

**A Novel Targeted Deletion of Mbd4 Reveals an Unappreciated Role in Class Switch
Recombination**

BY

FERNANDO GRIGERA
B.A. (Oberlin College) 2002

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Microbiology and Immunology
in the Graduate College of the
University of Illinois at Chicago, 2013

Chicago, Illinois

Amy L. Kenter, Advisor and Chair
William Walden
William Hendrickson, Microbiology and Immunology
Toru Nakamura
Elizaveta Benevolenskaya, Genetics and Biochemistry

TABLE OF CONTENTS

<u>CHAPTER</u>	<u>PAGE</u>
1 INTRODUCTION	1
1.1 Early B cell Development.....	2
1.2 Antibody diversification: SHM and CSR.....	3
1.2.1 Overview.....	3
1.2.2 Somatic hypermutation.....	4
1.2.3 Class switch recombination.....	7
1.3 AID initiates antibody diversification.....	10
1.3.1 Overview.....	10
1.3.2 AID Function.....	11
1.3.3 AID Targeting.....	12
1.3.4 AID Regulation.....	14
1.4 MMR and BER in SHM and CSR.....	16
1.4.1 Overview.....	16
1.4.2 BER in mammalian systems.....	17
1.4.3 Loss of Ung studies in human and mouse.....	18
1.4.4 The role of APE1 and APE2 in CSR.....	19
1.4.5 DNA Polymerase β	20
1.4.6 MMR in mammalian systems.....	21
1.4.7 Msh2 and Msh6 and Exo I.....	24
1.4.8 Mlh1 and Pms2.....	25
1.4.9 Competing for access to the U/G mismatch.....	27
1.4.10 Models of DSB formation and A:T basepair mutations.....	28
1.5 DSBs breaks recruit DNA repair factors.....	33
1.5.1 Overview.....	33
1.5.2 C-NHEJ and A-EJ.....	34
1.6 Mbd4.....	37
1.6.1 Overview.....	37
1.6.2 Catalytic properties of Mbd4.....	37
1.6.4 Mbd4 as a part of MMR.....	39
1.6.5 A role for Mbd4 in antibody diversification.....	40
2 METHODS AND MATERIALS	41
2.1 Construction of Mbd4 Targeting vector.....	42
2.2 Selection of Mbd4 targeted CH12.F3 Clones.....	43
2.3 Southern analysis.....	43
2.4 CH12.F3 cell culture and IgA CSR.....	47
2.5 Mice.....	48
2.6 Primary B cell culture and IgG1 and IgG3 CSR.....	48
2.7 Flow cytometry.....	49
2.8 Quantitative (q) RT-PCR.....	50
2.9 Proliferation assay.....	51
2.10 Western Analysis.....	51
2.11 LM-PCR Double strand breaks assay.....	52

TABLE OF CONTENTS (Continued)

<u>CHAPTER</u>	<u>PAGE</u>
2.12 Cloning of individual LM-PCR products.....	53
2.13 Cloning and identification of S μ -S α junctions.....	54
2.14 Plasmids.....	54
2.14.1 pCDNA ^{NSI*} -PGK-Cre.....	54
2.14.2 pCDNA ^{NSI*} -PGK-Mbd4.....	58
2.14.3 pMSCV transduction constructs.....	58
2.15 Transfection.....	59
2.16 Genetic Complementation and transduction.....	61
3 RESULTS: Characterization of the Mbd4 ΔE2-5 mouse.....	62
3.1 CSR is not perturbed in the Δ E2-5 Mbd4 KO mouse	63
3.2 MMR proteins are substantially reduced the Δ E2-5 Mbd4 KO mouse.....	64
3.3 Mbd4 modulates CSR in primary B cells	68
3.4 Ectopic expression of UNG2, but not UNG1, inhibits CSR	69
3.5 Exons 6-8 of Mbd4 are expressed in Δ E2-5 Mbd4 KO mouse.....	72
4 RESULTS: Characterization of the Mbd4 DE6-8 KO CH12.F3 cell line.....	75
4.1 Targeting the catalytic exons of Mbd4 for deletion	76
4.2 1A-12 Δ/Δ is an Mbd4 KO clone and has a diminished CSR phenotype.....	78
4.3 Appropriate proliferation, GLT, and AID expression.....	81
4.4 MSH2 and MLH1 are reduced in Mbd4 KO CH12.F3.....	85
4.5 Analysis of DSBs in the Mbd4 KO CH12.F3 Cell line.....	86
4.5.1 Reduced double strand breaks (DSBs) in the Mbd4 deficient cells.....	86
4.5.2 Skewing of DSBs to AG in the Mbd4 KO cell line.....	90
4.5.3 Skewing of S μ DBSs in the Mbd4 KO.....	93
4.6 S μ -S α microhomology length is increased in Mbd4 KO clone.....	96
4.7 Retroviral transduction and genetic complementation in Mbd4 KO cell line....	100
5 DISCUSSION.....	107
5.1 Summary.....	108
5.2 Deletion of exons 2-5 of Mbd4 did not produce a CSR phenotype.....	109
5.3 Failure to rescue the IgA CSR defect with genetic complementation.....	111
5.4 Concluding remarks.....	116
APPENDICES.....	119
Appendix A.....	120
Appendix B.....	121
Appendix C.....	129
Appendix D.....	130
CITED LITERATURE.....	131
VITA.....	142

LIST OF FIGURES

<u>FIGURE</u>	<u>PAGE</u>
1. Model for the generation of mutations during Phase I and Phase II of SHM...	6
2. Schematic of CSR in the Igh locus.....	8
3. Model of 5'- and 3'- directed MMR	23
4. Model of DSB formation.....	29
5. Model of SHM Phase II mutations.....	32
6. Construction of pLNTK-5' Mbd4 homology arm.....	44
7. Construction of pLNTK-3' Mbd4 homology arm.....	45
8. Construction of Mbd4 targeting vector.....	46
9. Construction of pCDNA-PGK-EGFP.....	57
10. Construction of pCDNA-PGK-Cre, pCDNA-PGK-Mbd4, and pCDNA-PGK	60
11. E Δ 2-5 deletion of Mbd4 gene does not influence IgG1 or IgG3 CSR. Mbd4 +/+ (WT) and Mbd4 Δ 2-5/ Δ 2-5.....	65
12. Loss of Mbd4 compromises stability of MMR proteins and this occurs post- transcriptionally.....	67
13. Nuclear UNG (UNG2) inhibits IgG1 CSR.....	73
14. Deleting exons 6-8 of Mbd4 in the CH12.F3 cell line.....	77
15. IgA CSR is diminished as a result of Mbd4 ablation in CH12.F3 cell line...	79
16. qRT-PCR analysis confirms lack of Mbd4 mRNA and induction in wild type.....	80
17. MBD4 is reduced in a heterozygous Mbd4 clones and abolished in a null Mbd4 clone.....	82
18. Mbd4 null CH12.F3 cells proliferate normally.....	83
19. Aicda gene transcript, μ GLT, and α GLT are induced normally after CIT stimulation in null (KO) Mbd4 CH12.F3 cells.....	84

LIST OF FIGURES (continued)

<u>FIGURE</u>	<u>PAGE</u>
20. MSH2 and Mlh1 proteins are diminished in the absence of MBD4.....	88
21. Mbd4 ablation leads to a substantial reduction of DSBs.....	89
22. Breaks are skewed towards AG in both CH12.F3 Mbd4- deficient cells and Msh2- and Mlh1- deficient primary B cells.....	91
23. DSBs skew toward S μ TR region in the absence of Mbd4.....	94
24. S μ -S α junctions from Mbd4 deficient CH12.F3 cells have longer microhomologies.....	99
25. Genetic complementation with MBD4 in IgA rescue does not coincide with expression of protein.....	101
26. Genetic complementation with MBD4 in Mbd4 deficient cells fails to rescue IgA CSR.....	103
27. pCDNA expression system failed to rescue IgA CSR defect in Mbd4 KO cells.....	106
28. Transcription in the Mbd4 Locus is complex and overlaps with CN781668....	120
29. Compendium of S μ -S α junction sequences.....	121
30. Plasmid constructs used for transfection or transduction experiments.....	129
31. PCR of Mbd4 and CN781668 transcripts.....	130

LIST OF ABBREVIATIONS

V(D)J- Variable, (Diversity), Joining
CSR- Class Switch Recombination
SHM- Somatic Hypermutation
AID- Activation Induced Cytidine Deaminase
Ig- Immunoglobulin
BER- Base Excision Repair
UNG- Uracil N-Glycosylase
DNA Pol β - DNA Polymerase β
APE1/2- Apurinic/apirymidinic Endonuclease1 /2
MMR- Mismatch Repair
MSH2- MutS Homolog 2
MLH1- MutL Homolog 1
MBD4- Methyl CpG Binding Domain protein 4
Ab- Antibody
Ag-Antigen
A- Adenine
C-Cytosine
G-Guanidine
T-Thymine
⁵MeC- 5-Methyl Cytosine
Exo1- Exonuclease I
DSB- Double-Stranded Break
SSB- Single-Stranded Break
HNPCC- Hereditary Non Polyposis Colorectal Cancer
HhH-GDP- Helix-hairpin-Helix Glycosylase Superfamily
GLT- Germline Transcript
C-NHEJ- Classical Non-Homologous End Joining
A-EJ- Alternative End Joining
HR- Homologous Recombination
PMS2- Post- Meiotic Segregation factor 2

SUMMARY

Methyl CpG binding domain protein 4 (Mbd4) is a mismatch DNA glycosylase is involved in rectifying T/G and U/G mismatches that result from deamination of 5-methyl cytosine (5meC) and cytosine (C) in DNA, respectively. Activation induced cytidine deaminase (AID) initiates class switch recombination (CSR) and somatic hypermutation (SHM), two processes that occur in B cells critical for expanding the primary antibody repertoire, by deaminating C, thereby converting it to uracil (U) and creating a U/G mismatch in DNA. Key intermediates, double-stranded breaks (DSBs) and Phase II SHM mutations are formed downstream of the U/G lesion and are not wholly dependent on the activity of uracil N-glycosylase (UNG) in CSR and SHM, respectively, suggesting that other glycosylases fulfill these roles. The aim of this thesis was to determine whether Mbd4 supplies glycosylase and/or scaffolding activity during CSR.

B cells harboring a deletion of exons 2-5 of the Mbd4 gene (Mbd4 $\Delta^{2-5\Delta 2-5}$) do not manifest any CSR phenotype. An investigation of the mismatch repair proteins, Msh2, and Mlh1, revealed that these proteins are substantially reduced in this genetic background. The lack of CSR phenotype is surprising with respect to MMR proteins, because these genes have been shown to be haploinsufficient for CSR, suggesting a potential modulator role of Mbd4. In contradistinction, a CH12.F3 cell line that I designed where exons 6-8 of the Mbd4 gene have been deleted (Mbd4 $\Delta^{6-8\Delta 6-8}$) show a reduced CSR phenotype, which is accompanied by reduced DSBs, a skewing of DSB positioning within S μ , and longer microhomologies at S μ -S α junctions. In addition MMR proteins were also affected by Mbd4 deficiency in the CH12.F3 cell line, highlighting the only similarity between these different Mbd4 deficient models.

I conclude that the Mbd4 is involved in CSR and it shares several features common to the MMR pathway. The discrepancy that exists between the different Mbd4 deletion models can be attributed to the incomplete nature of the Mbd4 deficiency in Mbd4^{Δ2-5Δ2-5} B cells. I determined that a residual transcript continues to be made in these cells and that it yields a stable truncated peptide that is likely protective for CSR.

CHAPTER 1
INTRODUCTION

1.1 Early B cell Development

The early stages of B cell maturation take place in the bone marrow from a common lymphoid progenitor. Stages are broken up into pre-pro-B, pro-B, pre-B, and finally immature B cells, whereupon cells exit the bone marrow and migrate to various peripheral organs of the immune system (1). A coordinated cascade of expression of various transcription factors closely follows the fate of the B cell as it completes one stage of development and enters the next. At several key points during development signaling through the B cell receptor (BCR) on the surface of the B cell also provides a cue that one stage of development has reached completion.

The BCR does not come pre-assembled; rather it has to go through several rounds of gene rearrangement before it is even expressed. Rearrangement occurs on three loci, the first two comprising the immunoglobulin light chain loci (*Igk* and *Igl*) and the third the immunoglobulin heavy chain locus (*Igh*). Multiple variable (V) and joining (J) segments are found in all three genes, while diversity (D) segments are unique to the *Igh* locus (2). Flanking each of the different V, D, and J segments, specific sequences called recombination signal sequences (RSSs) direct the recombinase-activating genes 1 and 2 (RAG1 and RAG2) to catalyze the V(D)J recombination reaction (3-5). RAG1 and RAG2 create double-stranded breaks (DSBs), which are repaired by the classical non-homologous end-joining pathway (C-NHEJ) (see section 1.5.2). V(D)J recombination proceeds in an ordered sequence of events that coincides with the completion of one of each of the stages of B cell development. D-J joining begins first on the *Igh* locus and takes place during pre-pro B to pro-B cell stage. As D-J joining reaches completion, V-DJ joining commences and a transition takes place from pro B to pre B stage, as now a functional heavy chain (IgM) associates with the surrogate light chain receptors, $Ig\alpha$ and $Ig\beta$, together forming the preB cell receptor (pre-BCR) which is expressed on the surface. Signaling

through the pre-BCR causes the pre B cells to convert from CD43⁺ to CD43⁻ and initiates V-D joining at the *Igκ* and *Igλ* loci (6). Once a functional light chain (in human and mouse mostly *Igκ*) has been assembled it is paired with a single heavy chain, and a tetramer is formed from two of these heterodimers to make a functional Immunoglobulin (Ig) that is expressed on the surface and marks the transition from pre-B to immature B cell. Immature B cells are then subjected to a rigorous selection in the bone marrow that shapes the primary repertoire of antibody (Ab) specificities through positive and negative selection. Deleting self-reacting B cells from this repertoire is the subject of central tolerance (reviewed in (7, 8)), which will not be discussed further here. Immature B cells that survive this process are now ready to exit the bone marrow where they will migrate to various organs of the peripheral immune system, such as, lymph nodes, Peyer's patches, and the spleen. Here they will become mature B cells and upon encounter with antigen (ag) will undergo another set of antibody gene alterations that further diversifies the Ig repertoire

1.2 **Antibody diversification: SHM and CSR**

1.2.1 **Overview**

V(D)J rearrangement establishes a primary repertoire of Ab specificities that can vary from species to species depending on the number of available germline-encoded V, D and J segments and the expression of terminal deoxynucleotidyl transferase (TDT) (which adds non-templated nucleotides at V-J and V-(D)J junctions) (9, 10). Chickens, which carry only one or very few functional V, J, and D segments do not generate a robust primary repertoire of antibody specificities through V(D)J joining. Chicken and rabbit generate diversity via another unrelated

process, termed gene conversion (reviewed in (11)). This is different from man and mouse, where numerous functional V, D, and J segments can be rearranged in V(D)J joining in conjunction with TDT to create an impressive primary repertoire ranging between 10^5 - 10^6 antibody specificities. However, no animal species generates a primary repertoire of antibody specificities large enough to match the much higher Ab specificity that will be needed to accommodate the vast array of antigens an animal will encounter throughout their lifetime. In order to develop this kind of repertoire, somatic hypermutation (SHM) achieves this feat, which is the subject of the next section.

