When the palindromic structure was disrupted, it reduced the V<sub>H</sub> region transcription and SHM, but CSR

was minimally affected as long as all four enhancers were present [111]. The finding that HS knock-in (i.e., without the palindromic spacer regions) fully restores CSR is certainly in agreement with that study. Many of the HS knock-in clones displayed the hyper-CSR phenotype, suggesting that there may be negative-regulatory elements in the 3'RR that are located in the spacer regions or are dependent on the overall architecture of the 3'RR. One possible element is the “like-switch” sequences in the 3'RR that mediate “suicide recombination” [112], which was proposed to play a role in B cell homeostasis [112]. The 3'RR-RMCE platform established in this study will be a useful tool for dissecting the *cis*-acting elements within the 3'RR that positively or negatively regulate CSR.

**CHAPTER 3: ELUCIDATING AID-MEDIATED DEAMINATION DURING CLASS  
SWITCH RECOMBINATION**

### 3.1 Introduction

Immunoglobulin (Ig) gene, encoding protein molecules commonly known as antibodies, undergo multiple rounds of genetic diversification to facilitate more effective pathogen clearance. Antigen stimulated B cells make changes in Ig genes via somatic hypermutation (SHM) and class switch recombination (CSR). SHM introduces point mutations at Ig variable (V) regions (encoding the antigen binding domain of Ig) for a chance to produce high affinity antibodies, whereas CSR changes the heavy chain constant (C) region resulting in an altered effector function of the antibody. CSR is achieved by generation and joining of DSBs within the switch (S) regions, deleting the intervening DNA, and juxtaposition of a new C region immediately downstream of the V region [1], [2].

Activation-induced cytidine deaminase (AID) is a B cell-specific factor that initiates both SHM and CSR [17], [18]. It is a single-strand specific cytidine deaminase that converts cytosines to uracils [1], [2]. During CSR, AID-generated uracils are recognized by Uracil DNA Glycosylase 2 (UNG2)-dependent base excision repair (BER) or MutS homolog 2 (MSH2)-dependent mismatch repair (MMR) [1], [2]. When uracils are generated within close proximity on opposing DNA strands, two closely positioned nicks across the DNA strands can occur, resulting in a double strand break (DSB). Joining of distal DSBs from the 'donor' and 'acceptor' S regions through non-homologous end joining (NHEJ) completes CSR and juxtaposes a new C region to the V region. Alternatively, MMR factors can recognize U:G mismatches and initiate a patch repair (i.e. migration of a strand break) that connects discontinuous strands of DNA strands located far apart and

ultimately results in a DSB. Loss of either UNG2 or MSH2 results in impaired CSR [61], [72], while loss of both results in total ablation of CSR [74]. UNG2 deficiency alone results in a severe CSR defect, whereas MSH2 deficiency only moderately impairs CSR [61]–[63], [72].

AID is not sequence specific, but prefers cytosines within the WRC (W=A/T, R=A/G) motif (also referred to as AID hotspots) [13], [24], [25]. The WRC motif is highly enriched in S regions, mostly in the form of AGCT. AGCT is one of the WGCW motifs (W=A/T), which the complementary strand is also WGCW, resulting in overlapping AID hotspots across the two DNA strands. From a previous study in the lab, it has been shown that the overlapping AID hotspots are important for efficient CSR [26], consistent with the general assumption that closely positioned nicks across the two strands initiate DSB formation. However, how AID deaminates S regions during CSR has not been studied in detail. The abundance and precise positions of AID deamination events during a typical CSR cycle is unknown. It is possible that AID deaminations are clustered at a S region (i.e. DSB by close by nicks) or scattered (i.e. rely on MMR-mediated strand excision). It is also unknown the number of and specific location of AID-generated uracils per S region per cell cycle. Due to this gap of knowledge, the process by which uracil repair in the S region leads to DSB formation remains mostly speculative.

Mutations in AID can result in type 2 hyper IgM syndrome type (HIGM2), where patients are characterized by normal or elevated level of IgM but lack IgG, IgA, and IgE. These patients are profoundly susceptible to bacterial infections and develop lymphadenopathies [113]. Most AID mutations that leads to HIGM2 are

inherited as autosomal recessive (AR) traits [19], [114]. Interestingly, several AID mutations (e.g.  $\Delta$ C10, S43P, L98R and R174S) have been reported to possess deaminase activity in biochemical assays but are defective for CSR *in vivo* [19], [91]. Why these AID mutants cannot support CSR despite being an active enzyme is intriguing.

In this study, we aimed to elucidate how AID-generated uracils are distributed in S regions as an important step towards a better understanding of how S region uracils are processed to result in DSBs. We determined the exact positions and abundance of uracils generated by the WT AID as well as a mutant AID, AID<sup>R190X</sup>, that is truncated and lacks 9 amino acids from its C-terminus. In previous experiments, AID<sup>R190X</sup> demonstrated higher enzymatic activity in bacterial antibiotic resistance reversion assays [115]–[118]. Most interestingly, some HIGM patients with the R190X mutation have one WT allele and one mutant allele; thus, the R190X mutation exerts a dominant negative effect [119]. Patients with the R190X mutation appear to have normal levels of SHM [119]. Thus, the R190X is a rare AID mutation that results in separate phenotypes for SHM and CSR.

## **3.2 Materials and Methods**

### **3.2.1 Reagents, cell culture, and Class Switch Recombination assay**

CH12F3 cells were cultured in RPMI 1640 medium supplemented with 10% (v/v) FBS and 50  $\mu$ M beta-mercaptoethanol. For CSR assays, viable CH12F3 cells were seeded at  $5 \times 10^4$  cells/ml in medium containing 1 mg/ml anti-CD40 Ab (eBioscience), 5 ng/ml IL-4 (R&D Systems), and 0.5 ng/ml TGF- $\beta$ 1 (R&D Systems) and grown for 72 hr. Cells were stained with an FITC- conjugated anti-mouse IgA Ab (BD Biosciences) and analyzed on an LSR II flow cytometer (BD Biosciences). CSR efficiency was determined as the percentage of IgA<sup>+</sup> cells. Rat anti-human/mouse AID Ab from eBioscience and human anti-mouse MSH2 Ab from Santa Cruz Biotechnology were used for Western blotting.

### **3.2.2 Gene targeting**

CRISPR/Cas9-mediated genome editing [101] was carried out by co-transfecting  $2 \times 10^6$  cells with 1 mg each of the hCas9 and gRNA vectors. Transfected cells were allowed to recover for 48 h before seeding into single cells in 96-well plates by limited dilution. Cultured clones were selected and analyzed by PCR, followed by Sanger sequencing of the PCR products to confirm the desired indels.

### **3.2.3 Recombinase-mediated cassette exchange**

Exchange vector (5ug) and Cre-expression vector (1ug) were co-transfected into  $2 \times 10^6$  cells by electroporation. Transfected cells were seeded in 96-well plates at  $\sim 100$  cells/well and grown for 72 hr before the selection agent ganciclovir (GANC) was added to a final concentration of 2 mg/ml. GANC-resistant colonies were screened for exchange event by PCR,

### **3.2.4 Uracil-DNA glycosylase assay**

$1 \times 10^6$  cells were collected and resuspended in 50 $\mu$ l 20mM Tris-Cl pH7.5, 100mM NaCl, 10mM EDTA, and 0.5% Igepal CA-630. The resuspended cells were frozen at  $-80^\circ\text{C}$  and thawed twice, then centrifugated at 15,000 rpm from 10 min to remove insoluble cell debris. Uracil-DNA glycosylase assays were carried out by mixing 2 $\mu$ l of the resulting supernatant (whole cell extract) with 5 pmoles of fluorescein-labeled oligonucleotide substrate in a final volume of 10 $\mu$ l for 10 min at  $37^\circ\text{C}$ . 1M NaOH was then added and incubated at  $95^\circ\text{C}$  to induce DNA strand breaks. The reaction was terminated by the addition of 11 $\mu$ l of formamide loading dye (Amersham Pharmacia) and the products were resolved on 10% TBE-urea polyacrylamide gels. The reaction products were visualized using a Typhoon phosphorimager (Amersham Pharmacia).

### **3.2.5 AID footprint analysis**

AID deamination was measured in CH12F3 cells, which are MSH2 and UNG2 deficient, and contain a 1.1kb long core Sa fragment inserted by RMCE to

replace the entire endogenous S $\alpha$  [26]. UNG2 and MSH2 deficiency prevents normal processing of AID-generated uracils in S regions, therefore these uracils remain in the S region of the cells. Upon DNA replication, any Us present within the DNA template strand will be paired with an A in the newly synthesized strand. Then, during the second round of DNA replication, the A in the template strand will pair with T, ultimately resulting in a C to T transition from the original sequence (G to A on the bottom strand). This C to T transition is an indication of AID-mediated cytosine deamination, and is referred to as an “AID-footprint” (Figure 11). These cells are stimulated with CIT medium (RPMI medium with CD40L, IL4, TGF $\beta$ 1) for the indicated length of time.

For bulk analysis, CIT medium was removed after the stimulation, then the core S $\alpha$  was amplified by PCR and cloned into a cloning vector for sequencing. For individual cell analysis, after stimulation, single cells were plated into 96-well plates, and cultured to increase the cell number, allowing for the core S $\alpha$  from each clone to be amplified and sequenced. Amplified core S $\alpha$  will contain two heterogeneous molecules; one from the top and one from the bottom DNA strand. Because AID is a deaminase for single stranded DNA, deaminated sites can be identified by overlapping peaks in the sequence chromatogram, with both the original and mutated base represented at a single position. Heterogeneity of the sequence is then confirmed by sub-cloning and sequencing the PCR product (Figure 12).

### 3.3 Results

#### 3.3.1 Time-lapsed accumulation of AID deamination

In order to illustrate AID deamination, or AID footprints, in CH12F3 cells, several genetic modifications are required. Uracils incorporated in DNA are normally processed by UNG2-mediated BER or MSH2-mediated MMR pathways. To preserve the AID footprints, both BER and MMR pathways need to be disrupted. Thus, *UNG2* and *MSH2* were sequentially disrupted using CRISPR-Cas9 genome editing. Additionally, endogenous S regions are long, repetitive, and GC-rich, which is difficult to amplify by PCR. Thus, the S region under investigation needs to be shortened to an extent that would support efficient CSR but amenable to PCR amplification. To this end, the endogenous S $\alpha$  was replaced with a 1.1kb “core” fragment using the recombinase-mediated cassette exchange (RMCE) technology [26]. The resulting cell line upon completion of the three genetic manipulations is hereafter referred to as S $\alpha$ <sup>core</sup>-DKO. The 1.1kb core S $\alpha$  represents the most uniformly repeated region of S $\alpha$  and has been shown to support CSR at 30% to 40% of the WT S $\alpha$  sequence [26].

To measure the accumulation of AID footprints over time, S $\alpha$ <sup>core</sup>-DKO cells were stimulated with CIT for either one week or four weeks (with necessary split once the culture reached confluency). Because CH12F3 cells can only switch from IgM to IgA, AID footprints were measured in the donor S $\mu$  and acceptor S $\alpha$  regions.

As shown in Figure 13A, each sequenced molecule contains at least one C to T conversion (an AID footprint) and up to more than ten mutations (Figure 13A).

We estimated that CH12F3 cells double every 12-16 hr when cultured in CIT-containing medium [78]. Therefore, over a span of a week, there are 10-12 generations. Based on this estimation, we calculated that an average of less than one (0.7-0.9) AID deamination event occurs at the 1.1 kb core S $\alpha$  per cell per generation. We observed AID footprints were roughly equally distributed between the top and the bottom DNA strands (48% vs 51% at one week and 44% vs 56% at four weeks of stimulation, respectively), and the AID footprints are evenly distributed across the entire S region (Figure 13C, 13D). This observation is consistent with the collapsed R-loop model in which a DNA secondary structure plays a major role in generating ssDNA substrates for AID (Figure 17).

In the core S $\alpha$  region, after one week of stimulation,  $6.8 \times 10^{-3}$  mutations per base had accumulated. When grown to four weeks under stimulation condition, the mutation frequency increased to  $1.3 \times 10^{-2}$  mutations per base, which is ~2-fold higher compared to cells grown for one week. The majority of deaminations occurred in AID hotspots at a similar frequency over time (89% and 80%, respect to time), whereas deamination at cold spots increased at four weeks of stimulation. This suggests a depletion of AID hotspots occurred over extended periods of stimulation and forced AID to seek non-AID hotspot motifs to deaminate (Figure 13C, 13D). We also examined a small region upstream of the donor S $\mu$ , called pre-S $\mu$ . This region was chosen in a number of earlier studies as a proxy to S $\mu$  due to the difficulty of PCR amplification of the intact S $\mu$  region. Similar to that of the core S $\alpha$ , AID deaminates pre-S $\mu$  fairly equally between the two DNA strands. There was a slightly higher number of deaminations at the AID cold spots (25% after one

week and 27% after four weeks of stimulation) (Figure 13F, 13G). These results indicate that AID targets both DNA strands and prefers cytosines in the WRC motif *in vivo*, in line with its WRC preference *in vitro* shown by numerous biochemical assays [25], [120].

### 3.3.2 AID footprints analysis at a single cell level

DSB formation at the donor and acceptor S regions is essential for CSR. AID initiates the DSB formation by targeting cytosines in S regions and converting them to uracils [1]. However, precisely which cytosines within any given S region are deaminated to uracils, especially the relative positions of uracils on the two DNA strands of the same molecule, have not been determined. This information is key to our understanding of the DSB formation mechanism. Like most studies reported so far, the experiments described above were done with pools of cells. During the PCR process, the top and bottom strands each gives rise to a separate PCR product. Thus, it is impossible to reconstitute AID footprints on both strands of the same DNA molecule. To elucidate the specific cytosines involved in DSB during CSR, we performed AID footprint analysis at the single cell level.  $S\alpha^{\text{core}}$ -DKO cells were stimulated with CIT medium for 24hr before cytokines were withdrawn, and the cells were plated in growth medium by limited dilution to develop single clones. The stimulation was carried out for 24hr because AID expression reaches the highest level at this time point [100]. The choice of cell clone versus single cell as the PCR template is due to the technical difficulty of amplifying  $S\alpha$  region from a single cell, especially when there is only one IgH allele

in CH12F3 cells (i.e. just one molecule) [26]. Within each cell clone, only two PCR products are generated, one from the original top strand and one from the original bottom strand. Therefore, the integrity of the top and bottom strand information from a single cell is preserved.

Using this method, a total of 307 PCR reactions (from individual cell clones) were sequenced by Sanger sequencing, of which 18 products showed one or more overlapping peaks in the chromatogram (at Cs, or Gs if deamination is on the bottom strand) (Figure 12), indicating two populations exist in that PCR reaction. These 18 PCR products were individually cloned into a plasmid, and transformed to *E. coli* to separate the two species within each PCR product. After Sanger sequencing of 8-12 bacterial clones from each PCR product, AID deamination profiles on both DNA strands from those individual B cells were reconstituted (Figure 14). In 16 cases, both top and bottom strand-derived sequences were successfully obtained. In the other two cases, only the WT sequences were found after sequencing the bacterial clones, suggesting that the initially observed overlapping peaks on the chromatogram are likely sequencing noises. All but one AID deamination occurred at WRC motif.

Out of the 16 molecules which top and bottom strands are no longer complementary due to AID deamination, 8 molecules had deamination events on only one of the strands, the other 8 molecules contained deaminations on both strands (Figure 14). From this limited number of molecules obtained, we made the following unexpected discoveries. First, the amount of AID deamination events occurred at the entire S $\alpha$  region within one cell cycle is rather sparse, despite the

peak expression of AID at that time point. It is important to note that AID-generated uracils in normal B cells do not persist to the next cell cycle. They are either faithfully repaired to be restored to the original Cs, or replicated over to generate C to T mutations. In this regard, only the uracils generated within a single cell cycle participate in DSB formation. Thus, our data suggest that DSB is formed from a very limited number of uracils generated by AID in each cell cycle. Second, there is only one molecule showing overlapping deamination at WGCW/WGCW motif (Figure 14, #14), a previously widely conjectured mechanism for DSB formation.

In CSR assay, S $\alpha$ <sup>core</sup> cells are capable of switching to ~5% at 24hr of stimulation. Out of 307 amplicons that were sequenced, approximately 3% of the molecules showed at least one mutation on both strands, which can be viewed as molecules that are potentially forming DSB. This suggests that even though AID is capable of targeting multiple cytosines, the minimum requirement for DSB formation might be as low as one deamination event on each DNA strand (Figure 14).

### **3.3.3 R190X AID in CH12F3**

Several AID mutations, (e.g., R190X, S43P, L98R, and R174S) described in HIGM2 syndrome patients, retain catalytic activity *in vitro* but cause defective CSR *in vivo*. Among these mutations, AID<sup>R190X</sup> is the most interesting because it exerts a dominant negative effect in human patients [119]. However, there has not been a cell line, or a mouse model for this particular mutation, to elucidate the molecular basis for the dominant negative effect.

To investigate the alterations of AID footprints caused by AID<sup>R190X</sup>, AID<sup>ΔE5</sup>Sα<sup>core</sup>-DKO was generated. The AID gene in these cells has been engineered by CRISPR/Cas9 genome editing to remove Exon 5. An earlier study using retroviral transduction to complement AID-deficient mouse splenic B cells has reported that AID<sup>ΔE5</sup> has a similar phenotype to AID<sup>R190X</sup> [121]. When overexpressed, AID<sup>ΔE5</sup> or AID<sup>R190X</sup> reduce CSR efficiency mediated by a WT copy of AID cDNA. The caveat of this kind of approach is that retroviral transduction often results in magnitude orders of overexpression of the transgene; leading to artifacts in many occasions [121]–[124]. In this study, two different methods (RMCE and CRISPR-Cas9 genome editing) were used to modify the AID gene in CH12F3 cells at its endogenous locus. These methods preserve the transcriptional regulatory elements intact and effectively avoid the artificial over-expression associated with retroviral transduction.

For RMCE, the entire AID coding region sequences (CDS) was replaced with a Puro-ΔTK positive/negative selection cassette (by homology-based gene targeting) followed by a cassette exchange with a synthetic DNA fragment containing the AID CDS with the R190X mutation (Figure 15A). Because the endogenous S regions are not shortened, direct AID footprint analysis is not practicable. Nevertheless, this engineered cell line is usable for testing for a potential dominant negative effect caused by the R190X mutation of AID. To generate AID<sup>R190X/+</sup> using RMCE, first the region spanning exons 2 to 4 of the AID locus was replaced with an RMCE cassette on one of the two alleles in the CH12F3 cell line. This design allows the mutation of any amino acid of AID except for the

first 3 a.a. encoded by Exon 1. Once AID<sup>RMCE/+</sup> cells were obtained, a cassette containing a fragment bearing the R190X mutation was used to replace the chromosomal cassette (Figure 15A). Loss of one chromosomal copy of AID results in a haploinsufficient phenotype [125]. If R190X has a dominant negative effect, then CSR efficiency would be reduced in AID<sup>R190X/+</sup> cells. Surprisingly, we observed AID<sup>R190X/+</sup> cells were capable of switching to IgA at ~50% of the WT level (Figure 15C). Therefore, instead of a dominant negative effect, the R190X mutation in this mouse cell line displayed a null phenotype in regards to CSR. These results contrast the reduced CSR observed in human patients with AID<sup>R190X/+</sup>, or mouse B cells with retroviral vector-mediated overexpression of AID<sup>ΔE5</sup>.

In parallel to the RMCE approach, CRISPR/Cas9 genome editing was used to generate AID<sup>ΔE5/+</sup> and AID<sup>ΔE5/ΔE5</sup> cells from WT CH12F3 cells. Two Cas9<sup>D10A</sup> nickases with targeting sites flanking exon 5 (Figure 15B) were introduced into CH12F3 cells. It is widely acknowledged that genome editing mediated by two proximal Cas9 nickases generates minimal off-target mutations in the genome [126], [127]. Similar to AID<sup>R190X/+</sup> cells generated via the RMCE approach, AID<sup>ΔE5/+</sup> cells are capable of switching to IgA at ~50% of WT levels while AID<sup>ΔE5/ΔE5</sup> cells do not switch at all (Figure 15C). These data strongly suggest that the C-terminal truncation of AID in this mouse B cell line behaves as a null phenotype rather than a dominant negative mutation shown in human patients. Because there was no antibody available to detect C-terminal truncated AID, the protein levels of AID<sup>R190X</sup> or AID<sup>ΔE5</sup> were not determined. Therefore, there is a possibility that the null

phenotype of AID<sup>R190X</sup> or AID<sup>ΔE5</sup> in CH12F3 cells is a result of the lack of truncated protein, which may be due to the instability of the protein.

In order to determine whether the C-terminal truncated AID deaminates S regions in CH12F3 cells, CRISPR/Cas9 genome editing was performed in Sα<sup>core</sup>-DKO cells to generate AID<sup>ΔE5/ΔE5</sup>Sα<sup>core</sup>-DKO cells, which were stimulated with CIT for one week and analyzed for AID footprints. The Pre-Sμ and Sα<sup>core</sup> regions were PCR-amplified and sequenced. Several published studies have examined the frequency of mutations caused by retrovirally expressed AID (or AID<sup>ΔC</sup>) in CH12F3 or mouse primary B cells [118], [121]. Typically, the C-terminal truncated AID produces more mutations at pre-Sμ region than the WT AID when retrovirally transduced. In contrast, we observed no significant deamination at both Pre-Sμ and Sα regions in cells with endogenous AID<sup>ΔE5</sup> (Figure 16), indicating that AID C-terminal truncation in CH12F3 cells represents a loss of function mutation.

### 3.4 Discussion

This study provides many valuable insights into how AID-deamination occurs within the entire S region, which have not been reported previously. First, we observed that AID is able to target top and bottom strands at a similar frequency. Because AID is a ssDNA deaminase, we had hypothesized AID to target the non-template (top) strand more frequently than the template (bottom) strand by forming an R-loop structure. One possible explanation for our results could be based on a recently reported crystal structure of AID. It was suggested that AID has two DNA binding grooves, resulting in higher binding affinity and deamination activity for structured DNA substrates such as branched or G4 structures [128]. This may allow for AID to bind both DNA strands at the same time and perform deamination at a similar frequency, potentially preferring a different secondary structure known as a collapsed R-loop. R-loops are known to collapse upon RNase H-mediated degradation to generate ss loops on both DNA strands due to switch repeats misalignment [53] (Figure 17). Second, we observed AID deamination evenly across the entire core S region. This is consistent to the mutation profile of V regions during SHM, where the mutation frequency rises up at ~150 bp downstream of the promoter and gradually tails off over a 1.5 kb region with a noticeable peak above the ~400 bp V region [129]. It is possible that the same distribution may occur in the S region based on the observation that small amounts of AID deamination were present at non-S sequence region positioned after the core S $\alpha$ . This data is also consistent with the hypothesis that AID associates and

travels with the RNA polymerase II complex, and AID's access to ssDNA relies on strand separation behind the polymerase due to superhelical tension [130], [131].

As AID footprints from a pooled population of cells can only show the *distribution* of deamination events, we took additional steps to determine the AID deamination action during one cell cycle. From the limited amount of data, we observed the AID deamination events mostly occurred distally during one cell cycle, This challenges two of the widely accepted hypotheses of AID action and DSB formation in S regions. One, DSB formation at S regions was thought to occur when the nicks are closely positioned on both the top and bottom strands, thus making overlapping AID hotspots the most likeable AID target to facilitate DSB formation [1]. Consistent with this hypothesis, it had been shown that short sequence motifs containing overlapping AID hotspots in the form of WGCW (e.g. AGCT) in the S regions are critical for efficient CSR [26]. In contrast, we only observed this pattern in 1 of the 16 molecules analyzed. Contrarily, our data showed only one molecule that had the pattern supporting this idea. Additionally, our data suggests distal nicks may contribute to DSB formation which could explain why the MMR pathway is also required for efficient CSR. As suggested previously [132], [133], EXO1 mediated strand excision might be a mechanism of migrating strand discontinuity on both strands to proximal sites , thus leading to DSB. There can be an alternative explanation for the lack of overlapping deamination at WGCW sites. For example, in collapsed R-loop structure, uracils between the top and bottom strands for many of the molecules shown in Figure 14 actually may be much close to each other (Figure 17), which is sufficient to generate a DSB. This

hypothesis needs to be tested experimentally in the future. Two, another general assumption in the field is that AID deaminates multiple sites during CSR for DSBs. However, our data suggest that the minimum requirement for DSB formation could be as low as one deamination event on each DNA strand.

Due to the disadvantages of Sanger sequencing, we explored the option of a new, or third generation, sequencing method called “Nanopore” sequencing. Sanger sequencing is low throughput, requires multiple steps in addition to PCR, and most importantly, may suffer from amplification biases introduced by PCR. In contrast, third generation sequencing is purported to provide ultra-long reads, is high throughput, and does not suffer from amplification biases. We explored Nanopore sequencing in an attempt to make more efficient progress on the project. Unfortunately, the data obtained using Nanopore was unreliable based on the control experiments. Thus, we will continue to use the original strategy despite a slower pace in order to yield a stronger and more reliable conclusion. Additionally, we plan to explore additional high throughput third generation sequencing strategies (e.g., PacBio) to obtain corroborating data for AID footprints at the single cell level.

In addition to the WT AID, we aimed to delineate the functional defect of AID mutations that cause HIGM2 syndrome. In this study, we focused on AID<sup>R190X</sup>, where the last 9 amino acids from its C-terminus have been deleted. This particular mutation was shown to have a dominant negative effect in human patients as well as in mouse B cells that overexpress this mutant protein[119], [121]. The C-terminus of AID encodes a nuclear export signal that shuttles AID out of the

nucleus [123], it is therefore somewhat counter-intuitive that the C-terminus of AID is required for CSR, which occurs in the nucleus. It has been reported that nuclear AID is subjected to proteasome-mediated degradation [134]. Therefore, C-terminal truncated AID (AID<sup>ΔC</sup>) is unstable due to its primary residence inside the nucleus. In studies where AID<sup>ΔC</sup> was transduced with a retrovirus, the overexpression masks some of the instability associated with the C-terminal truncation. The majority of studies have used fusion proteins at the C-terminus of AID which further stabilizes AID<sup>ΔC</sup> [123], [124], [135]. Our study raised the possibility of AID having a species-specific mechanism. It may be that truncated AID protein in human cells does not operate in mouse cells, accounting for the difference of the dominant negative effect. Recently, a mouse model of C-terminal AID truncation was generated (unpublished, personal communication with Dr. Jayanta Chaudhuri). In this mouse model, AID<sup>R190X</sup> behaves as a null allele, just like what we have observed in the CH12F3 cell line. It is possible the dominant negative effect shown in previous studies is an artifact of an overexpression. The overexpression of AID<sup>ΔE5</sup> showed an increased amount of deamination events in the donor S region, which correlates with the hypothesis that AID<sup>ΔE5</sup> loses the C-terminal nuclear export signal and accumulates in the S regions, inhibiting the binding of other DNA repair factors from completing steps in the CSR process (e.g. DSB formation, end-joining).

Our study indicates that the endogenous mouse AID<sup>R190X</sup> as well as AID<sup>ΔE5</sup> is unable to efficiently deaminate cytosines in CH12F3 cells. The reduced amount of AID deamination may be due to defects in AID protein stability or targeting to S

region. Another possibility is that even though AID<sup>R190X</sup> is stable and can be targeted to the S region, it may be unable to stay bound to deaminate cytosines. Further analysis of AID footprints made by AID<sup>ΔE5</sup> as well as other AID mutations using S $\alpha$ <sup>core</sup>-DKO cells, that were established in this study, may provide additional information regarding organismal differences of AID between mice and humans and why they cause defects in CSR.

## **CHAPTER 4: SUMMARY AND CONCLUDING REMARKS**

Class Switch Recombination (CSR) is one of the processes that diversifies immunoglobulin (Ig) molecules for the effective immune system. CSR is a region-specific event that changes the constant (C) region of the Ig molecule while maintaining the same variable (V) region, allowing the Ig molecule to recognize the same pathogen but alter its effector function. Many proteins are involved in the process, of which activation-induced cytidine deaminase (AID) is a B cell-specific protein that initiates the process. One of the unanswered questions in the field is how AID targets and acts on the switch (S) regions during CSR.

In Chapter 2, we provided evidence that the 3' regulatory region (3'RR) is not absolutely required for recruitment of AID to S regions in CH12F3 cell line. We showed deletion of the entire 3'RR in CH12F3 cell line resulted in a modest reduction rather than complete abolishment of CSR. This suggests 3'RR may have a more prominent role in chromatin remodeling compared to its role in transcription or AID-targeting. Additionally, replacing 3'RR region with just the 4HS sites resulted in hyper-CSR phenotype, indicating these sites are the most critical elements of 3'RR for CSR. This hyper-CSR phenotype also suggests the possibility of negative regulatory elements in the 3'RR. The 3'RR-RMCE platform established in this study will serve as a useful tool for dissecting these *cis*-acting elements within the 3'RR that positively or negatively regulate CSR.

In Chapter 3, we focused on elucidating how AID-generated uracils are distributed in S regions, which will provide a better understanding of the mechanism of double-stranded DNA breaks (DSBs) formation during CSR. Two DNA repair pathways are involved in processing AID-generated uracils, which can

ultimately result in nicks on ssDNA strand. It has been hypothesized that the nicks on the two strands need to be created at near proximity to lead to DSB formation that facilitates CSR. Additionally, it has been hypothesized that AID needs to deaminate many cytosines in S regions to uracils to effectively generate DSBs. Our data suggests that DSB formation does not necessarily need proximal nicks, or perhaps distal nicks can be brought to proximity (i.e., collapsed R-loop). Although AID is capable of deaminating multiple sites, the minimum amount of AID-generated uracils can be as low as one on each strand. This profile of AID deamination at the core S region at a single cell cycle level has never been shown before.

Finally, we aimed to delineate the CSR defect associated with the R190X mutation of AID, in which 9 amino acids are absent from the C-terminus. The C-terminus of AID includes a nuclear export signal (NES) that regulates the location and function of AID. Loss of this region somehow causes type 2 HIGM syndrome in human patients in a dominant negative fashion. Previously, there has been no cell line or mouse model to study the R190X mutation, making us the first group to have made a cell line to study this mutation. In our studies, the AID<sup>R190X</sup> did not exhibit the dominant negative phenotype in the mouse B cell line. Instead, the R190X acted as a null allele and resulted in aa haploinsufficient phenotype. In addition, AID footprints in the mouse B cell line with AID<sup>ΔE5/ΔE5</sup> demonstrated a basal level of deamination in both pre-S $\mu$  and S $\alpha$ . Thus, AID<sup>ΔE5</sup> is unable to deaminate cytosines in S regions even when there is an active catalytic domain present. This suggests there may be a species-specific mechanism of AID

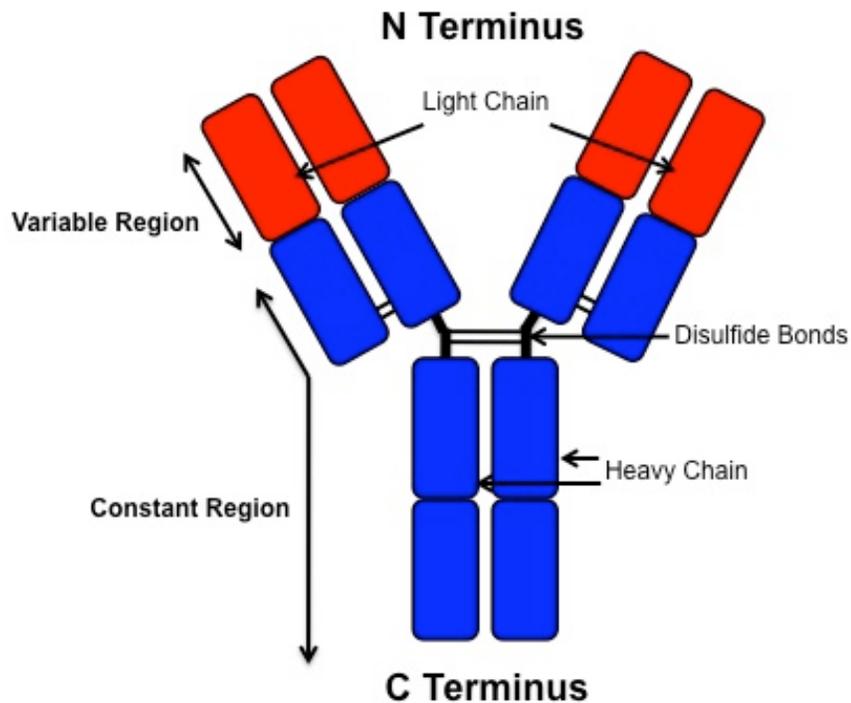
regulation, or AID protein itself, that allows AID<sup>R190X</sup> to be stable in humans and override the action of WT AID. In mice however, the AID<sup>R190X</sup>, loses this stability and the WT AID continues to function. To test this hypothesis, using RMCE strategy, an AID<sup>R190X</sup> fused with GFP at the C-terminus to help with protein stability could be generated in future studies. In addition to the data generated, the S $\alpha$ <sup>core</sup>-DKO cell line and AID footprint platform established in this study have many potentials. They will be useful tools for testing deamination of AID and AID mutations in the S region and targeted sequences. Also, the S $\alpha$ <sup>core</sup>-DKO cell line and AID footprint platform can be used to determine the effect of other proteins that may participate in AID action, targeting, stability, etc.

In summary, the studies presented in this dissertation have increased our understanding of the mechanism of CSR. In addition, the cell lines created as a part of these studies will serve as important tools to study the *cis*-elements responsible for the multiple proposed functions of 3'RR. The cell lines and methods developed in our studies can also be used in the future to elucidate factors that control AID activity and contribute to the defects that occur in patients with HIGM2 syndrome.

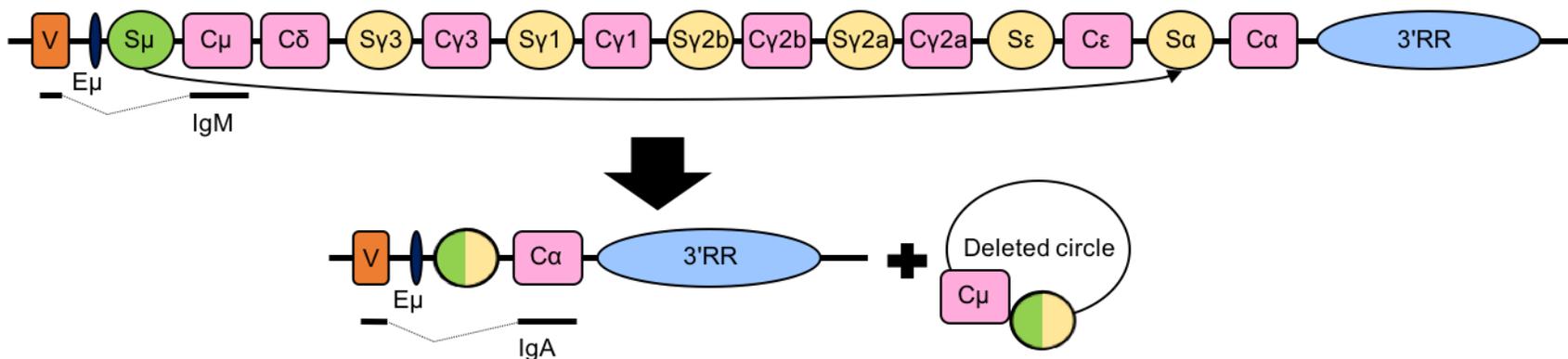
## **APPENDICES**

APPENDIX A: Chapter 1 Figures

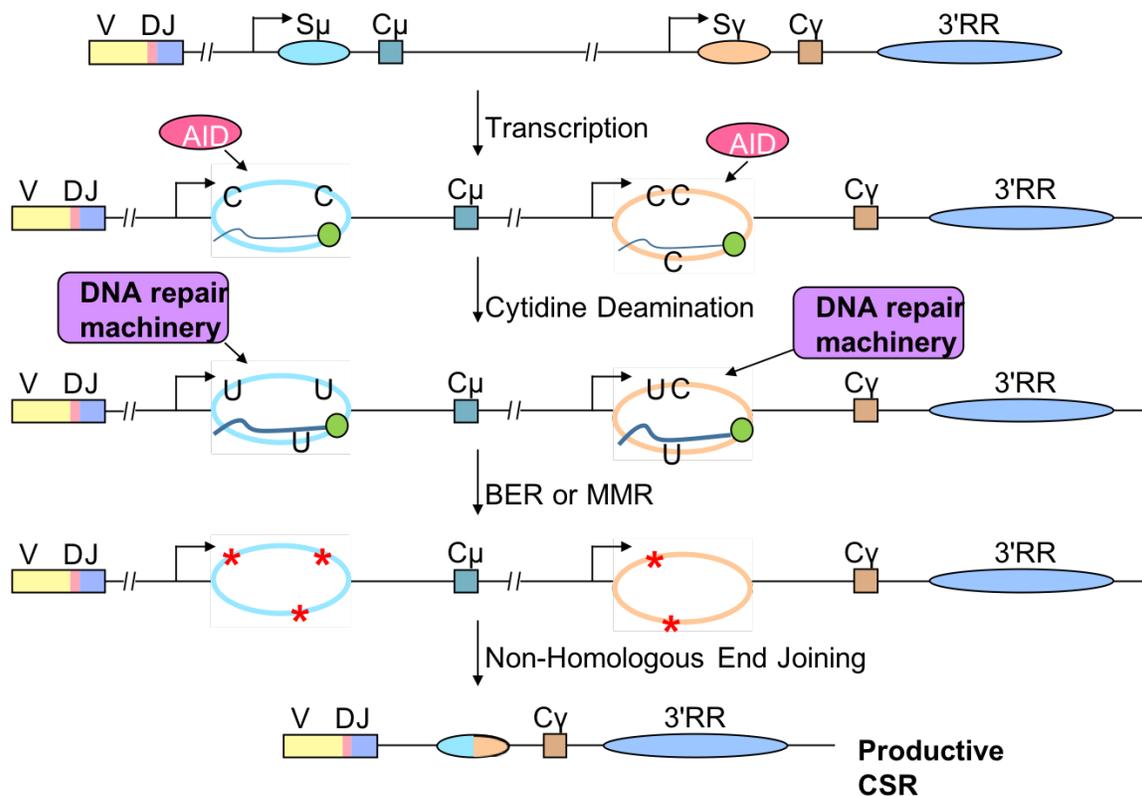
**Figure 1. Structure of an Immunoglobulin Molecule.** Immunoglobulin, or an antibody molecule consists of two heavy and two light chains which are joined by disulfide bonds. They have an N-terminal Variable (V) region (red) and a C-terminal Constant (C) region (blue). The polypeptide chains, light chains, heavy chains, disulfide bonds, variable and constant regions are indicated.