1.2.2 Somatic hypermutation

Inside the peripheral organs of the immune system structures termed germinal centers house a clonally multiplying population of B cells (12, 13). This impressive proliferation is in response to antigen and cytokines, mainly produced by T helper (T_H) cells), and is accompanied by a staggering rate of mutation (10^{-2} to 10^{-3} mutations per base pair) in V genes. This rate is 10^6 times higher than the rate of basal mutation in the cell (14). SHM is targeted to V genes in the *Igh*, *Igκ*, and *Igλ* loci and S regions in the *Igh* locus. Most of the mutations in V genes are confined to a 2kb region, which extends from 300bp flanking the 5' side of the V(D)J gene to 850bp in the flanking region of the 3' side (15). Mutations in S regions (both donor S_μ and acceptor ($S\gamma 1$, eg.)) span much larger distances, between 4-7kb, starting about 150 bp 3' of the I exon and extending for several kilobases (16).

Aptly named, somatic hypermutation (SHM) alters the V genes encoding the BCR, generating new Igs with higher affinity specificities that did not previously exist. In a process termed affinity maturation, BCR genes undergo several rounds of SHM, continually

accumulating mutations, in some cases leading to BCRs with higher and still higher affinity for Ag. B cells that express high-affinity BCRs bind to Ag better than other B cells in the germinal center population and concomitantly receive stronger a proliferation signal and thusly are preferentially expanded. This results in the generation of a secondary repertoire of superior Ab specificities that can further differentiate into memory and plasma cells (17).

There are two non-overlapping phases of SHM which give rise to the complete spectrum of mutations (reviewed in (18)). About half of all mutations occur in the first phase of SHM, termed Phase I, and affect G:C base pairs, the majority of which fall into the WRCY consensus, (where W=A or T; R= G or A, and Y= C or T) where the AGCT iteration is predominant. The coordinated action of activation-induced cytidine deaminase (AID) and uracil DNA or N-glycosylase (UNG) (covered in chapters 1.3.2 and 1.4.3, respectively) leads to transition and transversion mutations. AID converts deoxycytidine (dC) to deoxyuridine (dU), which creates a U/G mismatch that can in turn be recognized and excised by UNG to create an abasic (AP) site. Replication over uracil leads to C→T and G→A transitions, while replication over the AP site can lead to G→C and G→T transversions. Processing by REV1 and translesional (TLS) polymerases also contribute to the generation of transitions and transversions stemming from replication across the abasic site (summarized in Fig. 1).

The remaining half of SHM mutations occur in Phase II, and affect A:T basepairs. The mechanism here is different than the one from Phase I and involves recognition and excision of dU by the MutS α heterodimer (Msh2/Msh6) coupled to a gap-filling process that is carried out by an error-prone polymerase, pol η (fig. 1) (see also chapter 1.4.7, fig.1, and fig. 5). The distribution of mutations at the A:T basepair is not random, with more than twice as many

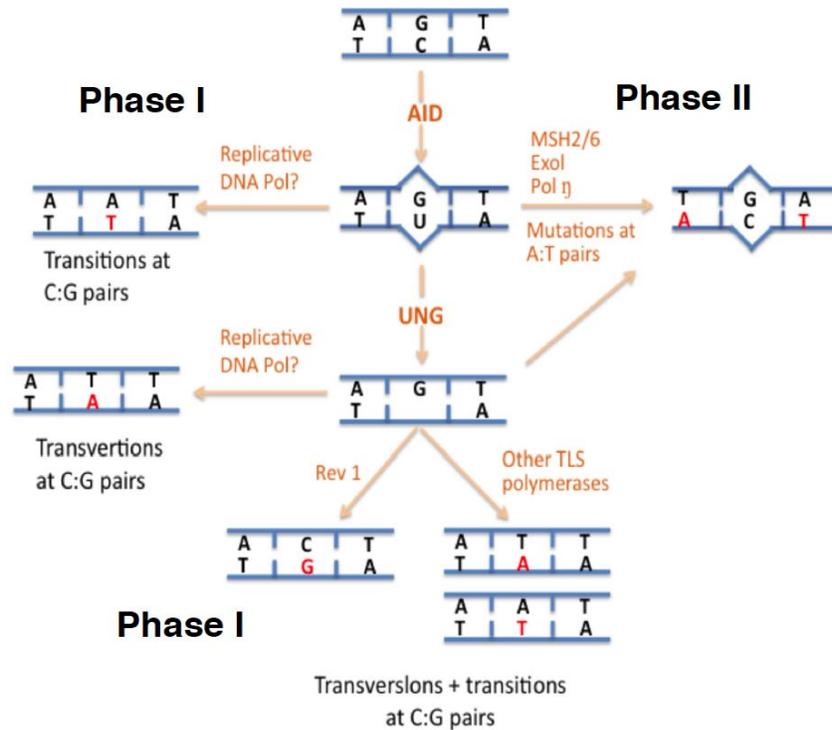


Figure 1. **Model for the generation of mutations during Phase I and Phase II of SHM.** AID deaminates dC . Phase I mutations affect C:G basepairs and are dependent on UNG. Replication across deoxyuridine can lead to transition mutations. Replication across the abasic site can lead to transitions or transversions. Rev1 and translesional polymerases (TLS) contribute to the mutation spectrum at C:G basepairs. Phase II mutations are dependent on the MutS α heterodimer (Msh2/Msh6), Exo I, and the error prone polymerase, pol η .

mutations occurring at A than at T on the template strand (19). Both phases of SHM are entirely dependent on the presence of the instigating U/G mismatch, which is also critical for initiating CSR. While mutations occur in S regions with high frequency during CSR the conversion of the U/G mismatch into double-stranded breaks (DSBs) is more important for promoting CSR.

1.2.3 Class switch recombination

CSR takes place inside of peripheral lymphoid organs and expands the pool of switched Ig-bearing cells. Unlike SHM, CSR only affects the *Igh* locus and occurs when the default IgM isotype is switched to one of five different downstream isotypes in mouse, IgG1, IgG3, IgG2a, IgE, IgA. This change is not accompanied by a change in the antibody specificity that was previously acquired in V(D)J joining. In mouse there are 8 C_H genes ($C\mu$, $C\delta$, $C\gamma3$, $C\gamma1$, $C\gamma2b$, $C\gamma2a$, $C\epsilon$, and $C\alpha$) together spanning approximately 220kb of the *Igh* locus, while the remaining nearly three megabases of the *Igh* locus houses numerous V, D, and J segments (20). The *Igh* locus is delimited at the 5' end by an intronic enhancer ($E\mu$) and at the 3' end by a series of DNA hypersensitive sites (HS3A, HS1,2, HS3b4, and HS5,6,7) collectively referred to as the 3' RR (21, 22). In between the enhancers are the different C_H genes, each containing (in the following order) a germline transcript (GLT) promoter (except for $C\delta$), an I-box, a switch (S) region, and a constant (C) region (fig. 2A).

The newly expressed constant gene maintains the original specificity of the Ab but provides it with new effector functions such as the ability to localize to mucosal epithelia (IgA), elicit an allergic reaction (IgE), and cross the placenta (IgG), for example (23). The CSR reaction occurs at the molecular level via an intrachromosomal deletion event that results in the excision

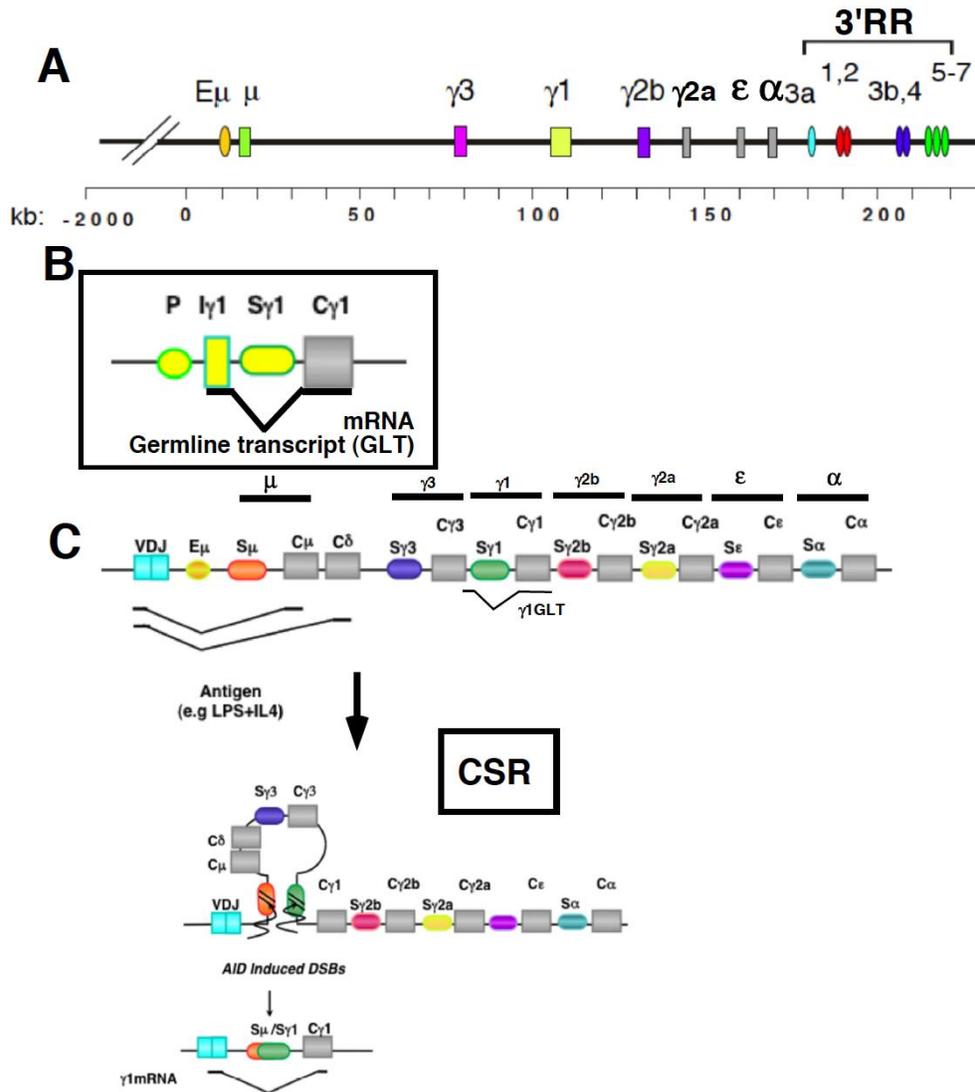


Figure 2. **Schematic of CSR in the *Igh* locus** **A**) A map of the 3' end of the *Igh* locus (drawn to scale) spanning ~220kb. Includes all C_H genes, the $E\mu$ enhancer, and the 3' RR. **B**) Diagram showing the components of the I-S- C_H locus ($\gamma 1$ is shown here as an example) and the resulting germline transcript (GLT). **C**) Cartoon depicts the *Igh* locus that is poised to undergo class switch to IgG1($C\gamma 1$) (not drawn to scale) upon stimulation with LPS +IL4.

of circularized intervening DNA and a rearrangement of the *Igh* locus. During CSR two distal switch (S) regions of DNA, which are immediately 5' of the constant (C) regions, come into close proximity to one another (24). DSBs are introduced into these regions by AID and resolved by DNA repair pathways (see chapters 1.3.3, 1.4.10, and 1.5). The example in figure 2C illustrates the molecular process of a B cell that is undergoing CSR from IgM to IgG1. Once the switch to another isotype is completed the rearranged *Igh* gene can be translated into a new heavy chain.

CSR is often used as an assay to study genes that are specific to antibody diversification in primary B cells. This is typically done using splenocytes or CD43- enriched quiescent B cells that are cultured in the presence of mitogens and cytokines that mimic conditions found *in vivo*. For example, addition of the mitogen, lipopolysaccharide (LPS) induces $\mu \rightarrow \gamma 3$ and $\mu \rightarrow \gamma 2b$ CSR, a T-cell independent process, while LPS and interleukin 4 (IL4) represses $\mu \rightarrow \gamma 3$ and $\mu \rightarrow \gamma 2b$ CSR, and instead induces $\mu \rightarrow \gamma 1$ and $\mu \rightarrow \epsilon$ CSR. Sometimes CD40L is included to induce CSR in B cells via the CD40 receptor, an interaction which normally occurs *in vivo* between B cells and T_H cells. Genetic mutations in either CD40 or CD40L lead to immunodeficiency class I Hyper IgM syndrome, a condition characterized by elevated IgM serum levels and an inability to make any of the other isotypes (25). The addition of inducers causes the transcriptional activation of different C_H genes, which precedes the actual recombination event, thus CSR to a particular isotype corresponds to specific transcriptional activation of a given C_H GLT promoter.

Induction of IgM to another downstream C_H isotype is tightly correlated to transcription through that C_H gene, and is preceded by the transcriptional activation through C_μ and the downstream C_H gene. Transcription is initiated at the C_H GLT promoters and extends through the 3' end of the C region. Splicing of I-box to C region, removes the S region, and yields a germline

transcript (GLT), which is a sterile transcript that does not get translated into protein (fig. 2B) (25). Curiously this splicing event is required for efficient CSR, but the reasons for this are unknown (26). These GLTs can be detected as early as 4 hours after induction of CSR and are absolutely required. The impact of loss GLTs on CSR is evident in mice in which the $\gamma 1$ promoter has been removed resulting in the absence of a $\gamma 1$ GLT and a complete failure to undergo $\mu \rightarrow \gamma 1$ CSR (26). Transcription through V segments is also required for SHM (27). Transcription creates a setting where a suitable substrate for activated-induced cytidine deaminase (AID) is revealed. The discovery of AID as the sole B cell specific factor and its role in initiating Ab diversification reactions is explored next.

1.3 AID initiates antibody diversification

1.3.1 Overview

AID was originally identified by a subtraction hybridization screen of cDNAs derived from the B cell lymphoma CH12.F3 under uninduced and induced-to-switch to IgA conditions (28). The follow-up to this paper defined AID as the critical mediator of two Ab diversification processes, SHM and CSR, that until then were not thought to be mechanistically related (29). It was further shown that a subset of patients suffering from a condition characterized by elevated IgM and an inability to make Abs of other isotypes (Class II Hyper IgM syndrome) harbored recessive mutations in the *Aicda* locus (30), thereby solidifying a connection to AID.

Deamination of cytosine by AID causes the inappropriate acquisition of uracil into DNA creating a U/G mismatch (31). This simple chemistry constitutes the bedrock of Ab diversification in mature B cells leading to hypermutation in SHM, but in the the case of CSR

causes the formation of a DSB (32). The function, targeting, and regulation of AID will be covered in the next section.

1.3.2 AID Function

The original discovery of AID as a novel member of the apolipoprotein B mRNA editing enzyme catalytic polypeptide-like (APOBEC) family led to the hypothesis that the role of AID in antibody diversification must involve the editing of an mRNA encoding a new endonuclease function (33-35). Most evidence now suggests that with the exception of APOBEC 1, which probably acquired the capability to deaminate RNA (36, 37), all the other members of the APOBEC family use DNA as a substrate (38). Certainly, where SHM and CSR are concerned mounting evidence supports the hypothesis that AID deaminates DNA (reviewed in (39)).

Recently, a provocative twist on the RNA-editing hypothesis was proposed that linked AID-dependent editing of an unknown micro RNA that binds to and inhibits the translation of the topoisomerase I (TOPI) mRNA (40, 41). Down regulation of TOPOI leads to increased SHM and is also proposed to create a temporary B DNA structure that is prone to cleavage, which is suggested to be the necessary DSB formation step in CSR. While it's possible that AID could deaminate RNA *in vivo*, the lack of a defined physiological target is still outstanding with regard to the TOPOI finding. As I will now demonstrate most evidence points to AID deaminating DNA.