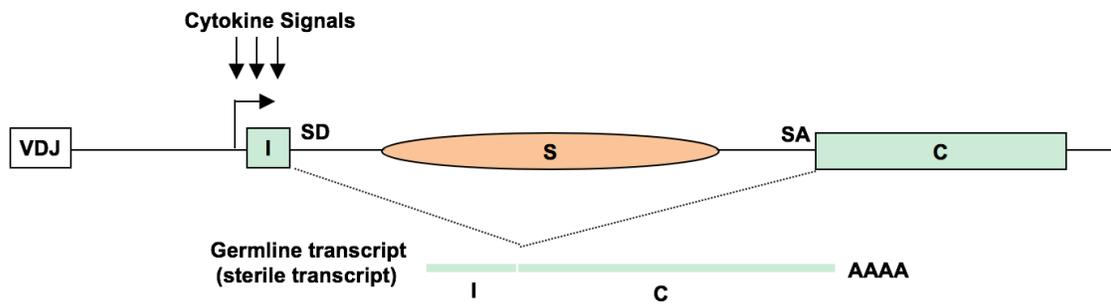
**Figure 2. Overview of Class Switch Recombination.** The Ig heavy chain locus is depicted. Orange rectangle represents variable (V) region. pink rectangles represent constant (C) region exons, colored circles represent switch (S) regions and blue oval represents 3'RR. Each S region precede its corresponding constant region. DSBs occur in donor ( $S_{\mu}$ ) and acceptor ( $S_{\alpha}$ ) then are joined together by NHEJ. The V region is juxtaposed to new constant regions; thus, CSR occurs. Intervening DNA is deleted.



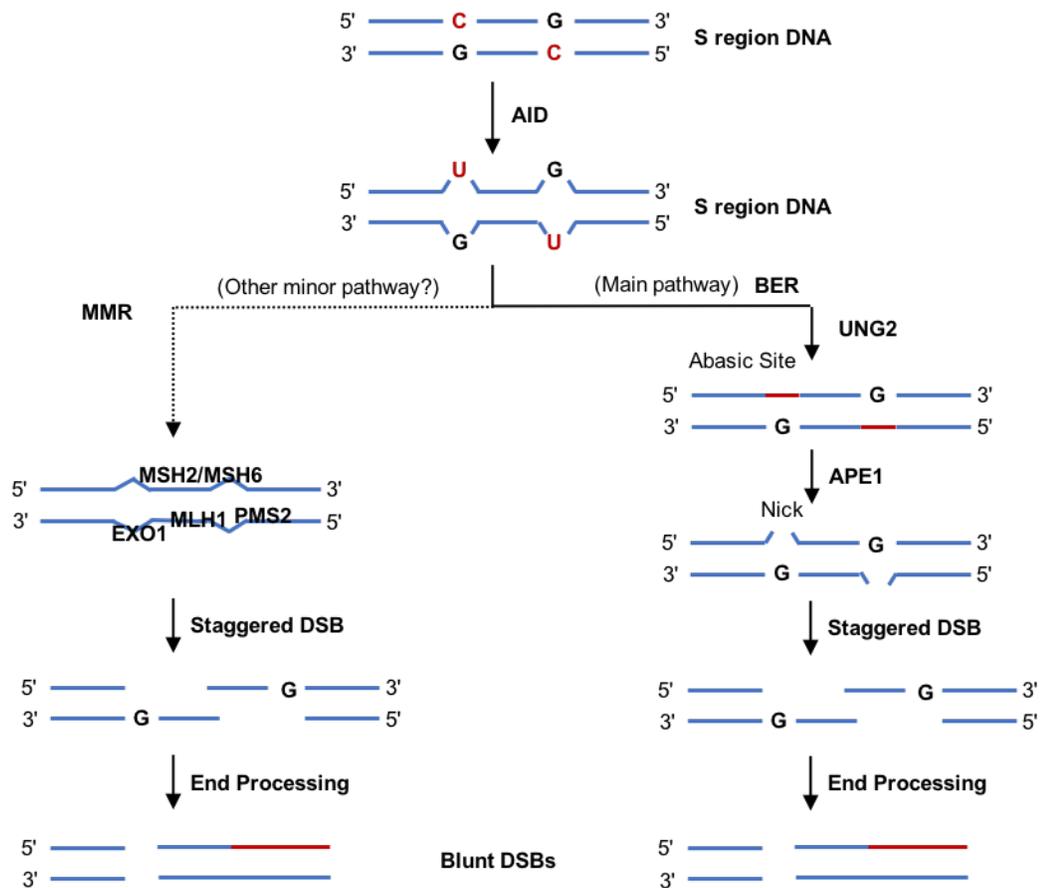
**Figure 3. Molecular Mechanism of Class Switch Recombination.** Cytokine-dependent promoter initiates the GLT. Transcription through S regions results in the formation of secondary structures such as R-Loops and G4s. This provides single-stranded DNA substrate for the action of AID, which converts cytosines into uracils. AID-generated uracils are recognized and processed by UNG2 mediated Base Excision Repair (BER) or MSH2-mediated Mismatch Repair (MMR) pathways. This generates nicks in the S region which can result in DSBs. DSBs are recognized and synapsed, a process that involves multiple factors such as  $\gamma$ -H2AX, 53-BP1 and ATM. The DSBs in the donor and acceptor S regions are joined by NHEJ pathways to complete the recombination process.



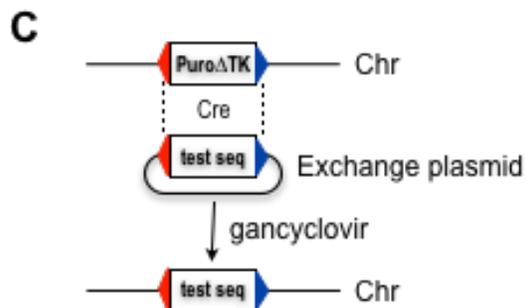
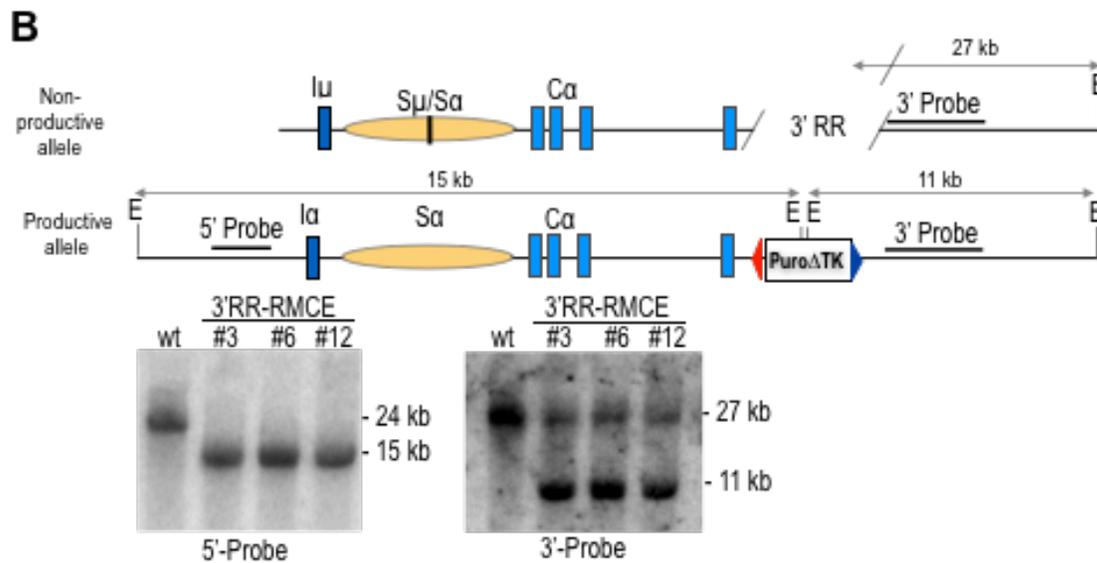
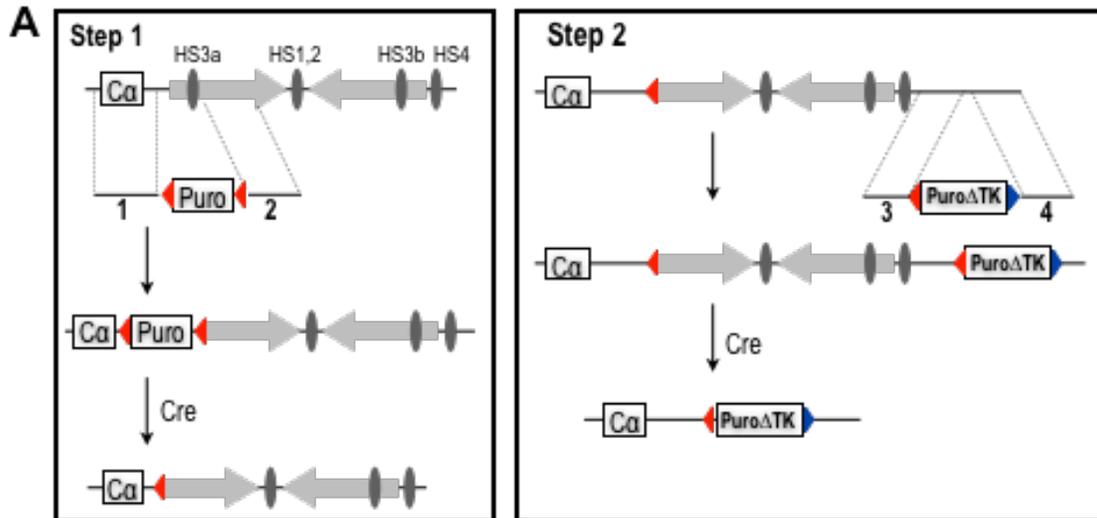
**Figure 4. Germline Transcription during Class Switch Recombination.** Upon cytokine stimulation, transcription activates by the cytokine-dependent promoter. The primary transcript consists of the I exon, intronic switch region and the constant region exons. The primary transcript is spliced to produce the germline transcript that consists of I exon and constant region exons. VDJ, rearranged variable region; I, I exon; S, switch region; C, constant region exons; SD, splice donor; SA, splice acceptor.



**Figure 5. Formation of Double Stranded DNA Break during Class Switch Recombination.** AID deaminates cytosines within S region into uracils. AID-generated uracils are processed through BER or MMR. In BER, which is considered to be the main pathway of uracil processing, UNG2 recognizes and removes uracils leaving abasic sites. These abasic sites are then targeted by APE1 which generates nicks resulting in a DSB. In MMR, MSH2/MSH6 heterodimer recognizes U:G mismatch, recruits MMR machinery including MLH1, PMS2 and EXO1. MMR induces gaps with long overhangs that are processed to generate a blunt ended DSB.



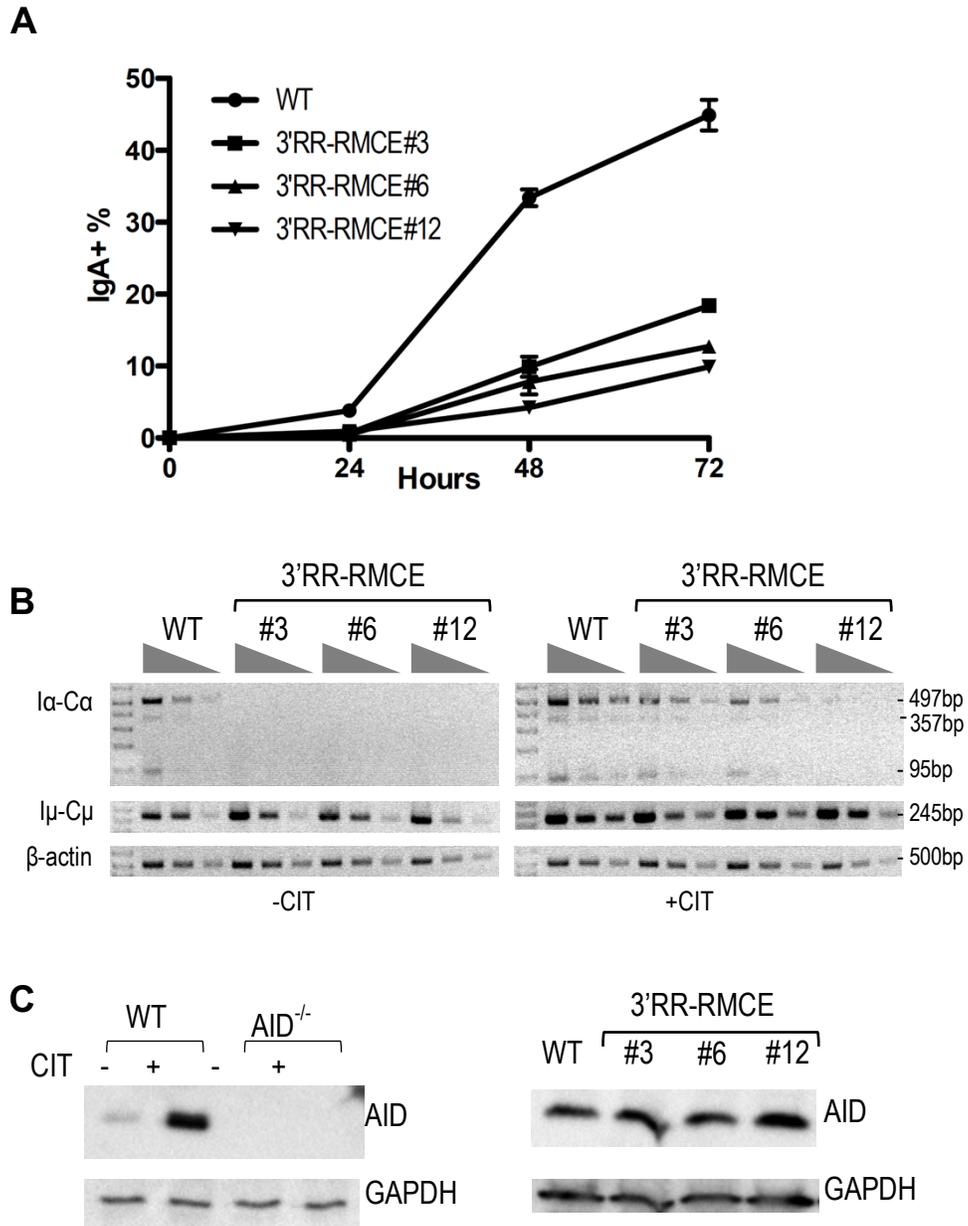
**Figure 6. Gene targeting to replace 3'RR with an RMCE cassette.**



**(Figure 6 cont'd.)**

**A.** Strategy to replace 3'RR with an RMCE cassette. Step 1: insertion of a LoxP site (red triangle) in front of 3'RR. Step 2: insertion of an RMCE cassette and deletion of 3'RR. **B.** Southern blot analysis of the targeted allele. The 5' probe only hybridizes to the productive allele. The 3' probe hybridizes to both alleles. **C.** RMCE. Exchange plasmid containing floxed mutant sequence and Cre expression plasmid are co-transfected into 3'RR-RMCE cells. Successful exchange events are enriched by counterselection against the TK gene using GANC. E, EcoRV.

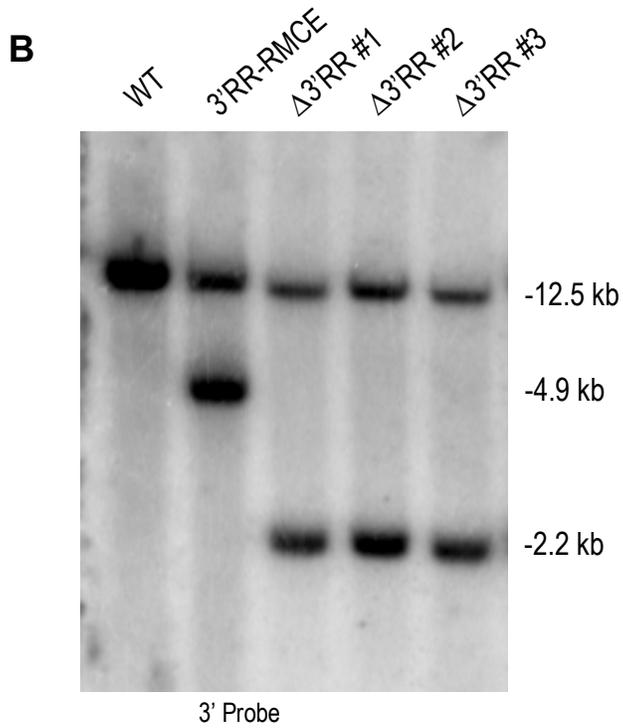
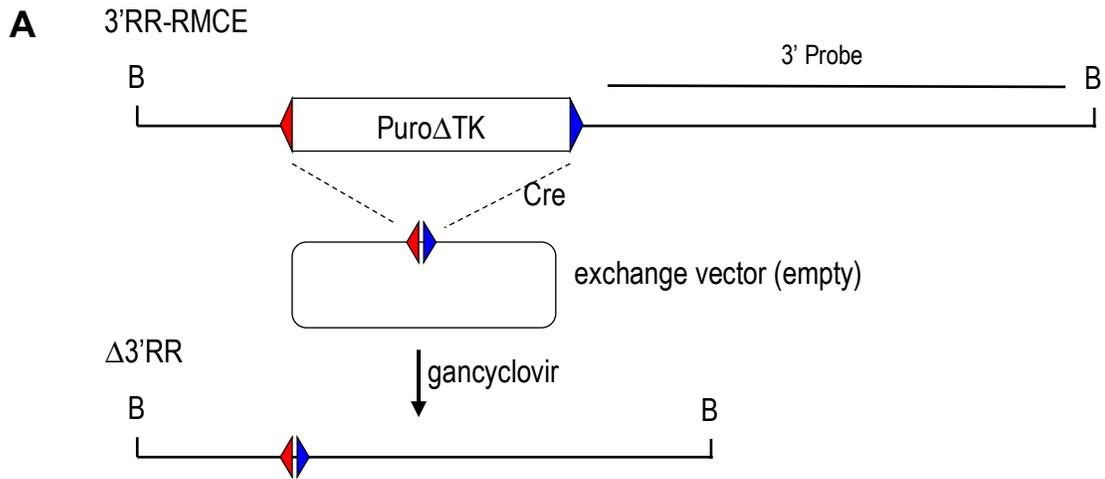
**Figure 7. Reduced Class Switch Recombination in 3'RR-deleted (from the productive allele) cells.**



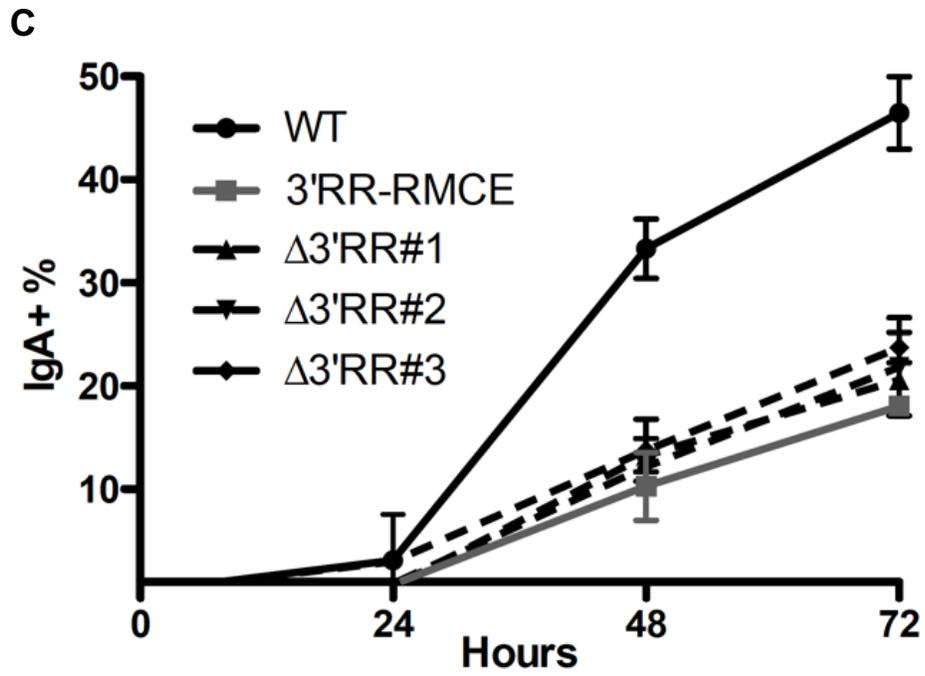
**(Figure 7 cont'd.)**

**A.** A time course of CSR over a period of 72 h for the three independent (#3, #6, and #12) clones from which 3'RR was deleted and replaced with an RMCE cassette (on the productive allele). Error bars indicate SE of three independent experiments. **B.** Semi-quantitative RT-PCR of GLTs (3-fold serial dilution of cDNA). Total RNA was harvested at 24 h after cytokine stimulation. **C.** Western blot analyses of AID expression in WT and 3'RR-RMCE clones. Cells were harvested after 24 h of stimulation. CIT, mixture of anti-CD40 Ab, IL-4, and TGF- $\beta$ 1 used to stimulate cells to undergo CSR.

**Figure 8. Class Switch Recombination in 3'RR-deficient cells after removal of the RMCE cassette.**



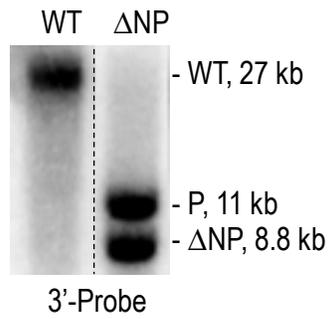
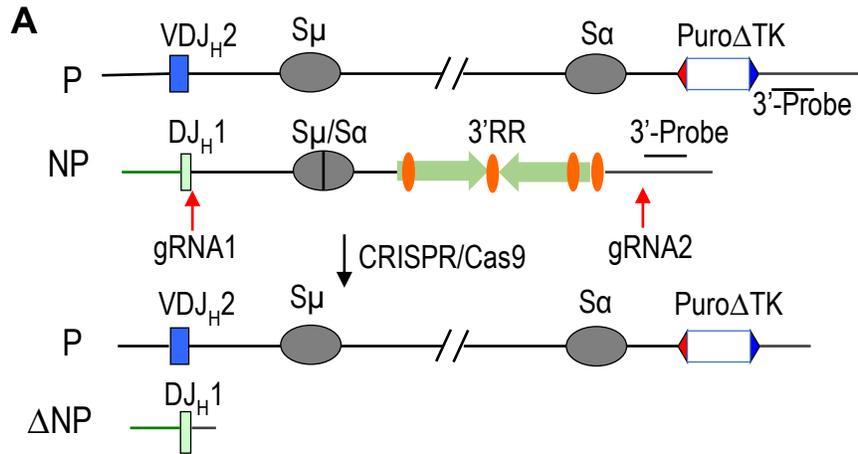
(Figure 8 cont'd.)



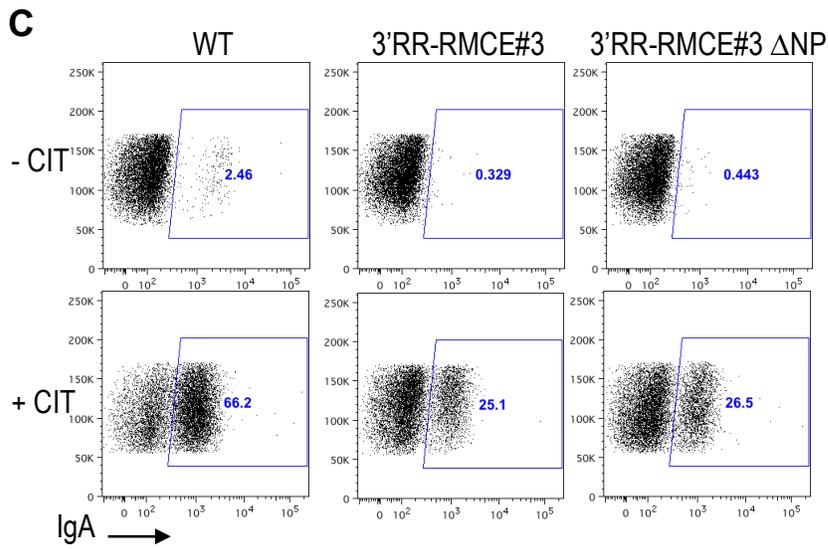
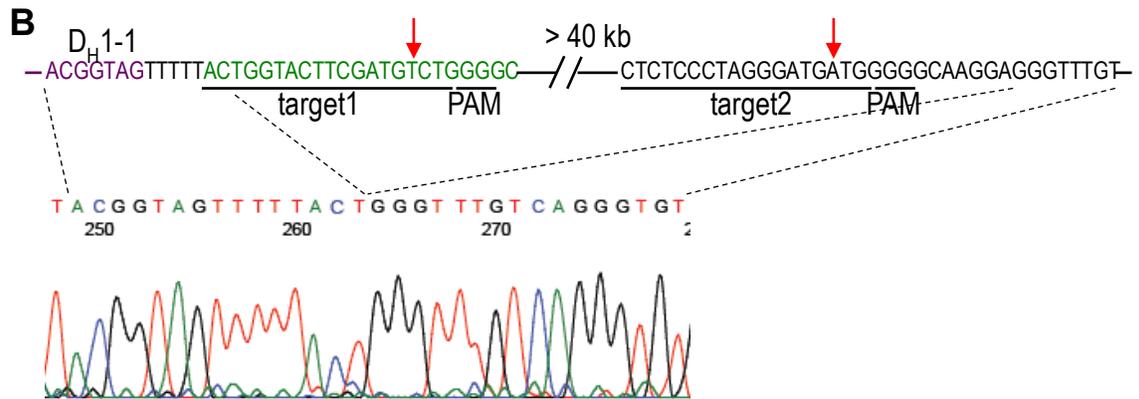
**(Figure 8 cont'd.)**

**A.** RMCE between the chromosome cassette and the empty cassette on the exchange vector results in the removal of the PGK-Puro $\Delta$ TK selection module from the chromosome. **B.** Southern blot analysis of three successfully exchanged clones ( $\Delta$ 3'RR #1, #2, and #3). **C.** A time course of CSR for the RMCE cassette removed clones (dashed lines) over a period of 72 h. B, Bgl II.

**Figure 9. Class Switch Recombination in cells with a deletion of 3'RR on both alleles.**



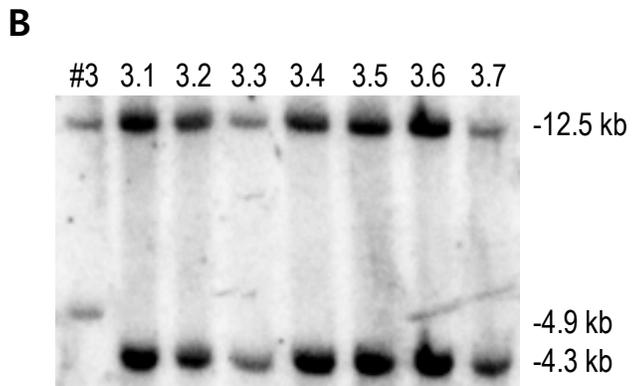
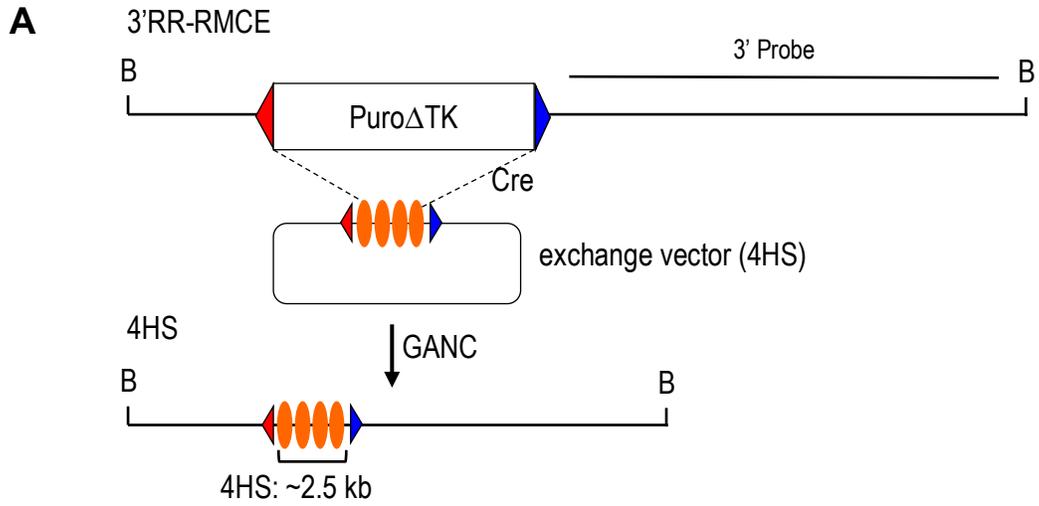
(Figure 9 cont'd)



**(Figure 9 cont'd)**

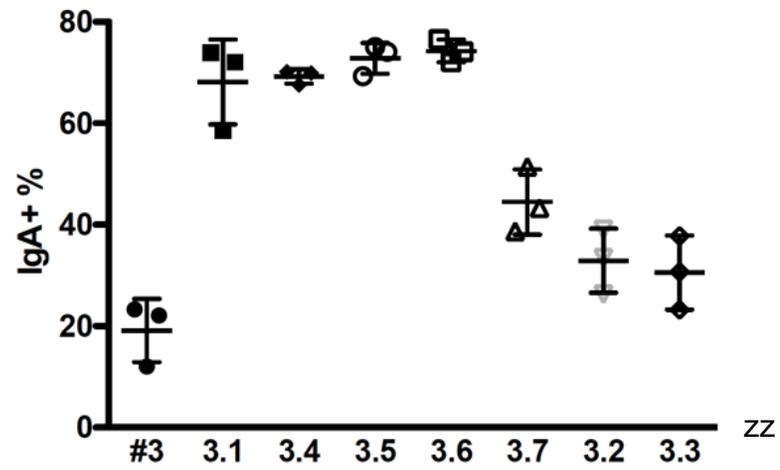
**A.** Deletion of 3'RR on the nonproductive (NP) allele by CRISPR/Cas9-mediated genome editing. The two red arrows indicate Cas9-targeting sites. The region between these two sites is deleted. The vertical dashed line on the Southern blot indicates the position from which two irrelevant lanes were removed. **B.** Sequences of the junction generated by CRISPR/Cas9-mediated deletion on the NP allele. Two red arrows indicate the original cleavage sites by the Cas9 nuclease. End resection is revealed by DNA sequencing of the final junction. **C.** CSR efficiency is unaltered upon the further removal of 3'RR on the productive allele. Boxed areas indicate post-switched IgA<sup>+</sup> populations.

**Figure 10. Knock-in of four DNase I HS sites restores Class Switch Recombination.**

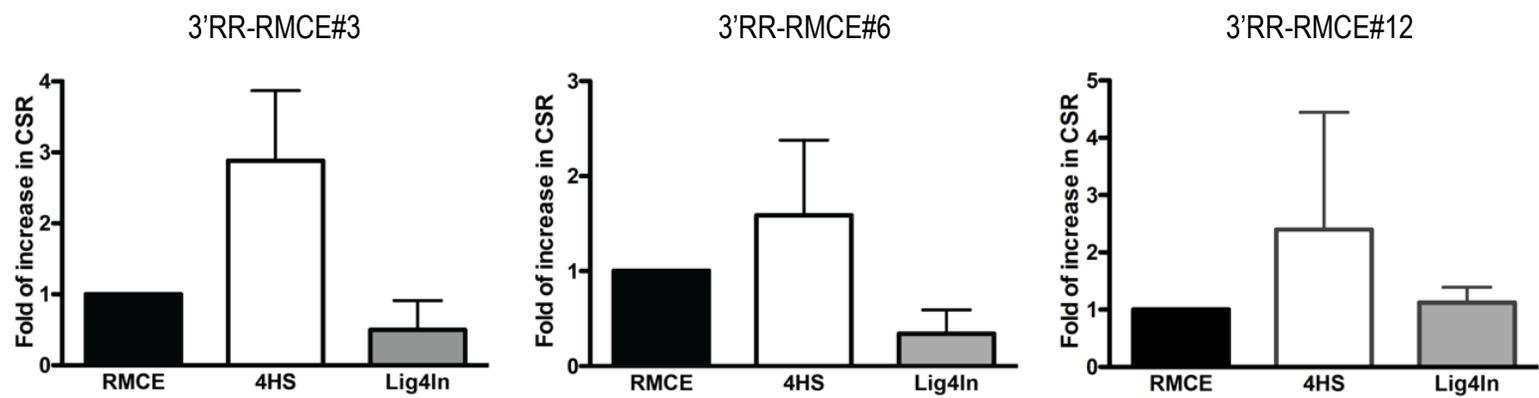


(Figure 10 cont'd)

C

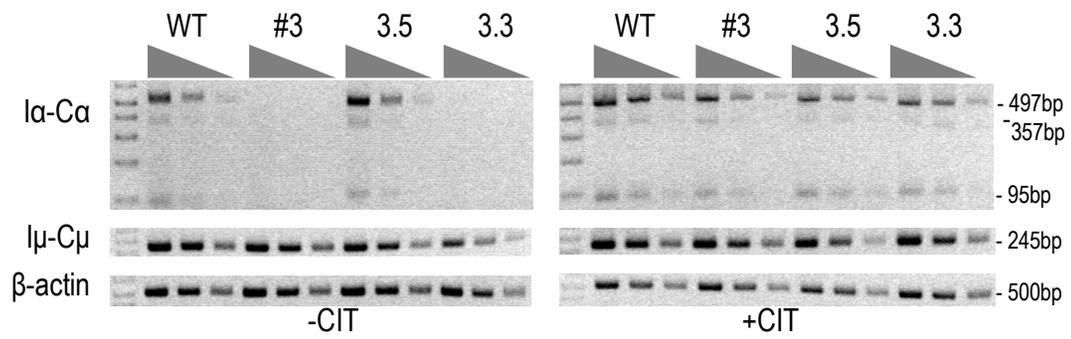


D

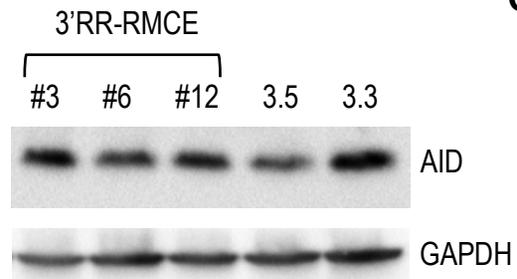


(Figure 10 cont'd)