AID is capable of deaminating a single-stranded DNA (ssDNA) substrate, but not double-stranded DNA (dsDNA) *in vitro* (42). If the substrate is transcribed, however, it was possible for AID to deaminate dsDNA (43), a finding that seems to corroborate the tight correlation found *in vivo* between AID activity and germline transcription of the constant (C)

genes of the *Igh* locus (reviewed in (44)). The latter finding suggested that AID gains access to its *in vivo* ssDNA substrate via transcription. Ectopic expression of AID in yeast and bacteria also leads to accumulation of C→T and G→A transition mutations (45-47), indicating that DNA was the likely substrate of AID.

While these reports all lend supporting evidence to the DNA deamination model, none of them were able to show directly that uracil residues in DNA are a result of AID deamination. The smoking gun, as it were, arrived on the scene only recently, as Maul et al., showed convincingly that U residues accumulate in immunoglobulin variable and switch regions in *Ung* deficient mice in an AID-dependent manner (31). The involvement of UNG in SHM and CSR is covered in chapter 1.4.

1.3.3 AID Targeting

A long-standing question in the field is how AID is targeted to V and S regions within Ig loci genes during SHM and CSR, respectively. One suggestion was that stable G-rich structures found within S regions, termed R-loops, which form extensively in S mammalian regions, form during transcription and grant AID access to its ssDNA substrate (48, 49). R-loops are proposed to form due to the high G/C content of the growing nascent mRNA GLT, which hybridizes to the template strand and allows the non-template strand to be targeted by AID. However, when S DNA from *Xenopus*, which has AT rich sequences, was used to replace S DNA from the mouse, CSR efficiency was somewhat reduced, but not entirely, suggesting that R-loops are not essential for targeting AID in CSR (50). In this report it was postulated that the ancestral AGCT motif, coupled to recognition by replication protein A (RPA) was what targeted AID to S regions.

One puzzling feature of AID targeting is that it targets both strands of the DNA duplex *in vivo*, a finding that could not be faithfully reproduced *in vitro*. It is possible to achieve deamination that approaches *in vivo* frequency using transcribed *in vitro* substrates in the presence of AID, RPA, and protein kinase A (PKA), but only on the non-template strand (51). Another possibility is that anti-sense transcripts, which were detected at low levels at both V and S regions, but not the corresponding C regions, could somehow target AID to the template strand (52). It's not clear, however, whether anti-sense transcripts are simply a secondary transcriptional event. New insight into how the template strand gets deaminated by AID comes from an exciting study of the RNA exosome, which is a conserved RNA degrading complex consisting of nine core subunits that can produce regions of ssDNA during transcription. In this study, partially purified AID-enriched exosome complexes were shown to induce the ability of AID to deaminate both strands of a transcribed *in vitro* dsDNA substrate (51). This reaction was totally dependent on transcription and highest in fractions containing the purified complex. Thus it appears the long sought solution to explain how both strands of the DNA duplex are targeted by AID may finally have arrived.

The sequence of DNA is also important and the existence of the WRCY hotspot and its reverse complement, RGYW, (where W=A or T, R =A or G, and Y=C or T) within V and S regions seems to attract AID activity, although deamination within these hotspot motifs, where SHM and CSR are concerned, never breaches more than 40-50% of all mutations and DSBs, respectively (53-56). Clearly, specific sequences and structures found in V and S regions do not tell the whole of story of how AID is targeted during SHM and CSR, respectively.

Our lab has also explored epigenetic marks that seem to couple AID activity to chromatin accessibility. During transcription the H3K4me3 mark is laid down promoter-proximal, usually

peaking around 200bp from the transcription initiation site and tapering off rather quickly after that (57, 58). The unusual persistence of the activating H3K4me3 mark, extending far into S DNA regions, but not C regions, coupled with stalling of RNA polymerase II may create a “window” of chromatin accessibility, wherein AID can find its ssDNA substrate and coordinate a crosstalk with DNA repair proteins (59, 60). Where transcription is involved in CSR, per se, the identification of the transcription elongation factor, Spt5, as an AID binding partner has furthered the notion that stalling of RNA polymerase II likely plays a key role in AID targeting (61). This is an exciting model because it provides a template that not only describes how AID targets primarily to specific regions within Ig genes, but also begins to explain how AID can target non-Ig genes. Ectopic overexpression of AID in a transgenic mouse model caused all of the mice to die of T cell lymphomas (62). AID is not expressed in T cell normally and in the transgenic mouse mutations were found in the TCR genes. Interestingly, genes that were actively transcribed in either T or B cells showed a similar level of AID-dependent mutation, making a case for active transcription serving as the main docking site for AID activity. The subject of how AID targets non-Ig genes is highlighted in (63) and not discussed further here..

Therefore, it is clear that AID targeting is linked to the transcription apparatus, but additional features, including the structure and sequence of the transcribed target, associated RNA degrading complexes, and the milieu of chromatin architecture are also important.

1.3.4 AID Regulation

As it has become clear that aberrant AID expression is involved in lymphomagenesis, the careful regulation, on several levels, of a promiscuous mutator makes logical sense. At the most basic level, expression of AID is controlled by a variety of B cell specific transcription factors,

including Pax5, STAT6, NF- κ B, E2A, and FOXO1 (64, 65). AID expression is down regulated by calmodulin inhibition of E proteins upon strong engagement of the B cell receptor with antigen (Ag) (66). The later finding illustrates nicely how the cell limits AID expression after it is not needed. In this case once a high affinity Ab has been generated there is no further requirement for AID to be active.

A dichotomy exists between different parts of the AID protein that seems to direct its activity towards either SHM or CSR. The N-terminus of AID is dispensable for CSR, while it is important for SHM (67). Conversely, the last ten C-terminal amino acids of AID comprise an evolutionarily conserved nuclear export signal (NES), the removal of which nearly abolishes CSR but does not affect SHM (68, 69). It is due to this NES that AID finds itself largely confined to the cytoplasm, a safeguard of sorts that may prevent AID from attacking non-Ig genes (70). AID is also regulated by phosphorylation at serine 38 by protein kinase A (PKA), which was detected at S regions independently of AID, and possibly coordinates AID activity with RPA at regions of ssDNA (51, 71, 72).

AID is also regulated at the transcript level in a process that likely involves translational blocking by micro RNA- mediated interference. Mutating the binding site for mir155 in the 3' untranslated region (3'UTR) of AID led to increased CSR at all isotypes that were investigated (73). Ablation of the *bic* gene, which contains the transcriptional unit of mir155, recapitulated the previous report's finding, and additionally was found to increase Myc-IgH translocations (74). This particular translocation is the signature of many B cell lymphomas that may harbor a dependence on AID (75-77).

Another level of regulation modulates protein levels of AID via ubiquitin- independent proteosomal degradation. Recently, a new interacting partner for AID, Regulin- γ (Reg γ), was

discovered. Ablation of Reg γ but not closely related family members, Reg α , or Reg β , modestly increased CSR to IgG1 and IgG3 (78). While degradation of AID seems to be largely dependent on the classical ubiquitin-dependent pathway, addition of MG-132, an inhibitor of the 26S proteasomal unit, revealed additional degradation of AID that was entirely dependent on Reg γ (78).

To summarize, AID deaminates dC in DNA converting it to U, thereby creating a U/G mismatch, which in B cells results in the generation of Ab diversity from mutations in SHM and DSBs in CSR. The absolute requirement of AID for initiating these processes in B cells must be counterbalanced by several levels of control to ensure that the dangerous chemistry deployed by AID does not wreak havoc on the genome. Clearly, the existence of a subset of B cell lymphomas the etiology of which can be traced to AID underpins the danger of unfettered AID expression.

Paradoxically, the very same mechanisms enlisted by the cell to combat mutation and DNA cleavage are indispensable to generating Ab diversification in SHM and CSR, respectively. Evidently, a careful balance has been struck between simple repair of the AID-induced lesion by various DNA repair pathways and generation of mutation and DSBs by the very same pathways. The subject of these pathways will now be explored.

1.4 MMR and BER in SHM and CSR

1.4.1 Overview

The lesion that is left behind in the DNA by AID is the same whether one is talking about SHM or CSR, but the steps that are taken to deal with this lesion result in qualitatively different

outcomes. Not surprisingly, the dependence of SHM and CSR on various different components of the BER and MMR pathways varies accordingly. The contribution that each of these pathways makes to SHM and CSR, whether there is competition for recognition of the U/G mismatch by these pathways, and lastly some models of how these pathways are involved mechanistically in CSR and SHM will be the topics of next section.

1.4.2 BER in mammalian systems

Base excision repair (BER) was first described in *Escherichia coli* as a complex process involving several enzymes that were initiated by cleavage of the N-glycosidic bond creating an abasic site in dsDNA, which is best defined as the apurinic/apyrimidinic (AP) site. As in bacteria, the mammalian equivalent of uracil DNA glycosylase (UNG) is responsible for catalyzing this first step of BER. BER can be distinguished from other DNA repair systems in that it is restricted to recognizing short, single nucleotide gaps in dsDNA. The recognition of a mispaired nucleotide is also limited to modifications that resemble the shape and size of the original nucleotide, such as those arising from deamination of cytosine and adenosine (reviewed in (79)). The key steps of BER can be summarized as follows: (1) base removal, (2) cleavage at the AP site, (3) processing of 3'-O-phosphate group to remove phosphodeoxyribose or phosphoglycolate, thereby liberating the 3'-OH group for the next step, (4) gap-filling by a DNA polymerase, and finally (5) DNA ligation. The players that are involved in each step can vary from species to species, but for the purpose of this thesis we will only be discussing those genes that have been shown to impact SHM and CSR in mammalian systems, namely, UNG2, apurinic/apyrimidinic endonucleases 1 and 2 (APE1 & APE2), and DNA polymerase β (DNA pol β).

1.4.3 Loss of Ung studies in human and mouse

The *Ung* locus encodes two forms of UNG in mouse and human, a mitochondrial isoform (UNG1) and a nuclear isoform (UNG2). These proteins are identical to one another with the exception of their N-terminal import sequences, which arise from distinct promoters and alternative splicing that direct them to either mitochondria, in the case of UNG1, or the nucleus, for UNG2 (80). Loss of UNG2 is thought to cause a severe reduction of CSR and a perturbation of SHM. I have added the clause “thought to cause” because it has to be noted that the UNG ablation in both mouse and human studies disrupted UNG1 as well, and a system that only destroys UNG2 does not currently exist. This caveat notwithstanding, ablation of UNG in a mouse model causes a 90-95% reduction of *ex vivo* CSR and loss of SHM Phase I transversion mutations, while SHM Phase II mutations are not affected (32). Cell lines derived from human patients with UNG deficiency also revealed a severe 99% deficit in *ex vivo* CSR (81). In another report, B cells that were obtained from human patients carrying inactivating mutations in UNG had higher content of uracil in genomic DNA and were not able to remove U from ssDNA substrates *in vitro* (82). Loss of UNG also leads to a moderate decrease in the amount of DSBs in $S\mu$ and $S\gamma3$ and a reduction of DSBs in AID hotspots in primary B cells activated with LPS, a finding that supports the participation of UNG in CSR in the steps immediately following AID deamination (55).

It is worth briefly mentioning that the enzymatic role of UNG in CSR is not without its share of controversy. Active site (D147N, N206V, and H270L) mutants that are attenuated for uracil excision (10^{-2} - 10^{-3}) were shown to rescue the CSR defect in *Ung* deficient mice (34). The mouse equivalent of the P2 patient mutation, F242S, from Imai et al. (81) was also capable of rescuing CSR (34). The former finding has been contradicted by another report that showed the

retroviral overexpression system used to deliver the UNG mutants is able to compensate for the diminished enzymatic activity (83). In this report a double mutant (D147N, H270L) that was incapable of rescuing CSR in both reports was shown not to possess enough uracil excision enzymatic activity, thereby refuting the claims made from the first report that uracil excision is dispensable chemistry in CSR. However, the mutant versions of UNG2 (D147G, and D147N) are better at CSR rescue than WT UNG2 (83), a finding that will be explored in detail in chapter 3.4.

1.4.4 The roles of APE1 and APE2 in CSR

UNG is a member the monofunctional glycosylase class of DNA glycosylases, which prevents it from performing the necessary chemistry to cleave the sugar-phosphate DNA backbone, which is instrumental in creating a single-stranded break (SSB). In mammalian BER this chemistry is provided by APE1 and APE2. The creation of SSBs is a critical intermediate step preceding formation of a DSB, therefore it follows that APE1 and/or APE2 should be expected to play a role in CSR. APE1 is critical for embryonic development, therefore *ape1*^{+/-} B cells, which exhibit defects in DNA repair, were used to study the effect of haploinsufficiency of *ape1* on CSR. Although APE2 does not possess as strong of an AP activity *in vitro* compared to APE1, it was also studied in this experiment. The *ape2* locus resides on chromosome X, so for all the experiments described below, male mice were used, thus providing the necessary haploinsufficiency with which to study CSR. A combined double mutant mouse, *ape*^{+/-}; *ape*^{Y/-}, was used as well. Summarizing results from CSR of 5 different isotypes showed that *ape1* haploinsufficiency causes a modest 20% decrease in CSR. Both the *ape2* single and the double *ape1/ape2* mutant B cells showed a stronger 40% decrease in CSR (84). In addition, a follow-up report demonstrated that DSBs are reduced in APE deficiency, especially in double *ape1/ape2*

mutant B cells (56). These data suggest that APE1 and APE2 proteins are both important in the steps that follow excision of uracil by UNG on the way to creating a SSB. The fact that a decrease in CSR can be linked to loss of just one allele of either APE1 or APE, suggests loss of both alleles might result in an even further reduction of CSR. It should be mentioned that another report found CSR from B cells that are null for *Ape2* was unperturbed, and in the same report, knock down of APE1 in the B cell lymphoma CH12.F3, did not affect CSR (85). Why a discrepancy exists between this report and the findings from Schrader et al. is not clear. However, loss of APE2 in activated primary B cells did cause a severe decrease in the frequency of SHM mutations while not affecting the position of mutations in AID hotspots (85). It's also possible that the CSR phenotype of APE deficiency is quite mild and therefore easy to miss if the sensitivity of an assay is off.

1.4.5 DNA Polymerase β

It is somewhat surprising to find that unlike deficiencies in UNG or APE proteins, which cause reductions in the efficiency of CSR, ablation of the BER polymerase, DNA pol β causes the opposite. Primary B cells deficient in DNA pol β show increased CSR to IgG2a, IgG2b, and IgG3 isotypes, and for reasons that are not entirely understood the effect is most pronounced in cells that were induced to switch to IgG2a (86). The finding that DSBs are increased in the absence of DNA pol β indicates that perhaps its primary role is to repair DSBs (56, 86). It was postulated in this report that the reason CSR proceeds to completion in primary B cells is that DNA pol β is overwhelmed by the sheer number of AP sites initiated by AID.

To conclude this section, let it be known that data from several independent labs all seem to concur that key proteins from BER are crucial mediators and modulators of CSR, and to a

lesser extent of SHM as well. The participation of BER in both CSR and SHM lends further support to the widely accepted DNA deamination model, originally proposed by Michael S. Neuberger. However, as you will see, BER does not operate in isolation during SHM and CSR, nor can it account for the full spectrum of SHM or CSR in B cells. To fill-in this veritable void in understanding MMR offers a clear insight.