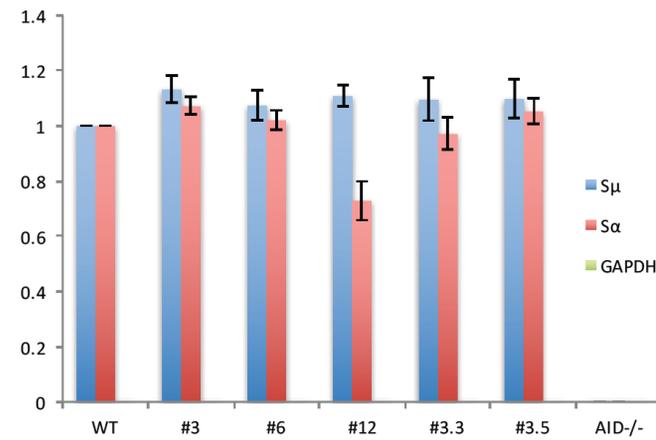
**E**



**F**



**G**

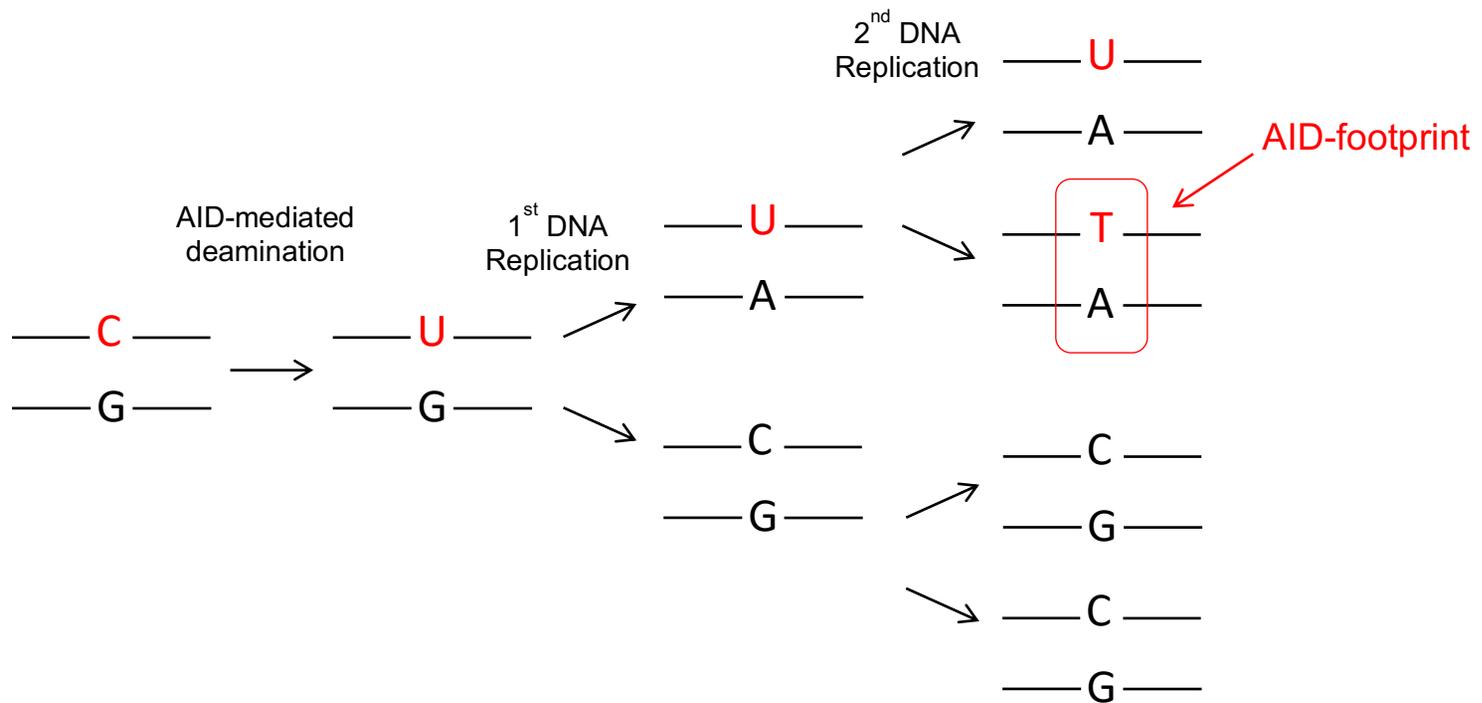


**(Figure 10 cont'd)**

**A.** A fragment consisting of four HS sites juxtaposed in tandem in the natural order was knocked-in by RMCE to replace the deleted 3'RR on the productive allele. **B.** Southern blot analysis of seven successfully exchanged clones. **C.** CSR of the seven clones contains the four HS sites in place of 3'RR. Four of the seven clones display a hyper-CSR (>WT) phenotype. All clones switch more efficiently than the parental 3'RR-RMCE clone (#3). **D.** RMCE with the four HS sites or an irrelevant DNA fragment derived from Lig4 intron (Lig4In) was performed in all three 3'RR-RMCE clones (#3, #6, and #12). **E.** Germline transcription measured by semiquantitative RT-PCR in two representative HS-exchanged clones (clone 3.5 displayed hyper-CSR phenotype, whereas clone 3.3 displayed only a moderate increase in CSR that was still lower than WT). **F.** Western blot analysis of AID expression in 3'RR-RMCE and HS knock-in clones. **G.** AID recruitment to S $\mu$  and S $\alpha$  by ChIP analyses. ChIP signals from WT cells stimulated for 16 h were set to 1. Error bars indicate the SE of three independent experiments.

APPENDIX C: Chapter 3 Figures

**Figure 11. Analysis of AID-footprints.** When uracil-processing elements are disrupted, AID-generated uracil is no longer processed. Upon DNA replication, AID-generated uracils can be identified by C to T (or G to A) transition.



**Figure 12. Occurrence of AID-generated mutation.** Representative chromatogram containing double peaks, an indication of heterogeneous PCR products. Sequence heterogeneity was confirmed by cloning of PCR product followed by Sanger sequencing.

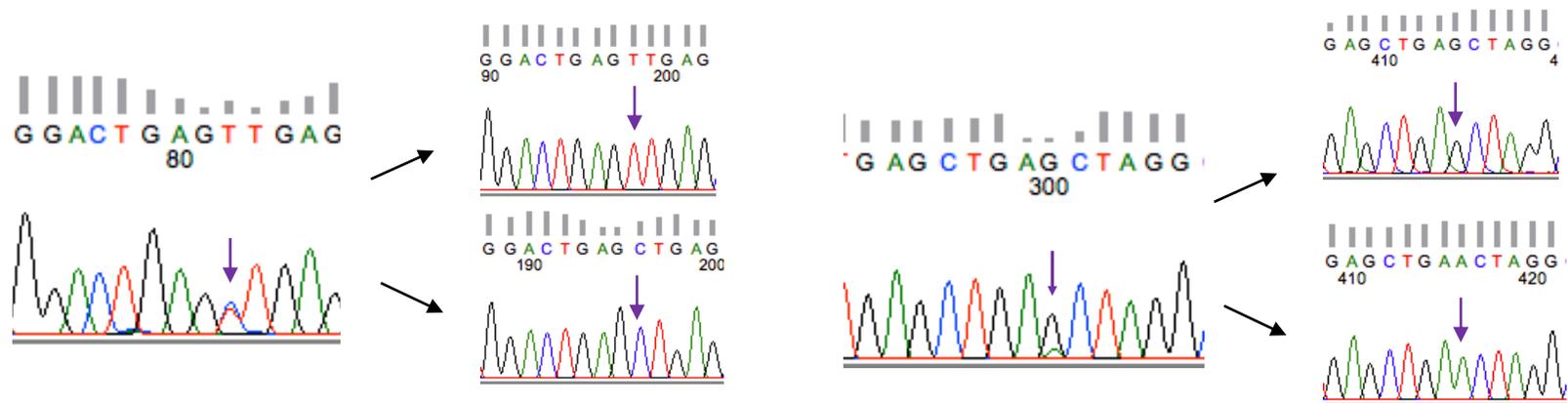
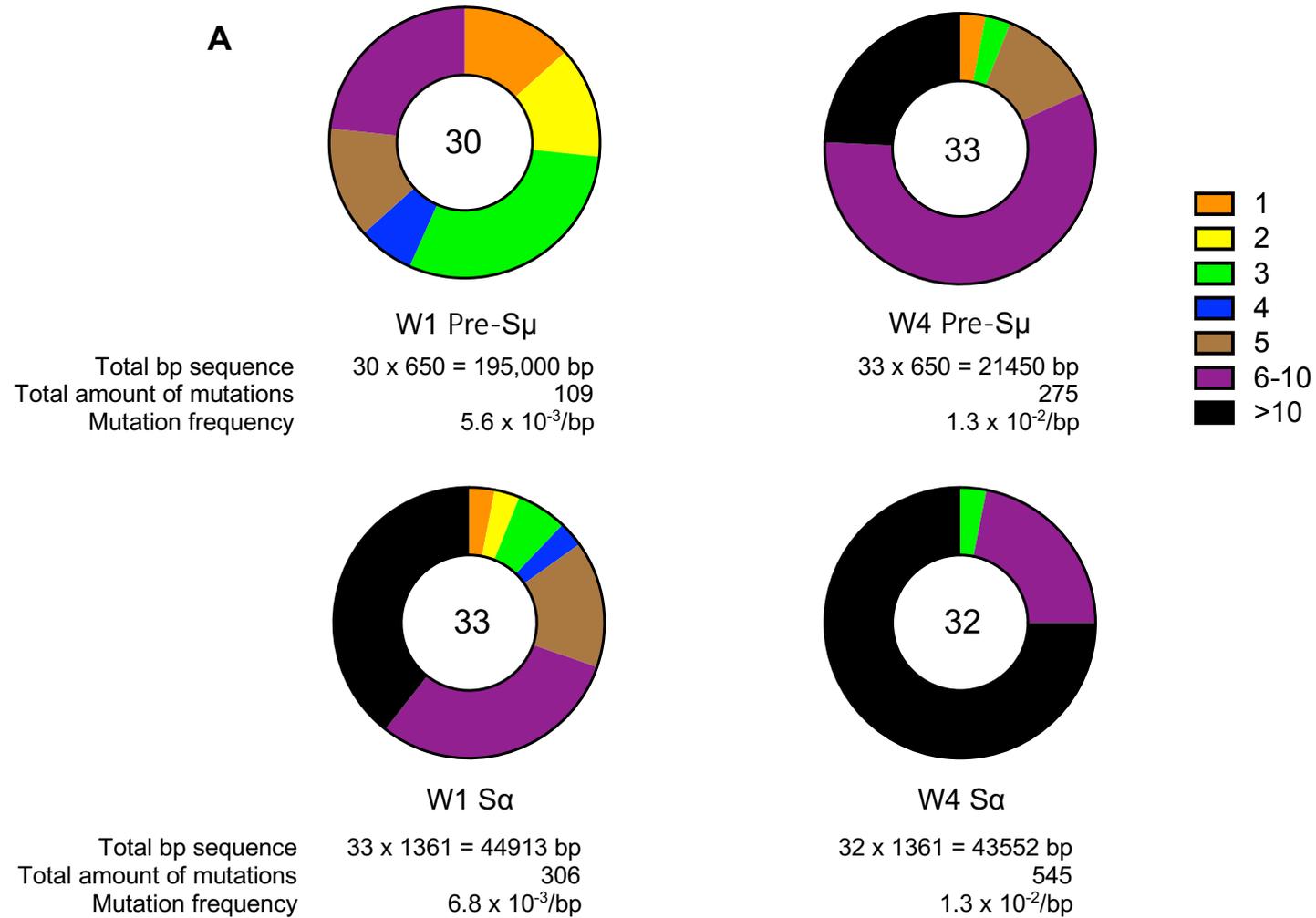
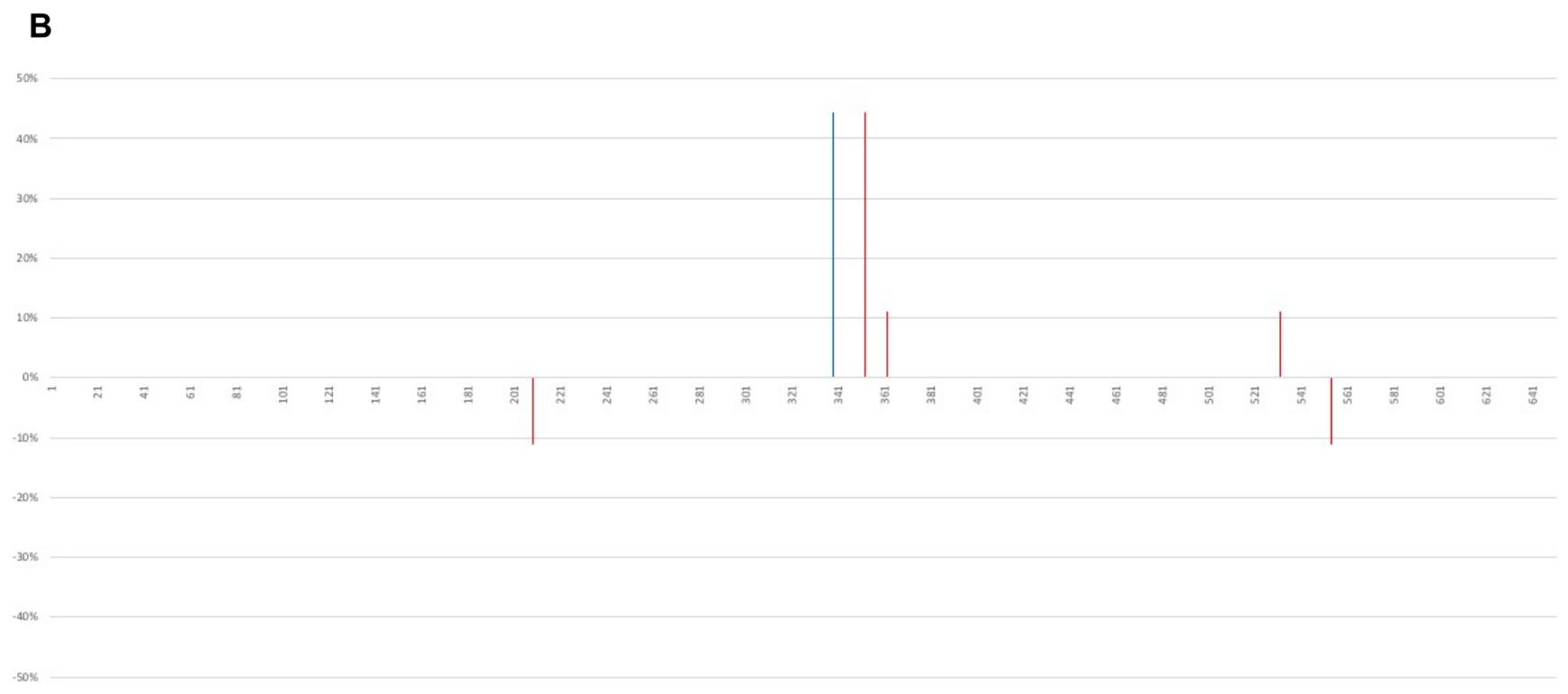


Figure 13. Analysis of AID footprints in pooled populations of S $\alpha$ <sup>core</sup>-DKO cells after 1 or 4 weeks of CIT stimulation.

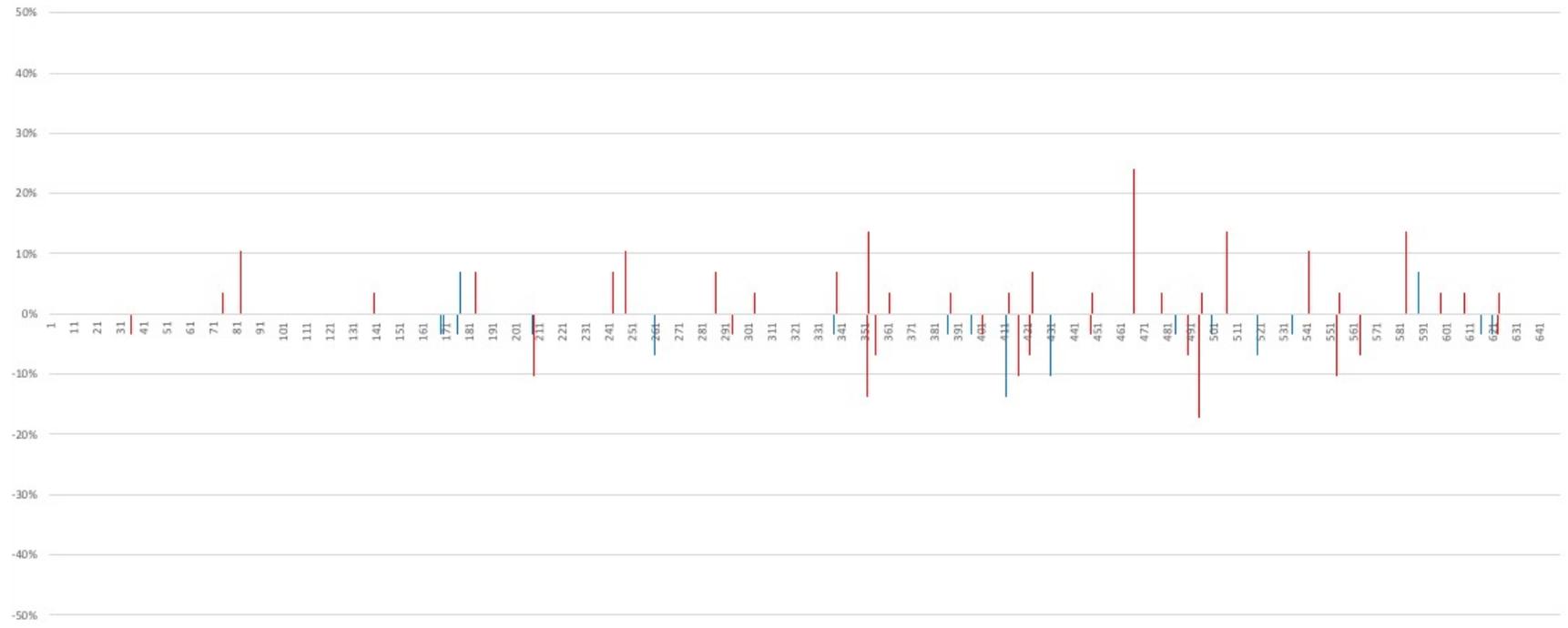


(Figure 13 cont'd.)



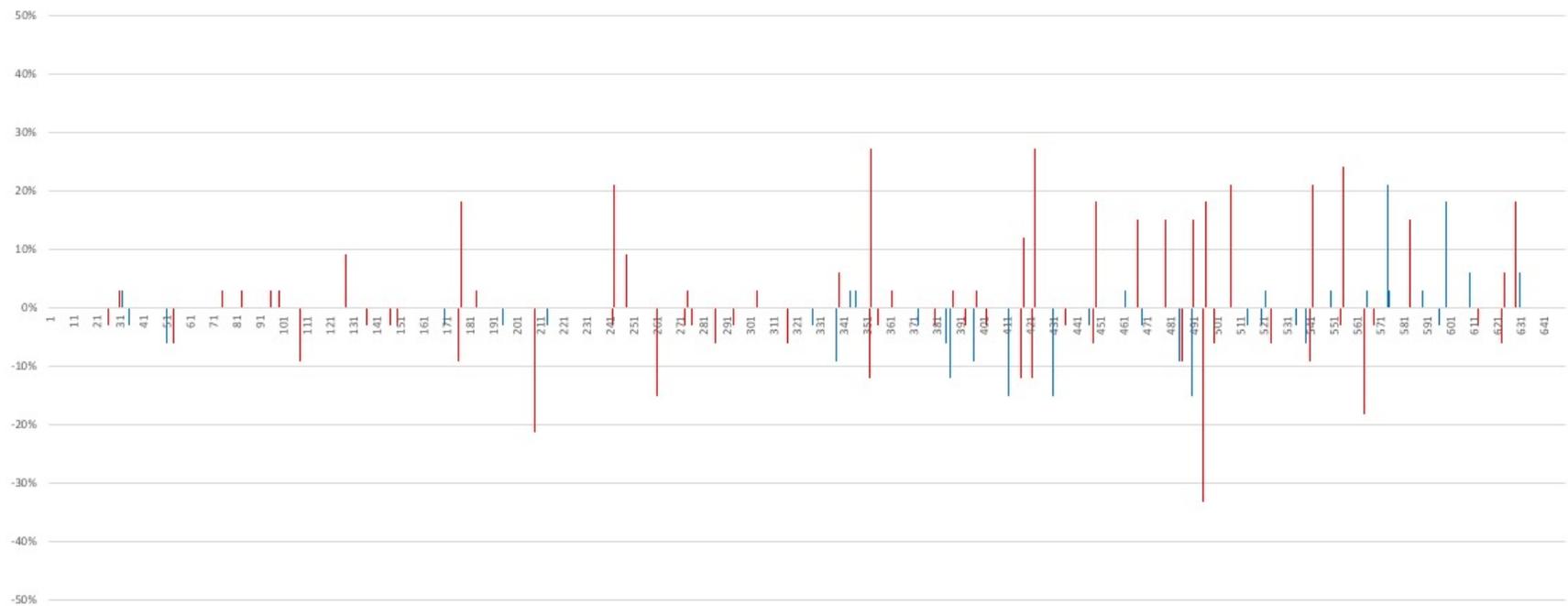
(Figure 13 cont'd.)

C



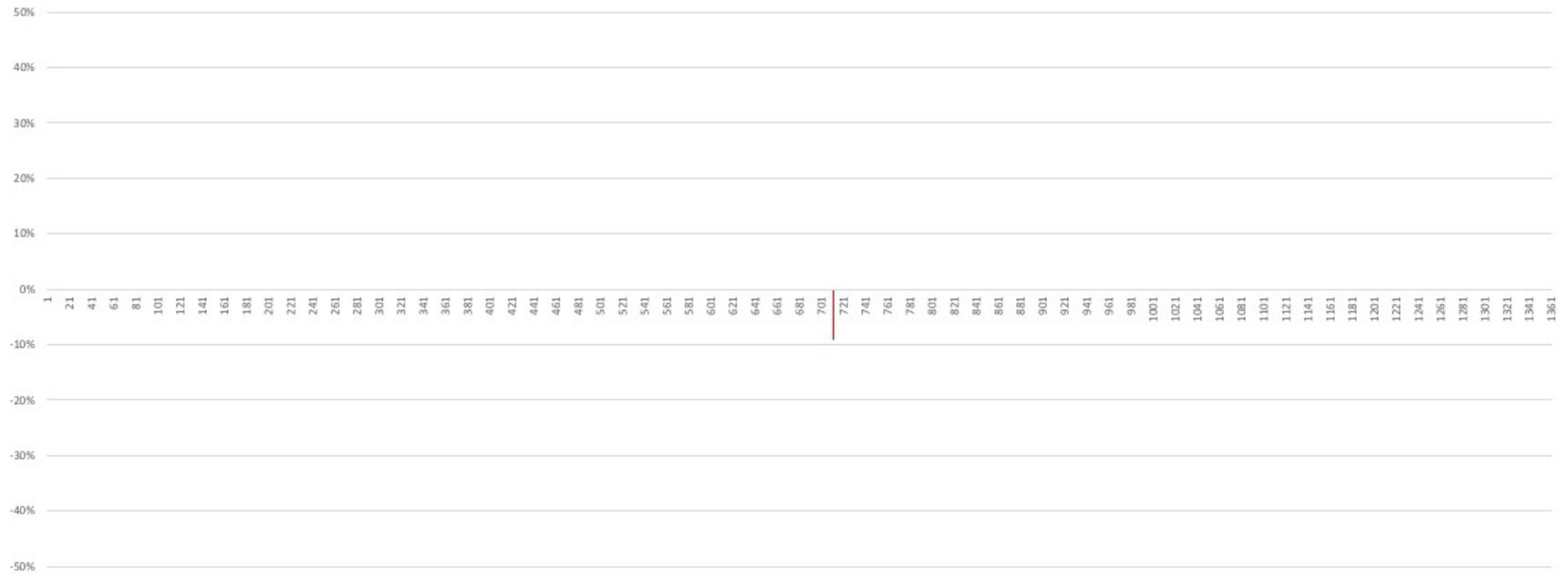
(Figure 13 cont'd.)

D



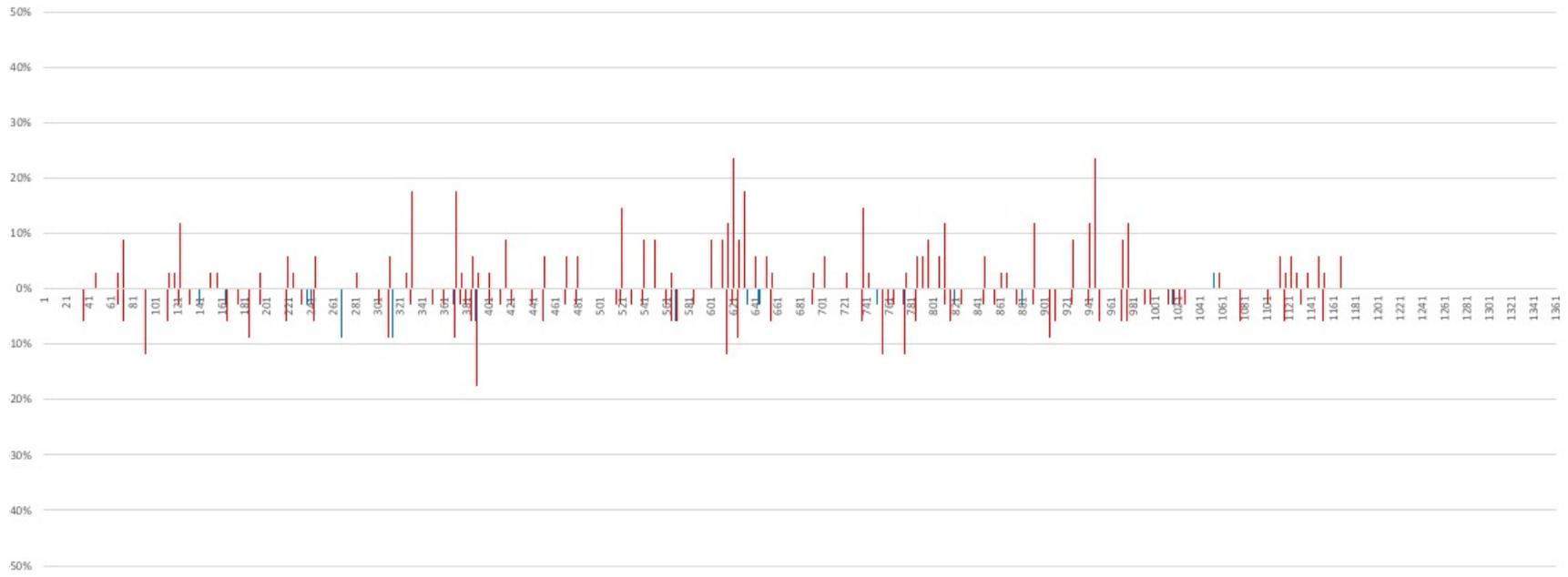
(Figure 13 cont'd.)

**E**



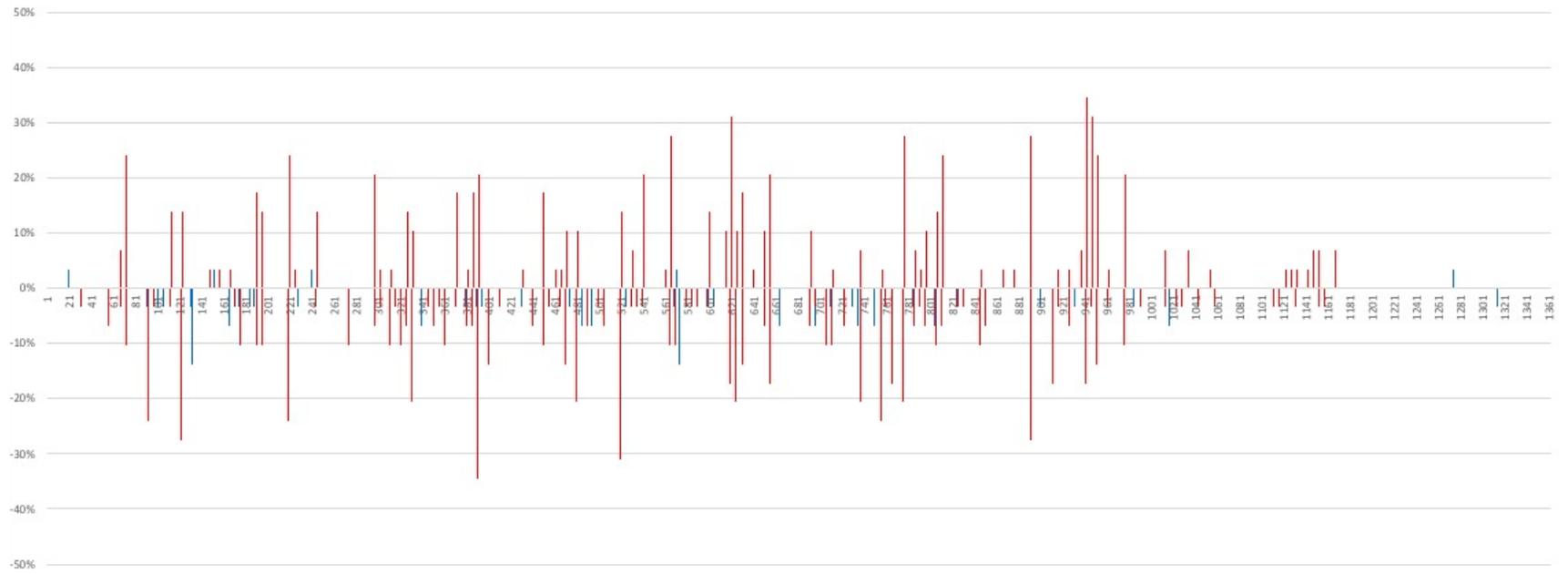
(Figure 13 cont'd.)

F



(Figure 13 cont'd.)

G



**(Figure 13 cont'd.)**

**A.** Mutation load among the Pre-S $\mu$  or S $\alpha$  region sequences from CIT stimulated S $\alpha^{\text{core}}$ -DKO cells. The number in the center of the pie chart is the number of sequences determined, with each slice of the pie indicating the proportion of sequences with 1 to more than 10 mutations. Mutation frequency per base shown under each pie chart. **B-G.** Distribution of AID footprints across Pre-S $\mu$  or S $\alpha$  region sequences from CIT stimulated S $\alpha^{\text{core}}$ -DKO cells at each timepoint. Each bar represents percentage of mutations at that location out of total number of molecules sequenced. Mutations at C shown on the top and mutations at G shown on the bottom of the strand. Red bars represent AID hotspots and blue bars represent AID coldspots. **B.** Pre-S $\mu$  sequences without CIT stimulation **C.** Pre-S $\mu$  sequences at one week of CIT stimulation **D.** Pre-S $\mu$  sequences at four weeks of CIT stimulation **E.** S $\alpha$  sequences without CIT stimulation **F.** S $\alpha$  sequences at one week of CIT stimulation **G.** S $\alpha$  sequences at four weeks of CIT stimulation. CIT, mixture of anti-CD40 Ab, IL-4, and TGF- $\beta$ 1 used to stimulate cells to undergo CSR.

**Figure 14. AID-footprints in Sa region.** Sa region sequences from Sa<sup>core</sup>-DKO cells stimulated with CIT for 24 hours. Each line is a sequence, with mutations at G on bottom strands and mutations at C on top strands. AID hotspots are indicated in red and coldspots indicated in blue.

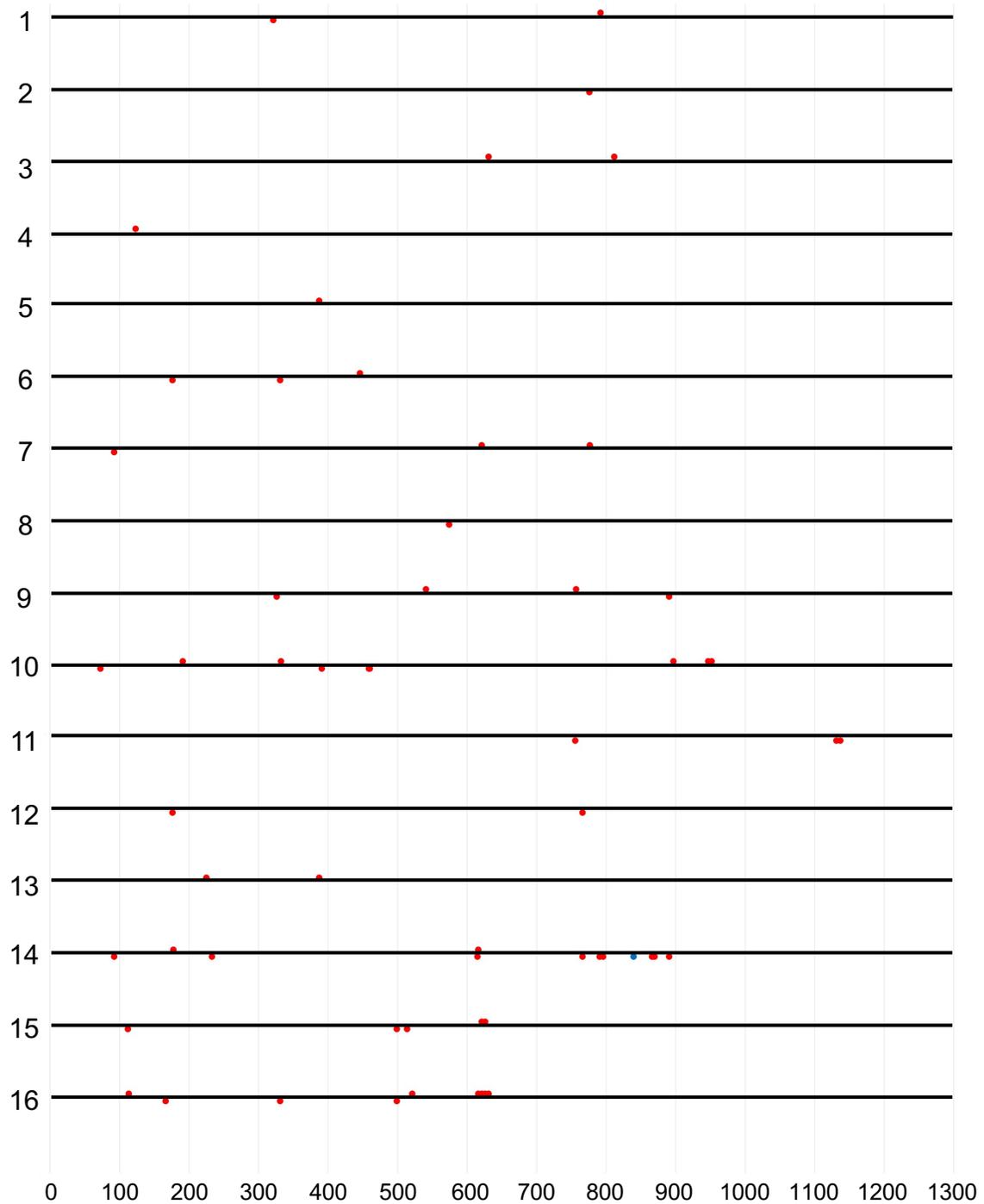
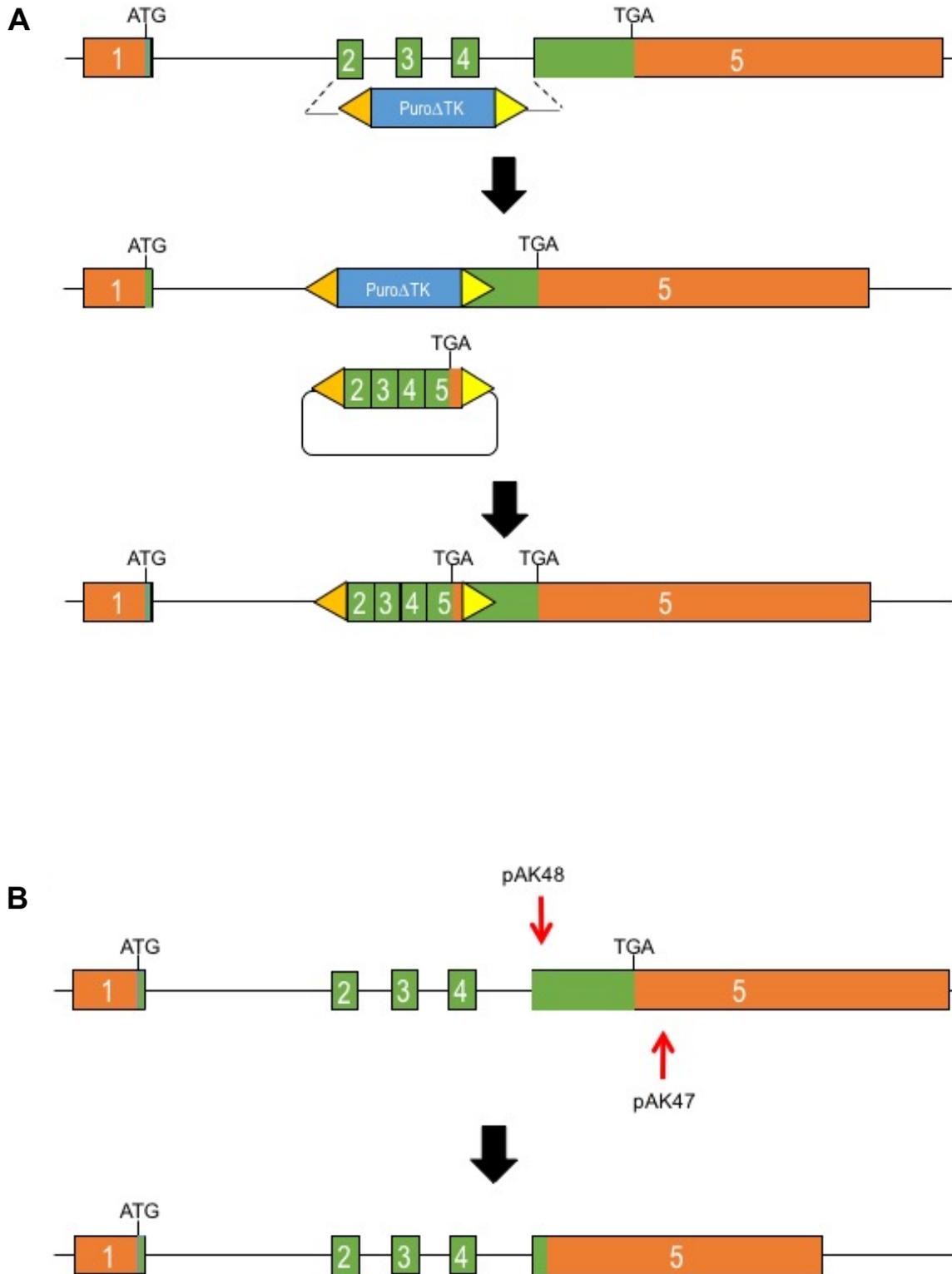
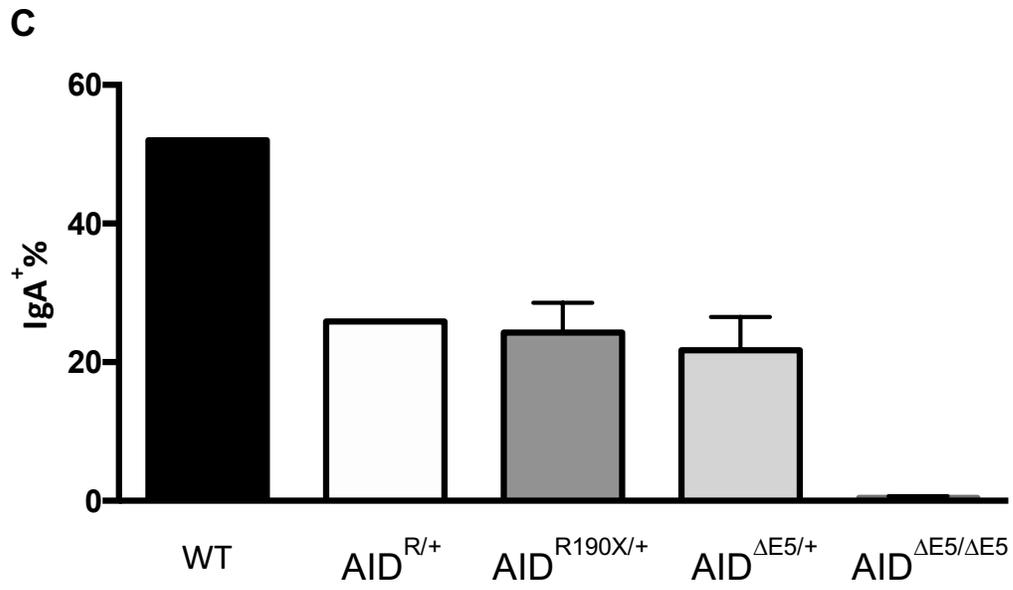


Figure 15. Genome editing at AID locus



(Figure 15 cont'd.)