1.4.6 MMR in mammalian systems

MMR is perhaps best characterized in the *Escherichia coli* (*e.coli*), using purified proteins in *in vitro* settings (87). In this system the MutS homodimer first recognizes a mismatch, followed by recruitment of the MutL homodimer, and activation of the endonuclease, MutH, which recognizes and incises an unmethylated GATC site on the newly synthesized, mismatch-containing strand of DNA (reviewed in (88)). In this system the unmethylated restriction site serves as the discriminating mark to distinguish between the two DNA strands. Unfortunately, outside of *e.coli* and closely related species of bacteria, this system does not seem to exist, and concurrently, no MutH homologue has been identified in any of the other prokaryotic or eukaryotic systems.

The lack of a discriminating mark with which to identify the error-containing strand, and more importantly the absence of an endonuclease function in Eukaryotic MMR has long puzzled scientists. One possible explanation is that strand discontinuity in the newly synthesized strand serves as the discriminating mark, such as those arising from discontinuity at 3' ends or termini of Okazaki fragments (88). Yet another conundrum stems from the requirement of the 5'→3' exonuclease activity of Exonuclease I (ExoI) in processing both 3'- and 5'- directed MMR. It is not understood how ExoI, which does not possess defined 3'→5' exonuclease activity could

orchestrate 3'- directed MMR. Moreover, without possessing any intrinsic endonuclease activities of their own, it was not understood how either the MutS α heterodimer (composed of MSH2 and MSH6) or the MutL α heterodimer (composed of MLH1 and PMS2) could provide the discriminating mark, which also serves as the entrance point for Exo I (reviewed in (89)). Salvation may have come from studies conducted in *Saccharomyces cerevisiae* and human where a latent endonuclease activity in the MutL α heterodimer, which was dependent on MutS α , PCNA, RFC, and ATP (90, 91), was described. The same studies also offered an explanation for

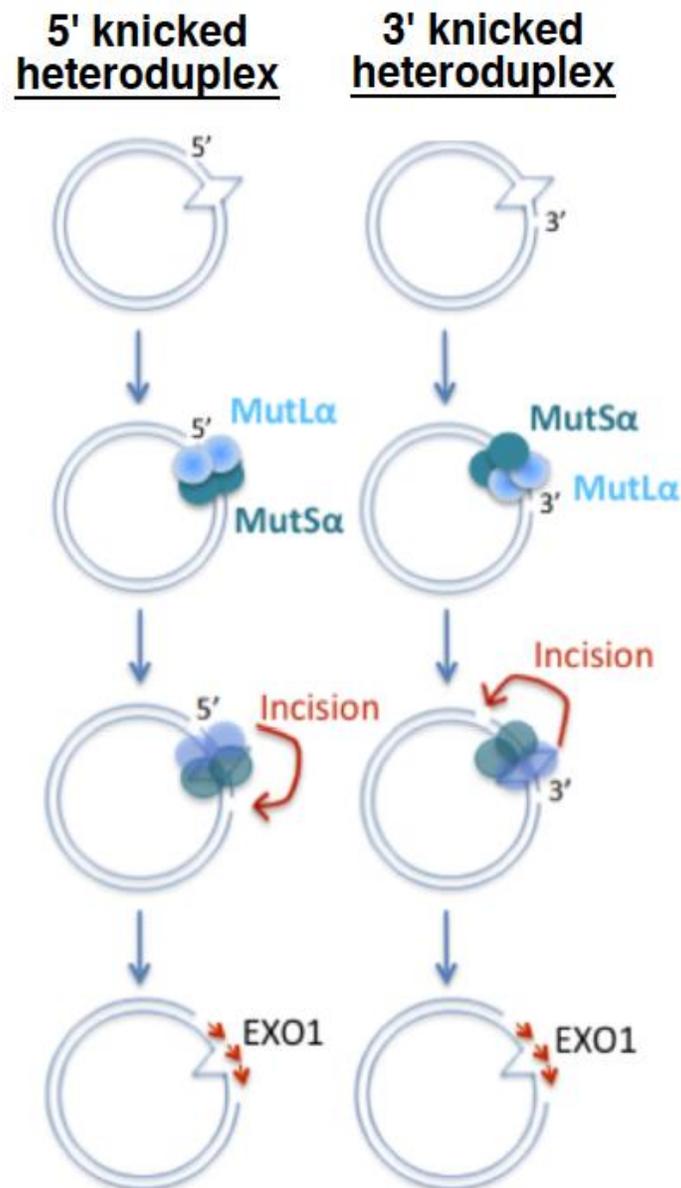


Figure 3. **Model of 5'- and 3'- directed MMR.** The MutS α heterodimer (Msh2/Msh6) recognizes a mismatch and recruits the MutL α heterodimer to the nearest strand discontinuity (nick). When the strand discontinuity is located 5' of the mismatch, Exo I can process past the mismatch via 5' \rightarrow 3' exonuclease activity. When the nick is located 3' of the mismatch MutL α makes an incision 5' of the mismatch via its putative endonucleolytic function, thereby providing the strand discontinuity. Now Exo I can proceed processing with 5' \rightarrow 3' exonuclease as usual.

how 3'-directed MMR can work in the absence of a discriminating mark, 5' of the mismatch, which is summarized in figure 3.

As is the case with BER, there are several proteins involved in mammalian MMR, with some members having distinct roles in one biological process, while others may not. We will only be discussing in detail those members that have a direct impact on SHM and CSR, namely mutS homolog 2 (Msh2), mutS homolog 6 (Msh6), exonuclease I (ExoI), mutL homolog 1 (Mlh1), and post-segregation meiotic factor 2 (Pms2).

1.4.7 Msh2 and Msh6 and Exo I

Together, Msh2 and Msh6 form the MutS α heterodimer, and disruption of one gene leads to diminished stability at the protein level of the other. In studies of *ex vivo* CSR ablation of these genes caused between a 40-80% reduction, varying among different isotypes, and this has been well documented by several labs (92-96). Serum Ig levels were reduced compared to WT in Msh2 and Mlh1 mouse knock outs and varied between a 50-90% reduction among different isotypes (93). When S μ -S α junctions, which represent the recombination point of S μ DNA with S α DNA from a downstream isotype (see chapter 3.6), were analyzed in Msh2 and Exo I deficient mice, compared to WT, scarcely any microhomology could be found at the junction (94, 97, 98). This was different than WT primary B cells, where typically nearly 2/3 of all junctions display some degree of microhomology. This analysis demonstrated that these MMR proteins were likely to be directly involved in the CSR reaction. The use of microhomology-versus direct-based end joining will be discussed in chapter 1.5.2.

Interestingly, Msh2, Msh6, and Exo I play a pivotal role in generating Phase II SHM mutations. Mutations occurring away from the U/G mismatch comprise about half of all SHM

mutations (reviewed in (18)) and generally affect A:T base pairs. It is these A:T mutations that are severely attenuated in the absence of the MutS α heterodimer or Exo I. (32, 93, 99). The finding that Msh2-Msh6 stimulates the low fidelity polymerase, pol η (100), loss of which is known to impact A/T mutations in SHM (101, 102), mechanistically links the MutS α with the generation of SHM Phase II mutations. Thus, a striking difference exists between Msh2, Msh6, Exo I, and the other MMR proteins, and that is the role that they play in SHM is exclusive to them.

1.4.8 Mlh1 and Pms2

Similar to what was described for Msh2 and Msh6, Mlh1 and Pms2 constitute the MutL α heterodimer, and they mutually influence the stability of each other. Importantly, the C-terminus of MLH1 is important for stabilizing PMS2 (103). Single deletion of either gene causes a reduction of CSR (92, 96, 104), which in accordance with the studies of loss of Msh2, varies, depending on isotype. Similar to Msh2 ablation, IgG2a and IgG2b CSR show a greater dependence on Mlh1 and Pms2 than the other isotypes (95, 96). When S μ -S α junctions from Mlh1 and Pms2 deficient mice were analyzed for microhomologies there was a statistically significant increase in the length of microhomologies compared to wild type controls (94, 98). Curiously, this trend is completely opposite to what was found for Msh2 and ExoI microhomologies. This suggests that the relationship between these different MMR proteins in CSR is not completely overlapping. Coupling this information with the apparent lack of a role for Mlh1 and Pms2 in SHM indicates that in spite of the decreased CSR profiles that are similar for ablation of all MMR proteins, Msh2 may function independently of Mlh1 and Pms2 in CSR.

Although an endonucleolytic function has been ascribed to Pms2 (104), it should be noted that E702K mutation reduces the protein to 20% of WT levels. Such a drastic change makes it impossible to deduce whether the reduction seen in CSR is simply not due to the reduced level of PMS2 protein and not because of its putative enzymatic function. Mice heterozygous for Mlh1 or Msh2 are 50% reduced for protein levels and already display a CSR defect, albeit a modest one, around 30% (personal communication, Janet Stavnezer). It is possible that an 80% reduction of PMS2, regardless of the E702K mutation, affects CSR to the same extent as mice that are null for Pms2. The explanation provided by the authors is that the loss of enzymatic function does not cause infertility of the mice harboring the mutation, even though CSR efficiency is diminished. It is true, not just with Pms2, but with Mlh1, and Msh5 as well, defects in male fertility, especially those associated with azoospermia (105) can be traced to defects in these proteins, however, it is not clear whether a 80% reduction in the protein level of WT Pms2 would cause infertility. As a matter of fact, Pms2^{-/+} animals, which are not sterile and have a 50% reduction of PMS2 protein are bred together to make Pms2^{-/-} mice. We know that a Pms2^{-/+} mouse displays a CSR defect and has a 50% reduction in of PMS2, so it's not clear whether the CSR defect in the Pms2^{E702K/E702} mouse, which has an even further reduction of PMS2 protein level, can be attributed to the endonucleolytic function of PMS2. It's possible that the authors may have simply misinterpreted the results

In a follow-up paper, the authors have generated a mutant form of Mlh1 harboring a G647A mutation that still allows for recognition of the U/G mismatch by the MutS α heterodimer but reduces CSR substantially (92). This mutation does affect the steady-state level of Mlh1, but destroys the ATPase activity of Mlh1, and because the endonucleolytic function of Pms2 in human and yeast systems was shown to be dependent on ATP (90, 91), the authors are

attempting to further the notion that Pms2 endonucleolytic function is important in CSR. It is not clear, however that the ATPase activity of Mlh1 in this study can be linked to Pms2 function during CSR.

Regardless of whether or not Pms2 actually encodes endonucleolytic function, both Mlh1 and Pms2 contribute substantially to the generation of DSBs during CSR. Compared to a WT background, Mlh1 shows a 75% reduction in breaks at S μ , while Pms2 shows an even greater loss, around 80-95% (106). Of all the MMR proteins, it seems that Pms2 contributes most to the generation of DSBs, as the DSB profile of Pms2 deficiency approaches what was found for AID deficiency (98).

In summary, both Mlh1 and Pms2 are critical mediators of the CSR reaction. The lack of involvement in SHM and the qualitative difference in microhomology usage at S μ -S α junctions distinguishes them from Msh2 and Msh6, and suggests they may function independently from these other MMR proteins during CSR. Msh2, Mlh1, and Pms2 are also instrumental in generating DSBs. However, whether or not a true endonucleolytic function, as it pertains to the generation of DSBs, can be ascribed to Pms2, remains to be seen. More elaboration of the endonucleolytic function of Pms2 in conjunction with different models of CSR is discussed in section 1.4.10.

1.4.9 Competing for access to the U/G mismatch

One thing is for certain, ablation of both Ung and Msh2 wipes out CSR to all isotypes, and reduces SHM mutations to one category, C \rightarrow T and G \rightarrow A transitions (93). The latter finding demonstrates rather nicely that without recognition of the U/G mismatch by either pathway, the lesion is simply replicated over, highlighting the massive number of AID deaminations that must

occur during SHM and CSR reflected in the impressive increase of transition mutations (93). Where SHM and CSR are concerned why is it that two different DNA repair pathways that process the same substrate, in this case a U/G mismatch that results from AID deamination, result in seemingly different outcomes? Perhaps more to point, why are two separate DNA repair pathways that are geared towards accomplishing the same goal needed for the full spectrum of these Ab diversification processes? In the physiological setting, it is difficult to ask whether these different pathways are competing for access to the U/G mismatch, but one model has been proposed (107). The creation of DSBs seems to be restricted to the G1 phase of the cell cycle, and paradoxically, MMR proteins, and UNG2 are upregulated towards the end of G1, even though they are presumed to participate during G1. If the U/G lesion is created in G1, then which pathway arrives at the scene first? Or do the BER and MMR pathways arrive simultaneously? The model from (106) postulates that MMR arrives at the U/G mismatch first, followed by BER. We will show preliminary data that begins to lightly tackle question in chapter 3.4, but for now, the next section will be devoted to describing the most accepted model of how BER and MMR contribute to DSB formation in CSR and to a lesser extent in SHM.

1.4.10 Models of DSB formation and A:T basepair mutations

Perhaps the best current explanation of how both BER and MMR pathways contribute to, rather than compete in SHM and CSR, can be found in the model proposed in Stavnezer et al. (108). This model postulates that if U/G mismatches on opposing strands of DNA are processed by BER (UNG and APE1/APE2) into SSBs and are close enough to each other in the DNA duplex then a DSB will spontaneously form. This type of DSB does not require any

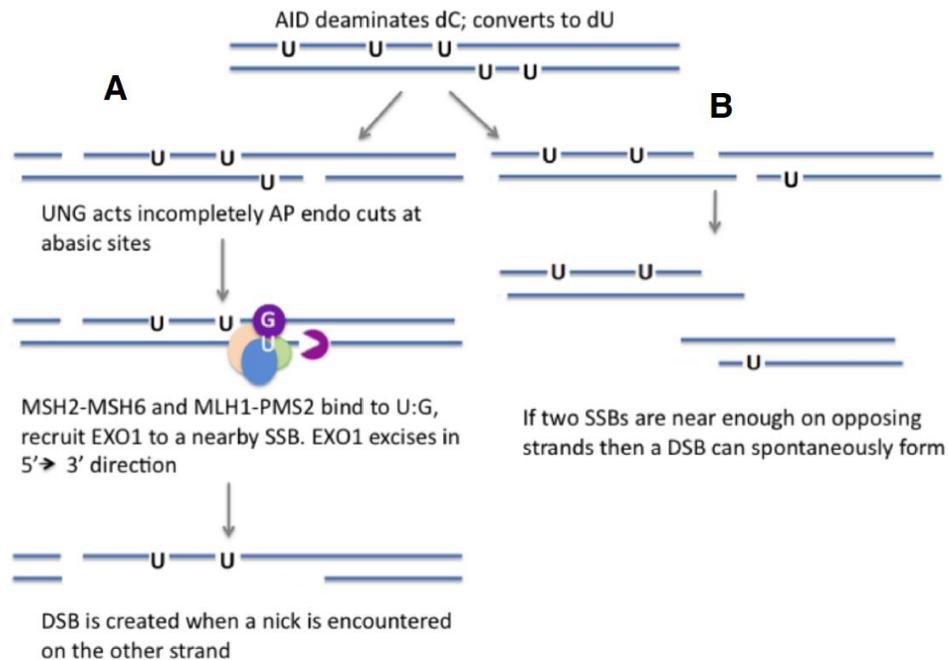


Figure 4. **Model of DSB formation.** AID deaminates dC creating a U/G mismatch. Mismatches are processed by UNG and APE 1 or 2 into single-stranded breaks (SSBs) **A)** MMR can be recruited to a U/G mismatch. The MutS α (Msh2/Msh6) and MutL α (Mlh1/Pms2) heterodimers recruit Exo I to a nearby SSB and processing begins in the 5' to 3' direction. Once a SSB is encountered on the other strand a DSB forms. **B)** When SSBs are close to each other on opposing strands of the DNA duplex a DSB can form spontaneously without any processing by MMR.

participation of MMR, and explains why in the absence of MMR proteins, a substantial amount of CSR still occurs (fig. 4B).