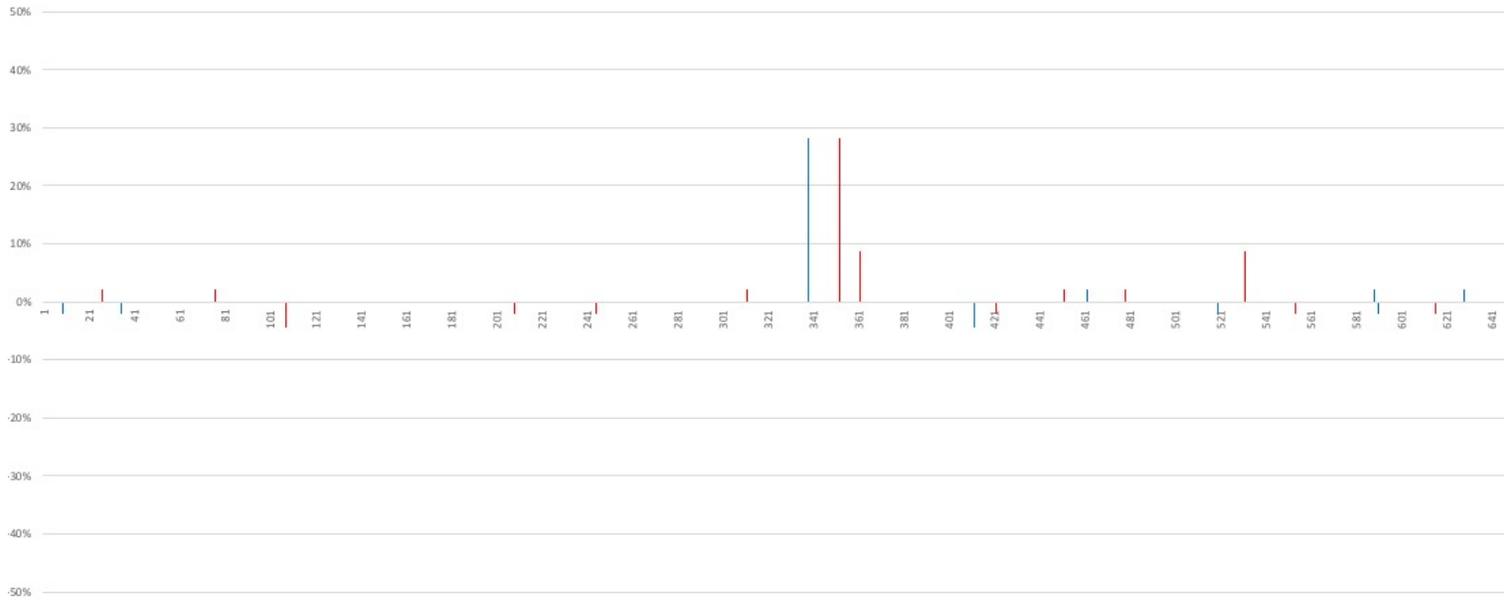


**(Figure 15 cont'd.)**

**A.** Gene targeting to replace AID with a RMCE cassette. Strategy to replace AID with a RMCE cassette. Exchanged plasmid containing R190X mutation sequence and CRE expression plasmid are co-transfected into AID<sup>R/+</sup> cells and selected with ganciclovir. **B.** CRISPR-mediated genome editing at AID locus. pAK48 and pAK47 are gRNA vectors that target Cas9 nickase to Exon 5 of AID locus which will generate  $\Delta E5$ . Coding regions are indicated in red. **C.** CSR efficiency of mutant AID cells. CSR efficiency of multiple clones containing AID<sup>R190X/+</sup>, AID <sup>$\Delta E5$ /+</sup> and AID <sup>$\Delta E5/\Delta E5$</sup> . CSR efficiency is determined as the percentage of IgA<sup>+</sup> cells after 72hrs of stimulation with CIT. CIT, mixture of anti-CD40 Ab, IL-4, and TGF- $\beta$ 1 used to stimulate cells to undergo CSR.

**Figure 16. Analysis of AID footprints in pooled populations of AID<sup>E5/ΔE5</sup>Sα<sup>core</sup>-DKO cells after 1 week of CIT stimulation**

**A**



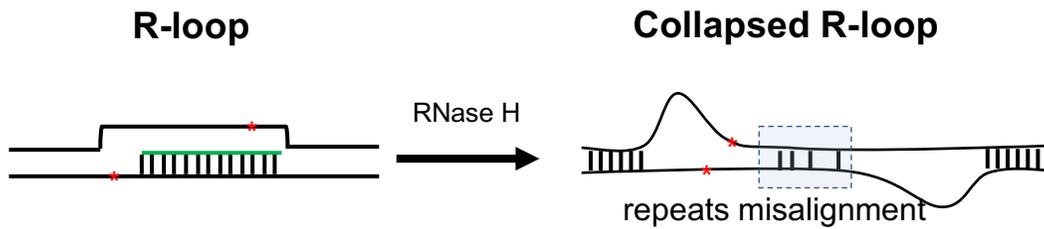
# B



**(Figure 16 cont'd.)**

**A-B.** Distribution of AID footprints across 5'S $\mu$  or S $\alpha$  region sequences from AID<sup>E5/ $\Delta$ E5</sup>S $\alpha$ <sup>core</sup>-DKO cells stimulated with CIT for 1 week. Each bar represents percentage of mutations at that location out of total number of molecules sequenced. Mutations at C shown on the top and mutations at G shown on the bottom of the strand. Red bars represent AID hotspots and blue bars represent AID coldspots. **A.** 5'S $\mu$  sequences at one week of CIT stimulation **D.** 5'S $\alpha$  sequences at one week of CIT stimulation. CIT, mixture of anti-CD40 Ab, IL-4, and TGF- $\beta$ 1 used to stimulate cells to undergo CSR.

**Figure 17. Model of DSB formation by distally position AID-generated uracils in a collapsed R-loop structure.** An example of distally positioned AID-generated uracils shown in red stars (\*). In R-loop, DNA strands are linear where uracils are positioned far from one another. Upon RNase H treatment, S region repeats misalign to form collapsed R-loop which can potentially bring the two uracils to proximity, thus leading to DSB formation.



## **BIBLIOGRAPHY**

## BIBLIOGRAPHY

- [1] J. Chaudhuri and F. W. Alt, "Class-switch recombination: interplay of transcription, DNA deamination and DNA repair.," *Nat. Rev. Immunol.*, vol. 4, no. 7, pp. 541–52, Jul. 2004.
- [2] K. Yu and M. R. Lieber, "Nucleic acid structures and enzymes in the immunoglobulin class switch recombination mechanism," *DNA Repair (Amst.)*, vol. 2, no. 11, pp. 1163–1174, 2003.
- [3] J. Stavnezer, J. E. J. Guikema, and C. E. Schrader, "Mechanism and Regulation of Class Switch Recombination," *Annu. Rev. Immunol.*, vol. 26, no. 1, pp. 261–292, 2008.
- [4] E. Pinaud *et al.*, *The IgH locus 3' regulatory region: Pulling the strings from behind*, vol. 110. 2011.
- [5] S. P. Methot and J. M. Di Noia, *Molecular Mechanisms of Somatic Hypermutation and Class Switch Recombination*, 1st ed., vol. 133. Elsevier Inc., 2017.
- [6] C. Gritzmacher, "Molecular aspects of heavy-chain class switching," *Crit. Rev. Immunol.*, vol. 9, no. 3, pp. 173–200, 1989.
- [7] T. Kataoka, T. Miyata, and T. Honjo, "Repetitive Sequences in Class-Switch Recombination Regions of Immunoglobulin Heavy Chain Genes," *Cell*, vol. 23, no. February, pp. 357–368, 1981.
- [8] F. Huang *et al.*, "Sequence Dependence of Chromosomal R-Loops at the Immunoglobulin Heavy-Chain S  $\mu$  Class Switch Region Sequence Dependence of Chromosomal R-Loops at the Immunoglobulin Heavy-Chain S  $\kappa$  Class Switch Region  $\square$   $\dagger$ ," vol. 27, no. 16, pp. 5921–5932, 2007.
- [9] R. Shinkura, M. Tian, M. Smith, K. Chua, Y. Fujiwara, and F. W. Alt, "The influence of transcriptional orientation on endogenous switch region function," *Nat. Immunol.*, vol. 4, no. 5, pp. 435–441, 2003.
- [10] A. A. Khamlichi, F. Glaudet, Z. Oruc, V. Denis, M. Le Bert, and M. Cogné, "Immunoglobulin class-switch recombination in mice devoid of any S $\mu$  tandem repeat," *Blood*, 2004.
- [11] A. A. Zarrin *et al.*, "An evolutionarily conserved target motif for immunoglobulin class-switch recombination," *Nat. Immunol.*, vol. 5, no. 12, pp. 1275–1281, 2004.
- [12] R. Shinkura, M. Tian, M. Smith, K. Chua, Y. Fujiwara, and F. W. Alt, "The

influence of transcriptional orientation on endogenous switch region function.," *Nat. Immunol.*, vol. 4, no. 5, pp. 435–41, 2003.

- [13] J. A. Hackney *et al.*, *Chapter 5 DNA Targets of AID. Evolutionary Link Between Antibody Somatic Hypermutation and Class Switch Recombination*, 1st ed., vol. 101, no. C. Elsevier Inc., 2009.
- [14] J. Stavnezer and C. T. Amemiya, "Evolution of isotype switching," *Semin. Immunol.*, vol. 16, no. 4, pp. 257–275, 2004.
- [15] M. Gellert, "V(D)J Recombination: RAG Proteins, Repair Factors, and Regulation," *Annu. Rev. Biochem.*, 2002.
- [16] W. Dunnick, G. Z. Hertz, L. Scappino, and C. Gritzmacher, "DNA sequences at immunoglobulin switch region recombination sites DEFINING SWITCH RECOMBINATION SITES VERSUS," vol. 21, no. 3, pp. 365–372, 1993.
- [17] M. Muramatsu, K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai, and T. Honjo, "Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme Figure 1. Induced Expression of AID in CH12F3-2 Cells," *Cell*, vol. 102, pp. 553–563, 2000.
- [18] H. Arakawa, "Requirement of the Activation-Induced Deaminase (AID) Gene for Immunoglobulin Gene Conversion," *Science (80-. )*, vol. 295, no. 5558, pp. 1301–1306, 2002.
- [19] P. Revy *et al.*, "Activation-Induced Cytidine Deaminase (AID) Deficiency Causes the Autosomal Recessive Form of the Hyper-IgM Syndrome (HIGM2)," *Cell*, vol. 102, no. 5, pp. 565–575, 2000.
- [20] M. Muramatsu, "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.*, vol. 274, no. 26, pp. 18470–18476, 1999.
- [21] A. Jarmuz *et al.*, "An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22," *Genomics*, 2002.
- [22] Q. Pan-Hammarström, Y. Zhao, and L. Hammarström, "Class Switch Recombination: A Comparison Between Mouse and Human," *Advances in Immunology*. 2007.
- [23] S. K. Petersen-Mahrt, R. S. Harris, and M. S. Neuberger, "AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification," *Nature*, vol. 418, no. 6893, pp. 99–104, 2002.

- [24] G. S. Shapiro and L. J. Wysocki, "DNA Target Motifs of Somatic Mutagenesis in Antibody Genes," *Crit. Rev. Immunol.*, 2012.
- [25] K. Yu, F. T. Huang, and M. R. Lieber, "DNA Substrate Length and Surrounding Sequence Affect the Activation-induced Deaminase Activity at Cytidine," *J. Biol. Chem.*, vol. 279, no. 8, pp. 6496–6500, 2004.
- [26] L. Han, S. Masani, and K. Yu, "Overlapping activation-induced cytidine deaminase hotspot motifs in Ig class-switch recombination.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 28, pp. 11584–9, 2011.
- [27] M. Liu *et al.*, "Two levels of protection for the B cell genome during somatic hypermutation," *Nature*, vol. 451, no. 7180, pp. 841–845, 2008.
- [28] L. Pasqualucci *et al.*, "Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas," *Nature*, 2001.
- [29] D. F. Robbiani *et al.*, "AID Is Required for the Chromosomal Breaks in c-myc that Lead to c-myc/IgH Translocations," *Cell*, vol. 135, no. 6, pp. 1028–1038, 2008.
- [30] H. M. Shen, A. Peters, B. Baron, X. Zhu, and U. Storb, "Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes," *Science (80-. )*, 1998.
- [31] S. Kracker and A. Durandy, "Insights into the B cell specific process of immunoglobulin class switch recombination," *Immunol. Lett.*, vol. 138, no. 2, pp. 97–103, 2011.
- [32] E. J. Pone *et al.*, "BCR-signalling synergizes with TLR-signalling for induction of AID and immunoglobulin class-switching through the non-canonical NF- $\kappa$ B pathway," *Nat. Commun.*, 2012.
- [33] G. Teng *et al.*, "MicroRNA-155 Is a Negative Regulator of Activation-Induced Cytidine Deaminase," *Immunity*, vol. 28, no. 5, pp. 621–629, 2008.
- [34] Y. Dorsett *et al.*, "MicroRNA-155 Suppresses Activation-Induced Cytidine Deaminase-Mediated Myc-Igh Translocation," *Immunity*, vol. 28, no. 5, pp. 630–638, 2008.
- [35] V. G. de Yébenes *et al.*, "miR-181b negatively regulates activation-induced cytidine deaminase in B cells," *J. Exp. Med.*, vol. 205, no. 10, pp. 2199–2206, 2008.
- [36] G. M. Borchert, N. W. Holton, and E. D. Larson, "Repression of human activation induced cytidine deaminase by miR-93 and miR-155," *BMC Cancer*, 2011.

- [37] K. Basso *et al.*, “BCL6 positively regulates AID and germinal center gene expression via repression of miR-155,” *J. Exp. Med.*, 2012.
- [38] Z. Xu, H. Zan, E. J. Pone, T. Mai, and P. Casali, “Immunoglobulin class-switch DNA recombination: Induction, targeting and beyond,” *Nature Reviews Immunology*. 2012.
- [39] J. Stavnezer and C. E. Schrader, “IgH Chain Class Switch Recombination: Mechanism and Regulation,” *J. Immunol.*, vol. 193, no. 5, pp. 2040–2040, 2014.
- [40] J. Stavnezer, “Immunoglobulin class switching,” *Current Opinion in Immunology*, vol. 8, no. 2. pp. 199–205, 1996.
- [41] S. Jung, K. Rajewsky, and A. Radbruch, “Shutdown of class switch recombination by deletion of a switch region control element,” *Science (80-. )*, 1993.
- [42] M. Lorenz, S. Jung, and a Radbruch, “Switch transcripts in immunoglobulin class switching.,” *Science*, vol. 267, pp. 1825–8, 1995.
- [43] K. Hein, M. G. Lorenz, G. Siebenkotten, K. Petry, R. Christine, and a Radbruch, “Processing of switch transcripts is required for targeting of antibody class switch recombination.,” *J. Exp. Med.*, vol. 188, no. 12, pp. 2369–2374, 1998.
- [44] A. Bottaro, R. Lansford, L. Xu, J. Zhang, P. Rothman, and F. W. Alt, “S region transcription per se promotes basal IgE class switch recombination but additional factors regulate the efficiency of the process.,” *EMBO J.*, vol. 13, no. 3, pp. 665–74, 1994.
- [45] G. Qiu, G. R. Harriman, and J. Stavnezer, “I( $\alpha$ ) exon-replacement mice synthesize a spliced HPRT-C( $\alpha$ ) transcript which may explain their ability to switch to IgA. Inhibition of switching to IgG in these mice,” *Int. Immunol.*, vol. 11, no. 1, pp. 37–46, 1999.
- [46] K. J. Seidl, A. Bottaro, A. Vo, J. Zhang, L. Davidson, and F. W. Alt, “An expressed neo(r) cassette provides required functions of the I( $\gamma$ )2b exon for class switching,” *Int. Immunol.*, vol. 10, no. 11, pp. 1683–1692, 1998.
- [47] L. Xu, B. Gorham, S. C. Li, A. Bottaro, and F. W. Alt, “Replacement of germ-line e promoter by gene targeting alters control of immunoglobulin heavy chain class switching,” *Proc. Natl. Acad. Sci.*, vol. 90, no. April, pp. 3705–3709, 1993.
- [48] J. Zhang, A. Bottaro, S. Li, V. Stewart, and F. W. Alt, “A selective defect in IgG2b switching as a result of targeted mutation of the I gamma 2b promoter and exon.,” *EMBO J.*, vol. 12, no. 9, pp. 3529–37, 1993.

- [49] J. P. Manis *et al.*, "Class switching in B cells lacking 3' immunoglobulin heavy chain enhancers," *J. Exp. Med.*, vol. 188, no. 8, 1998.
- [50] W. A. Dunnick, J. Shi, K. A. Graves, and J. T. Collins, "The 3' end of the heavy chain constant region locus enhances germline transcription and switch recombination of the four gamma genes.," *J. Exp. Med.*, vol. 201, no. 9, pp. 1459–66, 2005.
- [51] M. Cogné *et al.*, "A class switch control region at the 3' end of the immunoglobulin heavy chain locus.," *Cell*, vol. 77, no. 5, pp. 737–747, 1994.
- [52] C. Vincent-Fabert *et al.*, "Genomic deletion of the whole IgH 3' regulatory region (hs3a, hs1,2, hs3b, and hs4) dramatically affects class switch recombination and Ig secretion to all isotypes," *Blood*, vol. 116, no. 11, pp. 1895–1898, 2010.
- [53] K. Yu, F. Chedin, C.-L. Hsieh, T. E. Wilson, and M. R. Lieber, "R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells," *Nat. Immunol.*, vol. 4, no. 5, pp. 442–451, 2003.
- [54] D. Roy, K. Yu, and M. R. Lieber, "Mechanism of R-loop formation at immunoglobulin class switch sequences.," *Mol. Cell. Biol.*, vol. 28, no. 1, pp. 50–60, 2008.
- [55] G. A. Daniels and M. R. Lieber, "RNA: DNA complex formation upon transcription of immunoglobulin switch regions: Implications for the mechanism and regulation of class switch recombination," *Nucleic Acids Res.*, 1995.
- [56] R. Pavri *et al.*, "Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5," *Cell*, vol. 143, no. 1, pp. 122–133, 2010.
- [57] J. A. Daniel *et al.*, "PTIP Promotes Chromatin Changes Critical for Immunoglobulin Class Switch Recombination," *Science (80-. )*, vol. 329, no. 5994, pp. 917–923, 2010.
- [58] U. Basu *et al.*, "The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates," *Cell*, vol. 144, no. 3, pp. 353–363, 2011.
- [59] A. Stanlie, M. Aida, M. Muramatsu, T. Honjo, and N. A. Begum, "Histone3 lysine4 trimethylation regulated by the facilitates chromatin transcription complex is critical for DNA cleavage in class switch recombination," *Proc. Natl. Acad. Sci.*, 2010.
- [60] Q. Qiao, L. Wang, F. L. Meng, J. K. Hwang, F. W. Alt, and H. Wu, "AID

Recognizes Structured DNA for Class Switch Recombination,” *Mol. Cell*, vol. 67, no. 3, pp. 361-373.e4, 2017.