Conversely, when SSBs are far away from each other, the strands of the intervening DNA duplex cannot spontaneously melt apart, in which case processing by MMR is crucial. If the SSB is 5' of the mismatch, then MMR recruits and loads Exo 1 at the SSB and begins 5' → 3' processing, eventually encountering a SSB on the other strand and creating a DSB with long overhangs (fig 4A). This type of DSB would be expected to arise from regions in S_{μ} where there is a paucity of AID hotspots. This scenario is artificially enhanced in the S_{μ} TR KO mouse where a large portion of S_{μ} has been deleted that removes the majority of AID hotspots. The few hotspots that are left are not close to one another like they are in the S_{μ} TR region. Interestingly when MMR deficient mice are crossed with the S_{μ} TR KO mouse CSR is nearly abolished. The explanation for the lack of CSR is that because AID hotspots are far away from each other, and because MMR is not present to process SSBs, very few, if any DSBs are made (98). This is a relatively simple model that postulates MMR is only involved in processing pre-existing SSBs and does not actually create SSBs during CSR. In this model all SSBs are the result of BER processing of AID instigated U/G mismatches (combined action of UNG and APE1 and APE2). The fact that DSBs still occur in the absence of UNG suggests that MMR (or an MMR-related process) can create SSBs, but without a clearly defined endonuclease function, it's likely that other proteins would have to be involved. It could be that other glycosylases can work in lieu of or independently of UNG in creating some of the SSBs during CSR. This is certainly true in the Phase II of SHM, where a similar mechanism of introducing SSBs is probably required for the generation of A:T mutations, which has to be independent of Ung.

At first glance, it would appear that the model described by Stavnezer et al. should apply for SHM as well, but there are some key differences that do not fit. In Phase II of SHM several mutations are introduced at A:T basepairs. These mutations are dependent on Exo I and the MutS α heterodimer and curiously are not affected by loss of UNG. If MMR does not have the intrinsic ability to make SSBs (as proposed in the model from figure 4), and if UNG is not required for initiating SSBs during Phase II of SHM, then how is a SSB made that allows for loading of Exo I? The putative endonucleolytic function of the MutL α heterodimer would be an attractive candidate to supply the SSB required by Exo I, except that neither Pms2 nor Mlh1 deficiency affects mutations at A:T basepairs. Since APE 1/2 make SSBs after UNG has created an AP site, a role in creating the SSB that is required for Exo I during Phase II of SHM is probably unlikely. This leaves open the possibility that other DNA glycosylases participate in Phase II of SHM (fig. 5) (see chapter 1.6) It could be that a dependence on other DNA glycosylases is a feature that is unique to SHM, but the fact that MMR proteins (Msh2 and Exo1) are important for both CSR and SHM begs the question whether other glycosylases could also participate in CSR. If this were the case then one should simply be able to test for the effect that deleting these glycosylases would have on CSR. In this regard the DNA glycosylase Mbd4 is particularly interesting because it was originally identified as a binding partner of Mlh1 (see chapter 1.6).

In summary, BER and MMR contribute to the formation of DSBs during CSR, and are required for the full spectrum of mutations that are generated during SHM. Although the initiating lesion (the AID-induced U/G mismatch) is the same in both SHM and CSR, very different outcomes are achieved by varying participation of different BER and MMR components. Currently a unifying model that links SHM and CSR mechanistically does not exist.

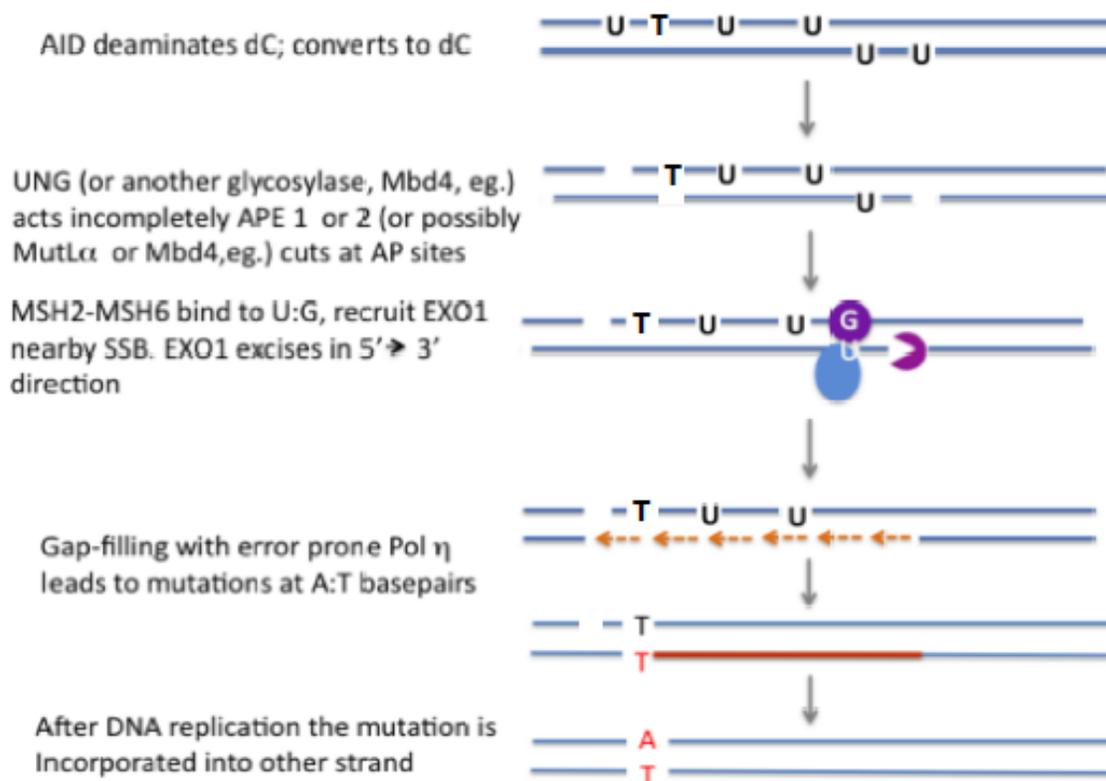


Figure 5. **Model of SHM Phase II mutations.** Various steps are described in the generation of mutations at A:T basepairs. The conventional model describes UNG as the sole DNA glycosylase, and loading of Exo I is thought to occur at an AP sites which can be processed by APE 1 or 2, but A:T mutations are not perturbed in the absence of UNG. Therefore we speculate the potential involvement of Mbd4 as an alternative DNA glycosylase for promoting AP sites. Other potential nicking enzymes are mentioned, including Mbd4, which has a putative lyase function and the MutL α heterodimer, which also has a putative endonucleolytic function.

The model in figure 4 is useful for understanding the formation of DSBs, but cannot be extended to SHM.

One obvious difference between SHM and CSR is that CSR is dependent on the creation of DSBs. Because DSBs are recognized in the cell as DNA damage, in the next section special attention will be given to various factors in DNA repair, which are required for completing the final stages of CSR. Almost no overlapping function exists for these factors in SHM, but some of these factors do have overlapping function in V(D)J joining, which also depends on DNA cleavage, which I will now briefly revisit.

1.5 DSBs breaks recruit DNA repair factors

1.5.1 Overview

A devoted system of detection and repair exists in the cell to protect the genome against a variety of DNA damaging sources. DSBs can result from external insults, such as ionizing and UV radiation, and from insults within, like free radicals that are released as the by-products of intermediary metabolism. Ab diversification in mammalian B cells, both in the case of V(D)J joining and CSR, takes advantage of a form of programmed DNA damage, which exploits many of same factors that are involved in detecting and repairing general cellular DNA damage. In this regard it is interesting to note that some form of immunodeficiency often accompanies many human diseases that are associated with defective DNA repair (109).

DSBs, whether self-inflicted or not, can be repaired in the cell by either one of two repair mechanisms. If a sister chromatid or homologous chromosome template is available, as during the S and G2 phases of the cycle, or during mitosis and meiosis, then homologous recombination

(HR) predominates and will result in high fidelity repair of the DSB. On the other hand, if the DSB should occur during the G1 phase of the cell cycle, as is the case for V(D)J joining and CSR, then the cell is equipped with non-homologous end joining (NHEJ), a more error-prone form of repair which directly pastes back together broken DNA ends, often times resulting in the loss or gain of nucleotides near the recombination site.

Studies of irradiation- induced DSBs demonstrated that one of the first (if not the first) players to arrive on the scene is the histone variant, H2AX. Within minutes H2AX gets phosphorylated by a PI (3)-like kinase, ataxia telangectasia mutated (ATM), which acts as DNA damage signal transducer, and recruits other DSB repair proteins, such as those constituting the MRN complex (Mre11, Rad50, and Nbs1) and BRCA1, forming DNA repair foci that are readily detectable by immunofluorescence (110). The MRN complex has also been implicated in recruiting ATM to broken DNA sites and is required for full activation of ATM (111). Relatedly, patients with hypomorphic mutations in Mre11 and Nbs1 show a disease phenotype similar to that of AT (112, 113). In AID KO mice DNA repair foci containing such proteins as γ -H2AX and Nbs1 do not form at the *Igh* locus upon CSR induction, which confirms that AID is responsible for creating DSBs (114).

1.5.2 C-NHEJ and A-EJ

Most (if not all) of the V(D)J DSBs that are introduced by RAG1 and RAG2 are repaired by NHEJ. In classical NHEJ (C-NHEJ) X-ray repair cross-complementing protein 4 (Xrcc4) functions alongside DNA ligase IV (Lig4) in the final step of resolving breaks that are blunt (or nearly blunt) ended). Ablation of either one of these proteins wipes out V(D)J joining, but surprisingly, CSR is only reduced by about half, which revealed that there is an alternative

pathway (A-EJ) for resolving DSBs that are created by AID. This pathway makes use of DSBs that underwent extensive resectioning resulting in long overhangs as evidenced by the analysis of S μ -S α junctions from activated primary B cells containing long microhomologies (115). This report was complemented by a study of Lig4 ablation in the B cell lymphoma CH12.F3 which demonstrated a complete loss of direct joins when S μ -S α junctions were analyzed (116). A-EJ is dependent on cTIP, and initially DNA ligase III (Lig3) and x-ray repair cross-complementing protein 1 (XRCC1), were suggested as well, but recent reports this year have demonstrated in B cells that neither XRCC1, nor ligase III are important for A-EJ (117, 118).

Other important members of C-NHEJ, such as Ku70/80 heterodimer, DNA dependent protein kinase (DNA-PK), artemis, and cernunnos (XLF), in general seem to have more crucial roles in V(D)J joining than in CSR, perhaps because A-EJ does not seem to contribute as a backup-up pathway to C-NHEJ in V(D)J joining. With the exception of Ku70/80, CSR is not as dependent on other C-NHEJ factors (119, 120). Although Ku70/80 may cause proliferation defects, which could in turn affect the rate of CSR, Ku70/80 deficient cells appear to undergo several rounds of cell division and yet fail at *ex vivo* CSR (121). A mouse engineered to be null for DNA-PK showed diminished CSR to all isotypes except γ 1 (122), and in this regard some redundancy between DNA-PK and ATM may exist (123). A scaffolding function for DNA-PK has also been proposed, as CSR was found to be normal in SCID mice harboring a hypomorphic mutation in DNA-PK that results in a truncated protein lacking catalytic function (DNA-PKcs) (124). Clearly, Artemis seem to be dispensable for CSR (125, 126), and as mentioned before ablation of Lig4 or XRCC4 only reduces CSR 2 fold, unlike in V(D)J which is wholly dependent on almost all C-NHEJ factors. Although XLF deficiency in human results in immunodeficiency-

associated disease (127), it was found not to be essential for either V(D)J joining or CSR in a mouse model (128).

Curiously, one study makes the claim that microhomology based end-joining catalyzed by A-EJ is dependent on MMR, since combined deficiency of Msh2, EXO I, or Mlh1 with Lig4 reduces CSR further than what is seen for Lig4 deficiency alone (129). This report also cites additional roles for Mlh1 in A-EJ as combined deficiency of Mlh1 and Lig4 produced $S\mu$ - $S\gamma$ 1 junctions with more direct joins than Lig4 deficiency alone (129). The opposite holds true for Mlh1 deficiency alone, where $S\mu$ - $S\gamma$ 1 junctions have longer microhomologies and fewer direct joins when compared to WT primary B cells (94, 95, 98). More on the impact of MMR deficiency on $S\mu$ - Sx junctions can be found in chapters 1.4.7 and 1.4.8.

Whether C-NHEJ or A-EJ is used to resolve DSBs that are instigated by AID in the *Igh* locus, the fate of the broken ends of DNA is not limited to *bona fide* CSR. For example, theoretically a DSB could be relegated in the same spot with no loss of intervening DNA and hence no CSR. Two DSBs within an S region can be relegated, with or without loss of intervening DNA, creating an intra-switch region deletion, resulting once again in no CSR. This phenomenon is not uncommon in $S\mu$ (130), and is elevated in the absence of the DNA repair protein p53-binding protein 1 (53BP1) (131). DSBs in the *Igh* locus can also result in translocations with other loci, although this event occurs rarely. With regard to the choice of employing C-NHEJ versus A-EJ to repair DSBs, 53BP1 protects against ATM-dependent resectioning of DSBs, and is therefore thought to shuttle DSBs into the C-NHEJ pathway (132). The mechanism that promotes *bona fide* CSR, that is to say, the joining of DSBs from two distal S regions, ahead of the other options mentioned above is not currently understood. Our lab has some data to suggest that 53BP1 function may be involved in directing this mechanism.

1.6 Mbd4

1.6.1 Overview

Two intriguing pieces of evidence originally focused our attention on a putative DNA glycosylase called Methyl CpG-binding domain Protein (Mbd4). First, a surprising role in DNA demethylation was recently proposed as the two-step erasure of Me⁵CpG, coupling the deaminating activity of AID with the T/G mismatch DNA glycosylase enzymatic function of Mbd4 (133). Subsequent reports demonstrated increased methylation of CpG sites in primordial germ cells (PGC) from mice that were deficient in AID in a process that would possibly link deamination of Me⁵CpG by AID to mismatch recognition and excision of the T/G mismatch by Mbd4 (134, 135). Secondly, Mbd4 has also been shown to interact with Mlh1, an MMR protein whose participation in CSR is well established (136). As a putative endonuclease and U/G mismatch DNA glycosylase, Mbd4 is a strong candidate to supply the nicking activity not provided by UNG and APE1/2 that models of CSR and SHM (see section 1.4.10) suggest is required. Indeed, the solved crystal structure of the C-terminal domain of Mbd4 (glycosylase domain) showed that Mbd4 belongs to the helix-hairpin-helix superfamily of bacterial DNA repair Nth/MutY/MIG/UVendo glycosylases/lyases (HhH GPD) (137).

1.6.2 Catalytic properties of Mbd4

As a member of the HhH-GPD family Mbd4 has the potential to be a bi-functional glycosylase, that is, in addition to excising the nucleobase from DNA, it can cleave the sugar-phosphate DNA backbone via lyase activity. Regarding Mbd4 lyase activity, the data have not been consistent. While one group found endonuclease and Mlh1-binding activity for Mbd4

(136), thereafter attributing to it a likely role in eukaryotic MMR analogous to the one fulfilled by MutH in the *e.coli* system of MMR, another group cited lack of endonuclease activity (138), instead documenting strong T/G and U/G mismatch DNA glycosylase activity. Additionally, like the bacterial DNA glycosylase *Mig*, Mbd4 has a tyrosine residue at the corresponding K120 position of *E. coli* endonuclease III (*Nth*). This basic residue is conserved among all *Nth* proteins and is crucial for AP lyase activity, therefore an endonuclease function for Mbd4 was dismissed (137, 138). However, it is clear that owing to distinctive homology found within MIG members (Mth, PaMig, APE0875, and Mbd4), AP lyase activity is present, especially when the appropriate substrate is used. This was demonstrated by uncoupling the DNA mismatch glycosylase activity from the lyase activity of *Pyrobaculum aerophilum* (*Pa*)*Mig*, using substrates that contained an AP site paired against guanine. (139). To my knowledge an AP site substrate was never tested directly in an Mbd4 nicking assay, only mismatched DNA substrates have been used, which would only address the glycosylase activity of Mbd4. Another more recent crystal structure of the Mbd4 glycosylase soaked with a mismatched DNA substrate in the active site described a chemistry that separates Mbd4 from other members, namely *Hogg* and *Nth* (140). The active site of Mbd4 is different than *Hogg* and *Nth*, and the glycosylase reaction proceeds via a unique intermediate that was not present in the crystal structures of HOGG and NTH. Therefore, although the question of whether or not Mbd4 actually possesses any endonucleolytic function is still up to debate, ruling out an Mbd4 lyase function based on potentially different chemistry from other members of the HhH-GPD family seems premature at this juncture.

While there is no clear consensus on the lyase function of Mbd4, several labs have confirmed that a DNA glycosylase function is present (137, 138, 140, 141). In all of these

studies, *in vitro* nicking assays have demonstrated that Mbd4 is efficient at removing U and T paired opposite G in double stranded DNA oligo substrates, and it is slightly more efficient at U removal, which is similar to what has been described for TDG (138).

1.6.3 **Mbd4 as a part of MMR**

Since its discovery, the question as to whether or not Mbd4 is an integral part of the MMR system remains controversial. Due to the ability to bind 5me-CpG/TpG mismatches, a putative endonuclease function, and the ability to interact with Mlh1, it was suggested that as a likely member of the MMR pathway (in a role akin to that of Mut H in *e.coli* MMR), Mbd4 probably functions as a tumor suppressor gene (136) in mammalian cells. This was demonstrated by overexpressing the C-terminal fragment of Mbd4 (glycosylase domain), which was proposed to act as a dominant negative leading to defective MMR and causing microsatellite instability (MSI) in human cell lines (136). MSI arises from down-regulated or aberrant expression of members of the MMR pathway (Msh2, Mlh1, Pms2, Pms1), presenting in 50-90% of hereditary non polyposis colorectal cancer (HNPCC) cases and 15% of sporadic colon cancer cases (142, 143). If Mbd4 functioned as a tumor suppressor gene like the other members of MMR, then one might expect to find mutations in Mbd4 in some of the remaining 10-50% of cases of HNPCC. Contrary to this, Bird et al., demonstrated that mutations in an MSI-sensitive Poly A tract (A₁₀) of exon 3 of Mbd4 are indeed found in 43% colorectal cancer cases, but in all of these cases there was already an apparent MMR repair defect (144). Sequencing the rest of Mbd4 did not reveal any additional mutations outside of the poly A tract in any of the 20 MSI-bearing tumor samples. This suggested that like another cohort of genes that are sensitive to MSI, such as BAX, TGFβRII, and IGFIIR, the A₁₀ tract in Mbd4 is the target of defective MMR, not the cause of it,

and that the main function of Mbd4 may instead be to counteract mutability of Me⁵CpG. In support of the latter, two independent groups, each crafting their own KO mouse, have demonstrated that loss of Mbd4 leads to a 2-3 fold increase in C → T and G → A transition mutations (145, 146). In both reports, the bulk of the mutations occurred at the CpG dinucleotide motif, which in many cases was methylated. This is not as severe as the 100-500 fold augmentation of such mutations in mice that have one or another form of defective MMR with an associated MSI phenotype (147, 148). Additionally, unlike inactivating mutations in members of the MMR pathway, mutation of Mbd4 alone in the mouse KO systems was not enough to promote tumorigenesis (147-149). However, when crossed to a mouse that is heterozygous for the APC allele, Mbd4 KO mice do show accelerated loss of heterozygosity (LOH)-associated-tumor progression, suggesting a role in modulating cancer progression (145, 146). However, the point is that loss of Mbd4 in current mouse KO models (145, 146) did not phenocopy loss of other MMR genes that are known to be tumor suppressing.

1.6.4 A role for Mbd4 in antibody diversification

We sought to determine whether based on the following lines of evidence Mbd4 could participate during CSR. (1) The original identification of Mbd4 as an Mlh1 -binding and -stabilizing protein gives reason to suspect an involvement in CSR. (2) The suggestion that other DNA glycosylases may play a role in SHM (and possibly CSR) is intriguing considering that Mbd4, like Ung, possesses strong U/G mismatch DNA glycosylase activity. Lastly (3), some models of CSR and SHM require an endonucleolytic function that is associated with creating SSBs that are important for loading Exo I. Mbd4 is related to a subfamily of HhH-GPD proteins called *Mig* of which one member, *Pa-mig* has demonstrated distinct lyase activity that is

decoupled from its DNA glycosylase function (138). Strong homology in the conserved domains of *Mig* family members suggests that perhaps this could apply for Mbd4 as well. These three pieces of information make a compelling case for studying the role of Mbd4 in CSR.

Unexpectedly, loss of Mbd4 impacts the stability of several MMR proteins, including Msh2, Msh6, Pms2, and Mlh1 (150). The Mbd4 KO cell line described in this thesis also shows a reduction of Msh2 and Mlh1 proteins (see figs. 17A, 17B). This somewhat complicates the study of a direct role of Mbd4 in CSR (and SHM), as you will soon see. However, the finding that Mbd4 impacts CSR, regardless whether directly or indirectly provides new insight into the mechanism behind Ab diversification in B cells.

CHAPTER 2

MATERIALS AND METHODS

2.1 Construction of Mbd4 Targeting vector

To create 5' and 3' homology arms, 2.85Kb and 5.34Kb regions, respectively, were PCR amplified from CH12.F3 genomic DNA using a 30 cycle amplification scheme (15 sec 94°C, 15 sec 62°C, 5 min 72°C) with Taq polymerase (MBI Fermentas). For the 5' homology arm, sense, 5'-GGGAAGCCATGGAAAACATC-3', and anti-sense, 5'-TGTCTTGGGAAAGGCCAACT-3' primers were used. For the 3' homology arm, sense, 5'-GTCGACATTGGCTCTGAGGAATTTTAGAT-3', and anti-sense, 5'-GTCGACAAAGATGAAAAAGGAGTTGGAG-3' primers were used. All primers were from IDT. PCR products, for the 5' and 3' fragments, were cloned into the TOPO-TA vector (Invitrogen) and sequenced. The 5' homology arm was digested with Eco R1 and cloned into a unique EcoR1 site of a pLNTK vector, which was altered to contain a single Eco RI site (fig. 6). The other Eco RI site was destroyed by PCR mutagenesis (Pfu Ultra) using sense, 5'-GCCTCTTGCGGGGACTTCTACCGGGTAG-3', and anti-sense, 5'-CTACCCGGTAGAAGTCCCCGCAAGAGGC-3' primers (fig. 6). Separately, the 3' arm homology, engineered with Sal I tails, was digested with Sal I out of the TOPO-TA vector and cloned into the Sal I site of pLNTK (fig. 7). The two separate pLNTK molecules, one containing the 5' homology arm, and the other the 3' homology arm, were then cut with Spe I, and religated to generate the complete targeting vector with both homology arms in the physiological orientation (fig. 8). Single mutations arising from PCR error in both 5' and 3' homology arms were sequentially reverted using a site-directed PCR mutagenesis 14-cycle program (15 sec 94°C, 10 sec 62°C, 25 min 68°C) with Pfu Ultra (Stratagene). To revert 5' homology arm mutation, sense, 5'-ACGCCTTTAAGCTTTAGTTTCTGAAAACCTTAAGGAATAC-3', and anti-sense, 5'-GTATTCCTTAAGTTTTTCAGAACTAAATCTTAAAGGCGT-3' primers were

used. To revert 3' homology mutation, sense, 5'-GCTATGAGATGAGATAAGGCTGCAGGGCAGGGCCAC-3', and anti-sense, 5'-GTGGCCCTGCCCTGCAGCCTTATCTCATCTCATAGC-3' primers were used. PCR products were digested with Dpn I (Invitrogen) to destroy the parent template and cloned into DH5-a competent cells (Invitrogen). One clone was identified by restriction fragment analysis.

2.2 Selection of Mbd4 targeted CH12.F3 Clones

CH12.F3 (2×10^6 cells) were electroporated (Amaxa Nucleofactor) with targeting vector (1mg) extracted with phenol/chloroform and linearized with Pvu I. Cells were allowed to recover for 24 hours, seeded into 96-well plates, and selected in G418 (750 μ g/ml) and gancyclovir (4 μ M). Following 10 days of drug selection, individual colonies were picked and allowed to expand for 2-3 days at which point a portion of the cells were frozen down for long-term storage; the rest was further expanded for another 2-3 days. Genomic DNA from each colony (3×10^6 cells) was prepared for Southern analysis. Cells were lysed in lysis buffer (1mM EDTA, 100mM NaCl, 0.5% SDS, and 10mM Tris), treated with RNase A (MBI Fermentas) for 3 hours at 37°, and then digested with Proteinase K (MBI Fermentas) overnight at 55°. Spooled DNA was allowed to rehydrate overnight and targeted clones were identified by Southern hybridization.

2.3 Southern analysis

To identify CH12.F3 clones targeted at the Mbd4 locus, DNAs were digested with Bam HI overnight and separated by 0.8% agarose gel at 40mA overnight. Gels were transferred to a nylon membrane (Biorad Zeta-probe 162-0197) for 16-24 hours and hybridized with probes. Probes were prepared using random hexamer

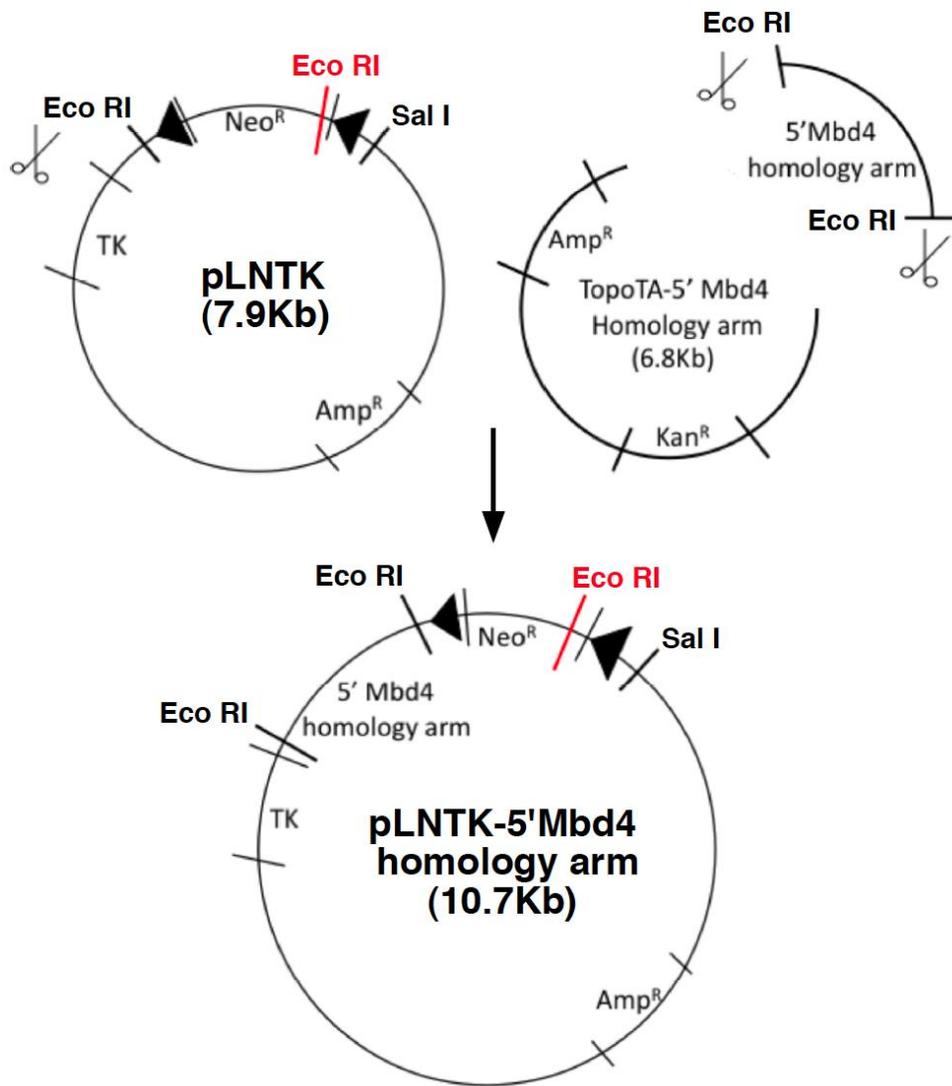


Figure 6. **Construction of pLNTK-5' Mbd4 homology arm.** To facilitate cloning of 5' Mbd4 homology arm an EcoRI site from pLNTK was destroyed (red). pLNTK and TOPO-TA vectors were digested with EcoRI and ligated. A clone was isolated that featured the 5' homology arm in the physiological orientation. LoxP sites flanking the Neo^R cassette are drawn as horizontal arrowheads. Scissors symbol indicates sites of cutting. Thymidine kinase gene (TK), Ampicillin resistance gene, (Amp^R), Kanamycin resistance gene, (Kan^R).