- [61] C. Rada, G. T. Williams, H. Nilsen, D. E. Barnes, T. Lindahl, and M. S. Neuberger, “Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice,” *Curr. Biol.*, vol. 12, no. 20, pp. 1748–1755, 2002.
- [62] K. Imai *et al.*, “Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination,” *Nat Immunol*, vol. 4, no. 10, pp. 1023–1028, 2003.
- [63] C. Rada, J. M. Di Noia, and M. S. Neuberger, “Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation,” *Mol. Cell*, vol. 16, no. 2, pp. 163–171, 2004.
- [64] F. A. Dingler, K. Kemmerich, M. S. Neuberger, and C. Rada, “Uracil excision by endogenous SMUG1 glycosylase promotes efficient Ig class switching and impacts on A:T substitutions during somatic mutation,” *Eur. J. Immunol.*, vol. 44, no. 7, pp. 1925–1935, 2014.
- [65] S. Masani, L. Han, and K. Yu, “Apurinic/aprimidinic endonuclease 1 is the essential nuclease during immunoglobulin class switch recombination.,” *Mol. Cell. Biol.*, vol. 33, no. 7, pp. 1468–73, Apr. 2013.
- [66] D. L. Ludwig *et al.*, “A murine AP-endonuclease gene-targeted deficiency with post-implantation embryonic progression and ionizing radiation sensitivity,” *Mutat. Res. - DNA Repair*, 1998.
- [67] S. Xanthoudakis, G. G. Miao, and T. Curran, “The redox and DNA-repair activities of Ref-1 are encoded by nonoverlapping domains.,” *Proc. Natl. Acad. Sci.*, 2006.
- [68] H. Fung and B. Dimple, “A vital role for Ape1/Ref1 protein in repairing spontaneous DNA damage in human cells,” *Mol. Cell*, 2005.
- [69] M. Z. Hadi and D. M. Wilson, “Second human protein with homology to the escherichia coli abasic endonuclease exonuclease III,” *Environ. Mol. Mutagen.*, 2000.
- [70] J. E. J. Guikema *et al.*, “APE1- and APE2-dependent DNA breaks in immunoglobulin class switch recombination,” *J. Exp. Med.*, vol. 204, no. 12, pp. 3017–3026, 2007.
- [71] J. E. J. Guikema, J. Stavnezer, and C. E. Schrader, “The role of Apex2 in class-switch recombination of immunoglobulin genes,” *International Immunology*. 2010.

- [72] M. R. Ehrenstein and M. S. Neuberger, "Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: Parallels with somatic hypermutation," *EMBO J.*, vol. 18, no. 12, pp. 3484–3490, 1999.
- [73] C. E. Schrader, W. Edelmann, R. Kucherlapati, and J. Stavnezer, "Reduced isotype switching in splenic B cells from mice deficient in mismatch repair enzymes.," *J. Exp. Med.*, vol. 190, no. 3, pp. 323–30, 1999.
- [74] K. Xue, C. Rada, and M. S. Neuberger, "The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in msh2<sup>-/-</sup> ung<sup>-/-</sup> mice.," *J. Exp. Med.*, vol. 203, no. 9, pp. 2085–94, 2006.
- [75] I. M. Ward *et al.*, "53BP1 is required for class switch recombination," *J. Cell Biol.*, vol. 165, no. 4, pp. 459–464, 2004.
- [76] B. Reina-San-Martin, S. Difilippantonio, L. Hanitsch, R. F. Masilamani, A. Nussenzweig, and M. C. Nussenzweig, "H2AX is required for recombination between immunoglobulin switch regions but not for intra-switch region recombination or somatic hypermutation.," *J. Exp. Med.*, vol. 197, no. 12, pp. 1767–1778, 2003.
- [77] C. Boboila *et al.*, "Alternative end-joining catalyzes robust IgH locus deletions and translocations in the combined absence of ligase 4 and Ku70.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 7, pp. 3034–9, 2010.
- [78] L. Han and K. Yu, "Altered kinetics of nonhomologous end joining and class switch recombination in ligase IV-deficient B cells.," *J. Exp. Med.*, vol. 205, no. 12, pp. 2745–2753, 2008.
- [79] C. Boboila *et al.*, "Robust chromosomal DNA repair via alternative end-joining in the absence of X-ray repair cross-complementing protein 1 (XRCC1)," *Proc. Natl. Acad. Sci.*, vol. 109, no. 7, pp. 2473–2478, 2012.
- [80] M. Dinkelman *et al.*, "Multiple functions of MRN in end-joining pathways during isotype class switching.," *Nat. Struct. Mol. Biol.*, vol. 16, no. 8, pp. 808–13, 2009.
- [81] M. Lee-Theilen, A. J. Matthews, D. Kelly, S. Zheng, and J. Chaudhuri, "CtIP promotes microhomology-mediated alternative end joining during class-switch recombination," *Nat. Struct. Mol. Biol.*, vol. 18, no. 1, pp. 75–79, 2011.
- [82] I. Robert, F. Dantzer, and B. Reina-San-Martin, "Parp1 facilitates alternative NHEJ, whereas Parp2 suppresses IgH/c-myc translocations during immunoglobulin class switch recombination," *J. Exp. Med.*, vol. 206, no. 5, pp. 1047–1056, 2009.

- [83] A. Xie, A. Kwok, and R. Scully, "Role of mammalian Mre11 in classical and alternative nonhomologous end joining," *Nat. Struct. Mol. Biol.*, 2009.
- [84] L. W. Arnold, N. J. LoCascio, P. M. Lutz, C. a Pennell, D. Klapper, and G. Haughton, "Antigen-induced lymphomagenesis: identification of a murine B cell lymphoma with known antigen specificity.," *J. Immunol.*, 1983.
- [85] D. Kunitomo, G. Harriman, and W. Strober, "Regulation of IgA differentiation in CH12LX B cells by lymphokines. IL-4 induces membrane IgM-positive CH12LX cells to express membrane IgA and IL-5 induces membrane IgA-positive CH12LX cells to secrete IgA.," *J. Immunol.*, vol. 141, no. 3, pp. 713–720, 1988.
- [86] M. Nakamura, S. Kondo, M. Sugai, M. Nazarea, S. Imamura, and T. Honjo, "High frequency class switching of an IgM+ B lymphoma clone CH12F3 to IgA+ cells," *Int. Immunol.*, vol. 8, no. 2, pp. 193–201, 1996.
- [87] G. Lenz *et al.*, "Aberrant immunoglobulin class switch recombination and switch translocations in activated B cell–like diffuse large B cell lymphoma," *J. Exp. Med.*, vol. 204, no. 3, pp. 633–643, 2007.
- [88] A. A. Jesus, A. J. S. Duarte, and J. B. Oliveira, "Autoimmunity in hyper-IgM syndrome," *J. Clin. Immunol.*, vol. 28, no. SUPPL. 1, pp. 62–66, 2008.
- [89] A. Melegari, M. T. Mascia, G. Sandri, and A. Carbonieri, "Immunodeficiency and autoimmune phenomena in female hyper-IgM syndrome," *Ann. N. Y. Acad. Sci.*, vol. 1109, pp. 106–108, 2007.
- [90] B. J. L. Hecht and J. C. Aster, "Molecular Biology of Burkitt ' s Lymphoma," *J. Clin. Oncol.*, vol. 18, no. 21, pp. 3707–3721, 2000.
- [91] Y. Minegishi *et al.*, "Mutations in activation-induced cytidine deaminase in patients with hyper IgM syndrome.," *Clin. Immunol.*, vol. 97, no. 3, pp. 203–210, 2000.
- [92] Y. Mu, C. Prochnow, P. Pham, X. S. Chen, and M. F. Goodman, "A structural basis for the biochemical behavior of activation-induced deoxycytidine deaminase class-switch recombination-defective hyper-IgM-2 mutants," *J. Biol. Chem.*, vol. 287, no. 33, pp. 28007–28016, 2012.
- [93] B. K. Birshtein, "Epigenetic regulation of individual modules of the immunoglobulin heavy chain locus 3' regulatory region," *Frontiers in Immunology*. 2014.
- [94] R. Bransteitter, P. Pham, M. D. Scharff, and M. F. Goodman, "Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase," *Proc. Natl. Acad. Sci.*, vol. 100, no. 7, pp. 4102–4107, 2003.

- [95] J. Chaudhuri, M. Tian, C. Khuong, K. Chua, E. Pinaud, and F. W. Alt, "Transcription-targeted DNA deamination by the AID antibody diversification enzyme," *Nature*, vol. 422, no. 6933, pp. 726–730, 2003.
- [96] S. K. Dickerson, E. Market, E. Besmer, and F. N. Papavasiliou, "AID mediates hypermutation by deaminating single stranded DNA.," *J. Exp. Med.*, vol. 197, no. 10, pp. 1291–6, 2003.
- [97] A. Sohail, J. Klapacz, M. Samaranayake, A. Ullah, and A. S. Bhagwat, "Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations," *Nucleic Acids Res.*, vol. 31, no. 12, pp. 2990–2994, 2003.
- [98] W. a Dunnick *et al.*, "Switch recombination and somatic hypermutation are controlled by the heavy chain 3' enhancer region.," *J. Exp. Med.*, vol. 206, no. 12, pp. 2613–2623, 2009.
- [99] L. Guglielmi, M. Le Bert, V. Truffinet, M. Cogné, and Y. Denizot, "Insulators to improve expression of a 3'IgH LCR-driven reporter gene in transgenic mouse models," *Biochem. Biophys. Res. Commun.*, vol. 307, no. 3, pp. 466–471, 2003.
- [100] E. M. Cortizas, A. Zahn, M. E. Hajjar, A.-M. Patenaude, J. M. Di Noia, and R. E. Verdun, "Alternative End-Joining and Classical Nonhomologous End-Joining Pathways Repair Different Types of Double-Strand Breaks during Class-Switch Recombination," *J. Immunol.*, vol. 191, no. 11, pp. 5751–5763, 2013.
- [101] P. Mali *et al.*, "RNA-guided human genome engineering via Cas9.," *Science*, vol. 339, no. 6121, pp. 823–6, 2013.
- [102] S. J. Ono, G. Zhou, A. K. F. Tai, M. Inaba, K. Kinoshita, and T. Honjo, "Identification of a stimulus-dependent DNase I hypersensitive site between the I $\alpha$  and C $\alpha$  exons during immunoglobulin heavy chain class switch recombination," *FEBS Lett.*, vol. 467, no. 2–3, pp. 268–272, 2000.
- [103] W. A. Dunnick, J. Shi, J. M. Zerbato, C. A. Fontaine, and J. T. Collins, "Enhancement of Antibody Class-Switch Recombination by the Cumulative Activity of Four Separate Elements," *J. Immunol.*, vol. 187, no. 9, pp. 4733–4743, 2011.
- [104] A. Garot *et al.*, "Sequential activation and distinct functions for distal and proximal modules within the IgH 3' regulatory region," *Proc. Natl. Acad. Sci.*, 2016.
- [105] V. Chandra, A. Bortnick, and C. Murre, "AID targeting: Old mysteries and new challenges," *Trends Immunol.*, vol. 36, no. 9, pp. 527–535, 2015.

- [106] N. Michael, H. M. Shen, S. Longerich, N. Kim, A. Longacre, and U. Storb, "The E box motif CAGGTG enhances somatic hypermutation without enhancing transcription," *Immunity*, vol. 19, no. 2, pp. 235–242, 2003.
- [107] A. Tanaka, H. M. Shen, S. Ratnam, P. Kodgire, and U. Storb, "Attracting AID to targets of somatic hypermutation," *J. Exp. Med.*, vol. 207, no. 2, pp. 405–415, 2010.
- [108] S. Zheng, B. Q. Vuong, B. Vaidyanathan, J. Y. Lin, F. T. Huang, and J. Chaudhuri, "Non-coding RNA Generated following Lariat Debranching Mediates Targeting of AID to DNA," *Cell*, vol. 161, no. 4, pp. 762–773, 2015.
- [109] P. Rouaud *et al.*, "The IgH 3' regulatory region controls somatic hypermutation in germinal center B cells.," *J. Exp. Med.*, vol. 210, no. 8, pp. 1501–7, 2013.
- [110] R. W. Maul *et al.*, "Spt5 accumulation at variable genes distinguishes somatic hypermutation in germinal center B cells from ex vivo-activated cells.," *J Exp Med*, vol. 211, no. 11, pp. 2297–2306, 2014.
- [111] A. Saintamand *et al.*, "Deciphering the importance of the palindromic architecture of the immunoglobulin heavy-chain 3' regulatory region," *Nat. Commun.*, vol. 7, p. 10730, 2016.
- [112] S. Peron *et al.*, "AID-Driven Deletion Causes Immunoglobulin Heavy Chain Locus Suicide Recombination in B Cells," *Science (80-. )*, vol. 336, no. 6083, pp. 931–934, 2012.
- [113] S. M. Gartler *et al.*, "Immunodeficiency with Hyper-IgM," in *Encyclopedia of Molecular Mechanisms of Disease*, 2009.
- [114] P. Quartier *et al.*, "Clinical, immunologic and genetic analysis of 29 patients with autosomal recessive hyper-IgM syndrome due to Activation-Induced Cytidine Deaminase deficiency," *Clin. Immunol.*, 2004.
- [115] V.-T. Ta *et al.*, "AID mutant analyses indicate requirement for class-switch-specific cofactors.," *Nat. Immunol.*, vol. 4, no. 9, pp. 843–848, 2003.
- [116] T. Kadungure, A. J. Ucher, E. K. Linehan, C. E. Schrader, and J. Stavnezer, "Individual substitution mutations in the AID C terminus that ablate IgH class switch recombination," *PLoS One*, vol. 10, no. 8, pp. 1–18, 2015.
- [117] R. M. Kohli, S. R. Abrams, K. S. Gajula, R. W. Maul, P. J. Gearhart, and J. T. Stivers, "A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase," *J. Biol. Chem.*, vol. 284, no. 34, pp. 22898–22904, 2009.

- [118] S. Sabouri, M. Kobayashi, N. A. Begum, J. Xu, K. Hirota, and T. Honjo, "C-terminal region of activation-induced cytidine deaminase (AID) is required for efficient class switch recombination and gene conversion," *Proc. Natl. Acad. Sci.*, vol. 111, no. 6, pp. 2253–2258, 2014.
- [119] K. Imai *et al.*, "Analysis of class switch recombination and somatic hypermutation in patients affected with autosomal dominant hyper-IgM syndrome type 2," *Clin. Immunol.*, vol. 115, no. 3, pp. 277–285, 2005.
- [120] R. Bransteitter, P. Pham, P. Calabrese, and M. F. Goodman, "Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase," *J. Biol. Chem.*, vol. 279, no. 49, pp. 51612–51621, 2004.
- [121] A. Zahn *et al.*, "Activation induced deaminase C-terminal domain links DNA breaks to end protection and repair during class switch recombination," *Proc. Natl. Acad. Sci.*, vol. 111, no. 11, pp. E988–E997, 2014.
- [122] T. Doi *et al.*, "The C-terminal region of activation-induced cytidine deaminase is responsible for a recombination function other than DNA cleavage in class switch recombination.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 8, pp. 2758–63, 2009.
- [123] S. Ito *et al.*, "Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 7, pp. 1975–80, 2004.
- [124] R. Geisberger, C. Rada, and M. S. Neuberger, "The stability of AID and its function in class-switching are critically sensitive to the identity of its nuclear-export sequence.," *Proc. Natl. Acad. Sci.*, vol. 106, no. 16, pp. 6736–6741, 2009.
- [125] M. Takizawa *et al.*, "AID expression levels determine the extent of cMyc oncogenic translocations and the incidence of B cell tumor development.," *J. Exp. Med.*, vol. 205, no. 9, pp. 1949–57, 2008.
- [126] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity," *Science (80-. )*, 2012.
- [127] T. W. W. Chiang, C. Le Sage, D. Larrieu, M. Demir, and S. P. Jackson, "CRISPR-Cas9D10A nickase-based genotypic and phenotypic screening to enhance genome editing," *Sci. Rep.*, 2016.
- [128] Q. Qiao, L. Wang, F. L. Meng, J. K. Hwang, F. W. Alt, and H. Wu, "AID Recognizes Structured DNA for Class Switch Recombination," *Mol. Cell*, vol. 67, no. 3, pp. 361-373.e4, 2017.

- [129] S. G. Lebecque and P. J. Gearhart, "Boundaries of somatic mutation in rearranged immunoglobulin genes: 5' boundary is near the promoter, and 3' boundary is ~1 kb from V(D)J gene," *J. Exp. Med.*, 1990.
- [130] A. L. Kenter, "AID targeting is dependent on RNA polymerase II pausing," *Seminars in Immunology*. 2012.
- [131] R. W. Maul, H. Saribasak, Z. Cao, and P. J. Gearhart, "Topoisomerase I deficiency causes RNA polymerase II accumulation and increases AID abundance in immunoglobulin variable genes," *DNA Repair (Amst)*., 2015.
- [132] P. D. Bardwell *et al.*, "Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice," *Nat. Immunol.*, vol. 5, no. 2, pp. 224–229, 2004.
- [133] J. Eccleston, C. E. Schrader, K. Yuan, J. Stavnezer, and E. Selsing, "Class switch recombination efficiency and junction microhomology patterns in Msh2-, Mlh1-, and Exo1-deficient mice depend on the presence of mu switch region tandem repeats," *J Immunol*, vol. 183, no. 2, pp. 1222–1228, 2009.
- [134] Z. Xu, E. J. Pone, A. Al-Qahtani, S.-R. Park, H. Zan, and P. Casali, "Regulation of aicda Expression and AID Activity: Relevance to Somatic Hypermutation and Class Switch DNA Recombination," *Crit. Rev. Immunol.*, 2012.
- [135] A. J. Ucher *et al.*, "Mismatch Repair Proteins and AID Activity Are Required for the Dominant Negative Function of C-Terminally Deleted AID in Class Switching," *J. Immunol.*, vol. 193, no. 3, pp. 1440–1450, 2014.