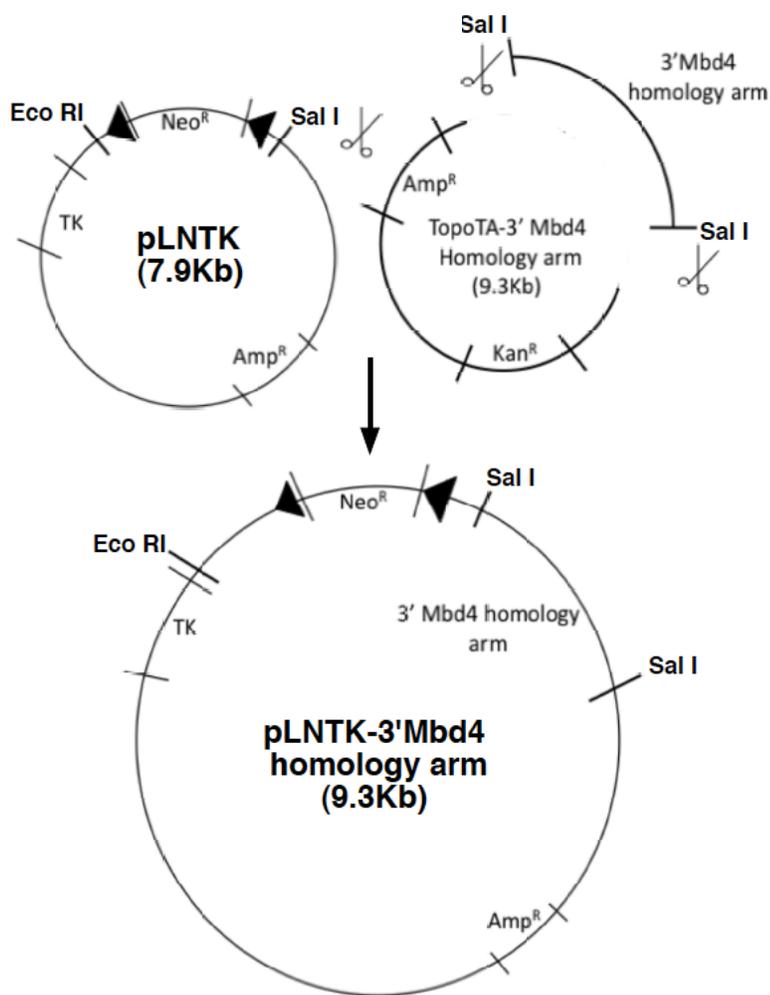


Figure 7. **Construction of pLNTK-3' Mbd4 homology arm.** pLNTK and TOPO-TA vectors were digested with SalI and ligated. A clone was isolated that featured the 3' homology arm in the physiological orientation. LoxP sites flanking the Neo^R cassette are drawn as horizontal arrowheads. Scissors symbol indicates sites of cutting. Thymidine kinase gene (TK), Ampicillin resistance gene, (Amp^R), Kanamycin resistance gene, (Kan^R).

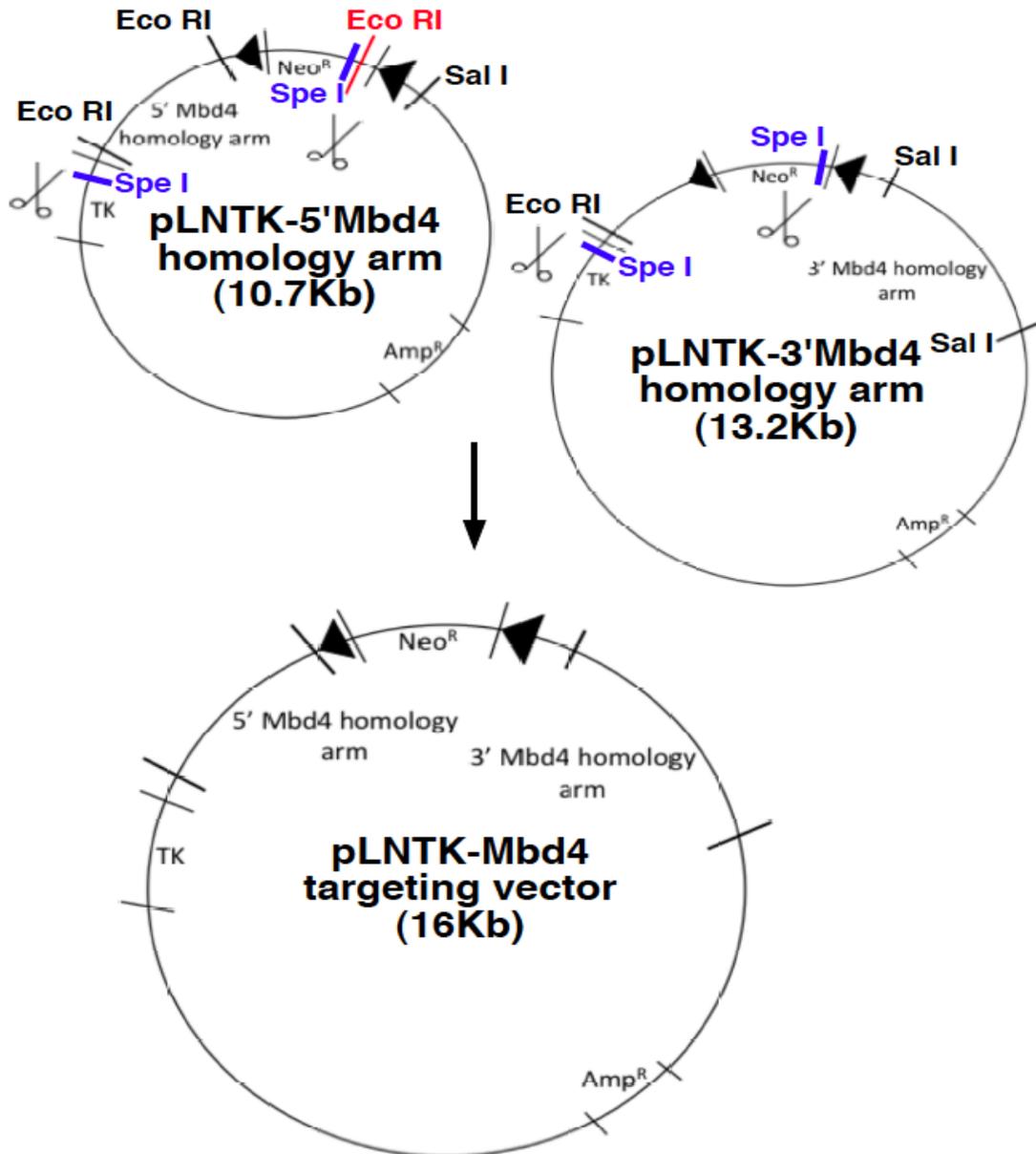


Figure 8. **Construction of Mbd4 targeting vector.** pLNTK-5'Mbd4 homology arm and pLNTK-3' homology arm were digested with SpeI and ligated. A clone was isolated that featured the 5' homology arm in the physiological orientation. LoxP sites flanking the Neo^R cassette are drawn as horizontal arrowheads. Scissors symbol indicates sites of cutting. Thymidine kinase gene (TK), Ampicillin resistance gene, (Amp^R), Kanamycin resistance gene, (Kan^R).

and Klenow fragment (MBI Fermentas), incorporating α -dCTP ^{32}P (15.44mCi/ml Perkin Elmer). For 5' Mbd4 probe, sense, 5'-GTGTGCCTCCTTTTTGGTTAATAC-3', and anti-sense, 5'-TTTTGCATGATGGCAAGGAC-3' primers were used. For 3' Mbd4 probe, sense, 5'-AGGACTGGTTCTAACCTGTTTCTG-3', and anti-sense, 5'-AGGCTACCATCCTCTCATTAGAAG-3' primers were used. The radiolabeled DNA probe was passed over a fine Sephadex G-25 column (Roche 11 273 922 001) to remove free nucleotides, and boiled for 5 minutes in 150 μl of pre-hybridization buffer (1% SDS, 50% formamide, 5X Denharts, 100mg/ml salmon sperm DNA, and 5X SSC). The boiled mix was then added to 50ml of pre-hybridization buffer and incubated with membranes at 42°C overnight in a hybridization rotisserie oven. Membranes were washed three times in 2X SSC, 0.5% SDS at 68°C for 15 minutes. A final wash in 0.5X SSC, 0.5% SDS was carried out at 68°C for 10 minutes, then membranes were briefly dried, wrapped in Syran Wrap, and exposed to a phosphoimager screen (Molecular Dynamics) overnight.

2.4 CH12.F3 cell culture and IgA CSR

CH12.F3 cells were thawed out of liquid nitrogen storage and allowed to recover at 37°C, 7.5% CO₂, for 48 hours in CH12.F3 media (RPMI supplemented with 10% fetal calf serum (FCS), 4mM glutamine, 50 μM β -mercaptoethanol, and pen/strep). Cells were spun down at 405 g using a refrigerated tabletop centrifuge (Eppendorf Model 5804 R), counted, and resuspended at a 1.25 x 10⁵ cells/ml in a final volume of 2ml. Cells were stimulated with IL4 (5ng/ml), TGF- β (0.2ng/ml), and CD-40L (25% vol/vol), referred to as CIT, to switch from IgM to IgA for 24 or 40 hours, whereupon cells were analyzed for surface IgA by flow cytometry. In cases where cell

extracts were needed for Western blot, the cell culture conditions listed above were scaled up by a factor of 4.

2.5 Mice

All mice were maintained in the BRL animal facility at the University of Illinois College of Medicine under specific pathogen-free conditions. $Mbd4^{\Delta_{2-5/+}}$ mice (Cortellino et al., 2003) were a gift from Dr. Bellacosa (Fox Chase Cancer Center, PA) and were backcrossed to the C57BL/6 background nine times in his lab. We bred heterozygous $Mbd4^{\Delta_{2-5/+}}$ mice to obtain $Mbd4^{\Delta_{2-5}/\Delta_{2-5}}$ KO mice. $Ung^{\Delta/\Delta}$ mice on a C57BL/6 background were a gift from Dr. Storb (University of Chicago, IL). All procedures involving mice were approved by the accredited Institutional Animal Care Committee of the University of Illinois College of Medicine.

2.6 Primary B cell culture and IgG1 and IgG3 CSR

Primary quiescent B cells were isolated from whole spleens from 8-12 week old mice. Briefly, splenocytes were treated with lysis buffer (Sigma R7757) to remove contaminating erythrocytes, incubated for 30 minutes on ice with anti-CD43 microbeads (Miltenyl-biotec 130-049-801), and allowed to flow through MACS Separation Columns (LD columns 130-042-901). Cells were washed three times with HBSS (Gibco) and resuspended at 1×10^7 cells/ml in MACS buffer (HBSS, 0.5% FCS, and 2mM EDTA) prior to loading onto columns. Columns are equilibrated with 2ml of MACS buffer prior to loading of cells. We typically load between $0.5-1 \times 10^7$ total splenocytes onto the columns in 0.5-1ml of MACS buffer. Quiescent (CD43-) B cells flow through the column after two successive 1 ml washes of MACS buffer. Recovery after

column elution ranges from $2.5-4 \times 10^7$ quiescent B cells. The remaining CD43+ splenocytes remain stuck to the magnetized column. Cells stimulated to switch to IgG1 are counted and resuspended at 8×10^5 cells/ml in B cell media (RPMI, 20% FCS, 4mM glutamine, 50 μ M β -mercaptoethanol, and Pen-strep) supplemented with 50mM LPS and 10ng/mL IL-4 (LPS+IL4). For IgG3, the stimulation conditions are identical, except that IL4 is not added. Cells, 2×10^6 , or 5×10^6 , are harvested after 48 hours for making RNA, or extracts for Western analysis, respectively. Cells, 1×10^6 , are harvested at 72 and 96 hours and prepared for flow cytometry for detecting surface B220, IgG1, and IgG3.

2.7 Flow cytometry

CH12.F3 (5×10^5 cells) or primary B cells (1×10^6 cells) were pelleted at 405 g, counted, washed once in FACS buffer (PBS, 0.125% sodium azide, 2% FCS), and spun down at 7000 rpm in a tabletop microcentrifuge (Eppendorf Model 5415 D). Cells were resuspended in 50 μ l of FACS buffer, and incubated on ice for 5 minutes with 1 μ l of Fc receptor blocker (0.5 μ g/ml) (BD Pharmingen 553142). An equal volume of primary antibody (Ab) staining solution (Primary Ab in FAC buffer) was added and allowed to incubate for 20 minutes in the dark, followed by three washes with 1ml of FACS buffer. Where needed a secondary antibody staining solution (Secondary Ab in FACS buffer) was added for 20 minutes in the dark and washed as described above. Cells were resuspended in 500 μ l and passed through FACS tubes with cell strainers (BD Falcon 352235). For IgG1 and IgG3 surface staining, 1 μ l of FITC rat anti-mouse IgG1 and IgG3 (0.5mg/ml each) (BD Pharmingen 553443 and 553403) were used. For IgA surface staining, 1 μ l of 16.67 fold diluted BIOT rat anti-mouse IgA (0.5mg/ml) (Southern Biotech) was used in combination with 1 μ l of 4.44 fold diluted Streptavidin-APC (0.2mg/ml) (eBioscience 17-4317-

82). For B220 surface staining, 4 μ l of APC anti-mouse CD45R/B220 (0.2mg/ml) (BioLegend 103212) was used. In cases where B cells were transduced with GFP-encoded virus, IgG1 surface staining was detected using 1 μ l of APC rat anti-mouse IgG1 (0.2mg/ml) (BD Pharmingen 550874). The flow cytometry analyses represent 10,000-15,000 events and were gated for live lymphoid cells determined by forward and side scatter with Cyan ADP and Summit software (Becton Dickenson).

2.8 Quantitative (q) RT-PCR

RNA was prepared using Trizol (Invitrogen) according to the manufacturer's instructions. To prepare the cDNA, 2ml of RNA (4 μ g) was used to program the reverse transcription reaction containing random primer (50mg/ml), DTT (2mM), dNTPs (1mM), reverse transcriptase (RT) (Superscript II) (200U), and a 5X buffer provided by the Superscript II manufacturer (Invitrogen), in a final volume of 10 μ l for 1 hour at 42°. To inactivate the RT, 30 μ l of 1/10 TE buffer was added and incubated at 72° C for 7 minutes according to manufacturer's specifications. The following primers were used in qRT-PCR (ABI 7600 FAST SYSTEM): Mbd4, sense, 5'-ACGATCTCCTGTCAAAAACC-3', and anti-sense, 5'-GCCCTTAGAAACCGAAGCACA-3'; 18S RNA, sense, 5'-TTGACGGAAGGGCACCACCAG-3', and anti-sense, 5'-GCACCACCACCCACGGAATCG-3' (151); μ GLT, sense, 5'-CTCTGGCCCTGCTTATTGTTG-3', and anti-sense, 5'-GAAGACATTTGGGAAGGACTGACT-3'; α GLT, sense, 5'-CCTGGCTGTTCCCCTATGAA-3', and anti-sense, 5'-CGGAAGGGAAGTAATCGTGA-3'; AID, sense, 5'-CCATTTCAAAAATGTCCGCT-3', and anti-sense, 5'-CAGGTGACGCGTAACACC-3' (152). A standard curve was generated using a cDNA

template from CH12.F3 cells stimulated with CIT for 24 hours. CT values were plotted versus the Log of serially diluted template (ng). Experimental cDNA samples were diluted accordingly to fall into the linear range of the standard curve. When analyzing primary B cells by q RT-PCR, a standard curve using cDNA from B cells activated for 48 hours with LPS/IL4 was used instead.

2.9 Proliferation assay

The fluorescent dye, Carboxyfluorescein diacetate succinimidyl ester (CFSE) (Invitrogen C1157) was used to monitor cell proliferation in CH12.F3 cells. Cells were washed in a PBS buffer containing 0.2% BSA, counted, and 1×10^6 cells were resuspended in 1.5ml of the PBS buffer. Equal volumes of cell suspension and staining solution (30 μ M CFDA-SE in PBS buffer) were mixed to achieve a final CFSE concentration of 15 μ M. Cells were incubated in the dark for 30 minutes at 37° C, quenched with 12 ml RPMI media containing 20% FCS, and pelleted at 405 g. Cells were then washed once more with the quenching media, pelleted, and resuspended in 2ml of CH12.F3 media and incubated for 3 hours. This period allows for any unincorporated CFSE dye to diffuse out of the cells. Cells were then prepared for flow cytometry. The incorporated CFSE dye was detected at 3 and 24 hours to assay for dilution of the dye overtime using a CYAN ADP machine with Summit software.

2.10 Western Analysis

Cytoplasmic and nuclear protein extracts were prepared as described (153), with one modification. Briefly, 2×10^6 cells were lysed in 100 μ l (instead of 400 μ l) of buffer A with freshly added 1mM DTT and protease-inhibitors (Sigma P8340). Proteins were resolved by SDS-PAGE. A 6X SDS loading dye solution was added to cytoplasmic or nuclear extracts then

heated to 99°C for 10 minutes prior to loading. Stacking gel was composed of 12% polyacrylamide, Tris-Cl- pH 6.8, and 1% SDS. The running gel was made with 10% polyacrylamide, Tris-Cl-pH 8.8, and 1% SDS. All gels were run for 4 hours at 130 volts in a SDS buffer (Tris-CL -pH7.5, 1% SDS). The gel was then transferred to a nitrocellulose membrane (Millipore) for 3.5 hours and placed into a blocking solution (Tris-Cl- pH 7.5, 150mM NaCl, 0.2% Tween, and 5% milk) overnight. Membranes were incubated with antibodies [anti Mbd4, rabbit anti-mouse polyclonal (1:5000) (Abcam 12187); anti Mlh1, rabbit anti-mouse monoclonal (1:10,000) (Abcam 92312); Msh2, (Santa Cruz sc-494); anti β -actin, mouse anti-mouse monoclonal (1:10,000) (Sigma A8316)] for 2-4 hours at 25° C in probing solution (Tris-Cl- pH 7.5, 150mM NaCl, 0.2% Tween, and 2% milk), followed by three 10-minute washes with TBS-T (Tris-Cl- pH 7.5, 150mM NaCl, 0.2% Tween). Next, membranes were probed for 1 hour with a goat anti-rabbit secondary antibody (1:10,000) conjugated to horseradish peroxidase (HRP), followed by three more 10-minute washes with TBS-T and final 10-minute wash with TBS. For β Actin, a goat anti-mouse secondary Ab was used instead. Membranes were then briefly blotted with paper towels and placed in plastic encasings. HRP activity was visualized via chemiluminescence-optimized films (Denville Scientific E3012) using a chemiluminescent substrate (Pierce 34080). Membranes were stored in TBS overnight for future probing.

2.11 LM-PCR Double strand breaks assay

CH12.F3 cells were lysed in agarose plugs as a source of genomic DNA as previously described (55). For LM-PCR, sense, 5'-GCAGAAAATTTAGATAAAAATGGATACCTCAGTGG-3' (5' S μ primer) (55), and anti-sense, 5'-GCGGTGACCCGGGAGATCTGAATTC-3' (L.1 primer) (55), oligos were used. Hot

Start Taq polymerase (1.4 units) (Qiagen), 1.5mM MgCl₂, 400uM dNTPs (Fermentas), and 5μL of agarose plug DNA was used to program a touchdown PCR: denaturation for 5 minutes at 95°, followed by the touchdown stage (denaturation at 95° C/ 1 minute of annealing at 65° C/1.5 minutes of elongation 72° C), which stepped down in annealing temperature 1° every 2 cycles, until reaching 61°, then 28 cycles of a regular stage (denaturation at 95°/ 1 minute of annealing at 61°/1.5 minutes of elongation 72° C), and lastly a final elongation for 5 minutes at 72° C. PCR products were separated on a 1.5% agarose gel overnight at 40mA, prepared for Southern analysis, and hybridized with the Sμ probe. The Sμ probe was generated as previously described in the Southern analysis section from a 431bp amplicon using sense, 5'-GCAGAAAATTTAGATAAAAATGGATACCTCAGTG-3'(55), and anti-sense, 5'-GGCTTTCTCTGTCTGGGTCA-3', oligos.

2.12 Cloning of individual LM-PCR products

LM-PCR products, from three independent PCR reactions, were treated with excess dATP (1mM) and Taq polymerase (Fermentas) for 30 minutes at 72° in a total volume of 10μL. An aliquot (2μL) of this reaction was cloned into TOPO-TA (Invitrogen), ethanol precipitated, resuspended in 10μL of water, and 6μL was transformed into DH5-a competent cells. Ampicillin resistant (amp^r) bacteria were plated on amp plates at a density of 100-200 colonies per Petri dish (Fisher). A colony lift hybridization procedure was used to identify colonies carrying Sμ-specific plasmids. Briefly, colonies were replica plated onto nylon membranes (Amersham Hybond N⁺ RPN82B). The membranes were then denatured in 0.5M NaOH for 45 seconds, neutralized in 1.0M Tris (pH 7.6) for 45 seconds, and DNA was precipitated on the membranes in 1M Tris-1.5 NaCl (pH7.6) for 45 seconds. Finally, the membranes were vigorously washed in

the precipitation solution to remove protein debris. Membranes were then baked in an oven at 80° for 2 hours and incubated at 42° in the pre-hybridization buffer (same as for Southern) for 1 hour. The S μ probe was prepared, as described in Southern analysis section, and incubated with membranes overnight. To visualize positive colonies, membranes were quickly air-dried, wrapped in Syran Wrap, and overlaid with autoradiography-sensitive films (Denville Scientific E3012) overnight. S μ - positive colonies were then picked from the original plates, minipreps were prepared, and DNA was sequenced using an ABI 3730 XL capillary machine.

2.13 Cloning and identification of S μ -S α junctions

Three independent preparations of spooled genomic DNA, (400ng) pooled from CH12.F3 WT or Mbd4 KO cells that were activated with CIT for 24 hours, was used to PCR amplify S μ -S α junctions using sense, 5'-GGTATCAAAGGACAGTGCTTAG-3', and anti-sense, 5'-GTCTGCCCGGTCTAGGTAAG-3' oligos. The PCR protocol was as follows: denaturation for 2 minutes at 94°, 38 cycles of denaturation for 15 seconds at 94°/annealing for 30 seconds at 60°/elongation for 3 minutes at 68°, plus 10 seconds each successive cycle, using Expand Long template polymerase (3.75 units) (Roche), along with Buffer 2 (2.75mM MgCl₂, Roche), and 400uM dNTPS (Fermentas). PCR products were cloned into the TOPO-TA vector. Colonies containing an S μ -S α junction were identified using the colony lift hybridization procedure and were verified by sequencing.

2.14 Plasmids

2.14.1 pCDNA^{NSI*}-PGK-Cre

To remove a neo^R cassette flanked by loxP sites from correctly targeted Mbd4 KO clones the Cre recombinase sequence was cloned into pCDNA3.1 (+). pCDNA^{NSI*}-PGK-Cre vector was constructed in several steps. pMSCV (Clontech) was linearized by digestion with ClaI (MBI Fermentas) for 1 hour at 37°, heat inactivated for 20 minutes at 70°, treated with the Klenow fragment (MBI Fermentas) and 100mM dNTPS each (MBI Fermentas) for 30minutes at 37° to produce a blunt end, and ethanol precipitated. The DNA was then resuspended in restriction buffer, digested with EcoRI (MBI Fermentas) for one hour at 37°, and heat inactivated for 20 minutes at 80°, releasing a cassette containing a phosphoglycerate kinase (PGK) promoter and green fluorescent protein (GFP) (fig. 9A). pCDNA 3.1(+) (Invitrogen) was linearized by digestion with XbaI (MBI Fermentas), ends were filled in with the Klenow fragment, and finally digested with EcoRI. Digested pMSCV and pCDNA 3.1(+) (10ng each) were ligated (T4 ligase-Invitrogen) together overnight overnight at 14°, ethanol precipitated, resuspended in 10µl of water, and 4µl was used to transform DH5-a competent cells (Invitrogen). Colonies selected on amp plates (100mg/ml) were analyzed and a correct clone was identified, which we called pCDNA-CMV-PGK-GFP (Fig. 9B). In the next step, the CMV promoter was removed from pCDNA-CMV-PGK-GFP by digestion with BglIII (MBI Fermentas) and EcoRI. To remove restriction enzymes, digested DNA was extracted with phenol-chloroform and the ends were filled in with Klenow, recircularized by ligation overnight, followed by selection on amp plates to obtain pCDNA-PGK-GFP (fig. 9B). The Cre cDNA was PCR amplified (Pfu Ultra-Stratagene) from pCMV-intron-Cre/Neo (a gift from Dr. Roberta Franks). To introduce an efficient Kosak sequence we engineered a forward primer with a tail to include a 5' HindIII site and a good mammalian Kosak (ACCATGA), thereby retaining serine as the first amino acid following methionine from P1 bacteriophage Cre. The reverse primer was designed to contain a

3' Apa I restriction site and internal Nsi I and Cla I restriction sites, to be used for future cloning applications. The Cre PCR reaction was incubated at 72° for an additional 30 minutes using Taq and 1mM dATP (MBI Fermentas) and cloned into TOPO-TA. The Cre insert was digested from TOPO-TA using HindIII and ApaI and pCDNA-PGK-GFP was digested using HindIII and ApaI (fig. 10A). Digested plasmids (10ng each) were ligated together in a volume of 10µl for 1 hour at 25°, ethanol precipitated, resuspended in 10µl of water, and 4µl was used to transform DH5-a competent cells (Invitrogen). Bacteria were plated onto amp plates. The next day amp plates were replica plated onto kanamycin (kan) plates (50mg/mL) and grown overnight at 37°. This counter-selection procedure removes TOPO-TA contaminating plasmids from the screening process, which is critical since we do not gel purify inserts from either of the vector backbones. The correct clones were isolated following restriction fragment analysis. Lastly, to facilitate future cloning applications involving Mbd4 cDNAs, an NsiI restriction site downstream of the multiple cloning site (MCS) was destroyed using PCR mutagenesis (fig.10B) (Pfu Ultra, Stratagene) using sense, 5'GAAGTATGCAAAGCATGCACCTCAATTAGTCAGCAACCAGGTGTGG-3', and anti-sense, 5'-CCACACCTGGTTGCTGACTAATTGAGGTGCATGCTTTGCATACTTC-3' oligos. PCR products were digested with DpnI (Invitrogen) for 30 minutes at 37° to destroy the parent template and this mixture was ethanol precipitated. After transformation and selection on amp plates the correct clones were identified via restriction fragment analysis, thereby establishing pCDNA^{NSI*}-PGK-Cre (fig. 10b), which could now be used as a platform for introducing Mbd4 cDNAs into the pCDNA-PGK format.

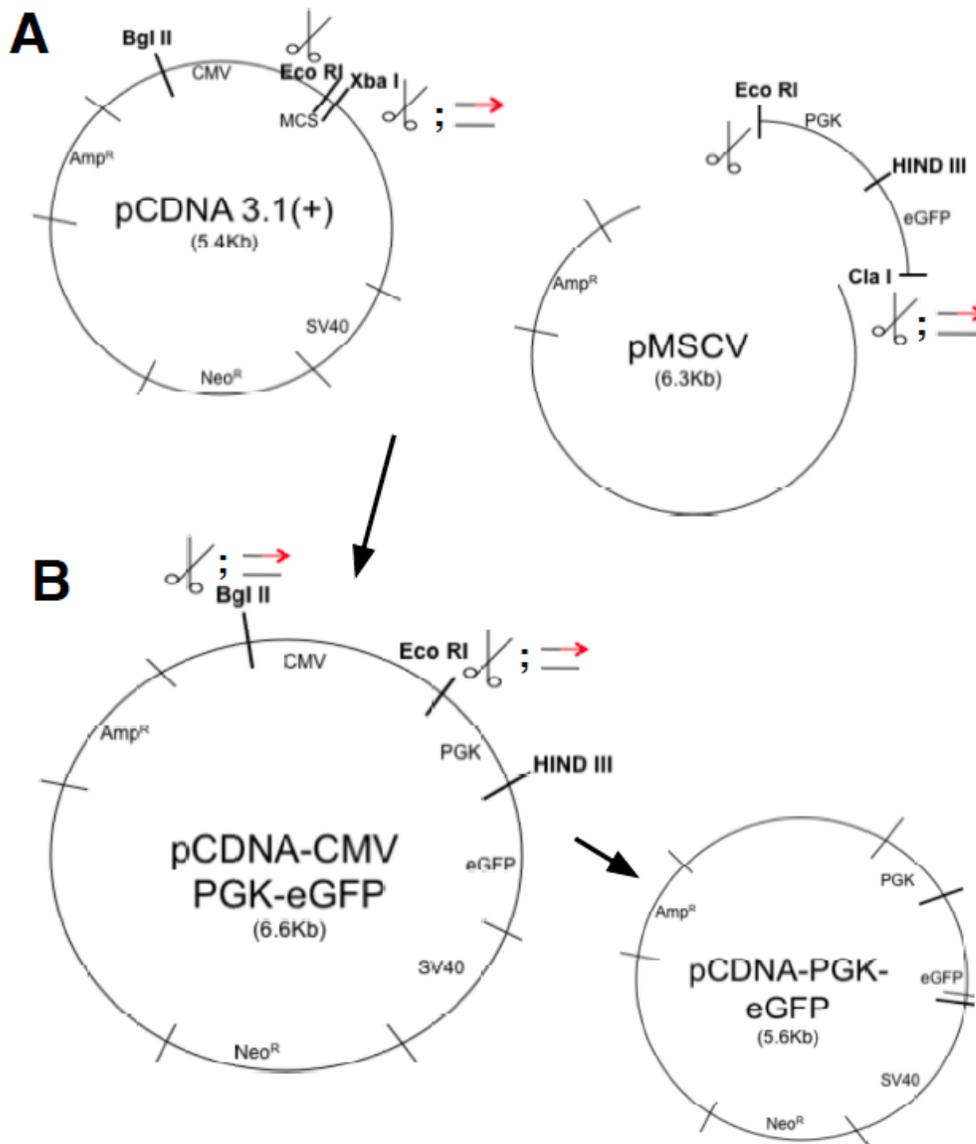


Figure 9. **Construction of pCDNA-PGK-eGFP.** Scissors symbol indicates sites of cutting. Thymidine kinase gene (TK), Ampicillin resistance gene, (Amp^R), Kanamycin resistance gene, (Kan^R). → symbol indicates fill-in with Klenow. **A**) pCDNA 3.1(+) was cut with XbaI and ends were filled in with Klenow, followed by cutting with EcoRI. pMSCV was cut with ClaI and filled in with Klenow, followed by cutting with EcoRI. Cut plasmids were ligated together to create pCDNA-CMV-PGK-eGFP. **B**) CMV promoter was removed from pCDNA-CMV-PGK-eGFP by cutting with BglIII and EcoRI. Ends were filled in with Klenow followed by recircularization with T4 Ligase.

2.14.2 pCDNA^{NSI*}-PGK-Mbd4

For genetic complementation in the Mbd4 KO cell line we cloned the cDNA for Mbd4 into pCDNA. To make pCDNA^{NSI*}-PGK-Mbd4, cDNAs encoding full-length and various iterations of Mbd4 were PCR amplified using oligos with modified tails. (PhusionTaq-Finnzymes). cDNAs was prepared using RNA that was isolated from primary B cells that were stimulated for 48 hours with LPS+IL4. PCR reactions were incubated at 72° for an additional 30 minutes using Taq and 1mM dATP (MBI Fermentas) and cloned into TOPO-TA. The correct TOPO-TA-Mbd4 clones were identified via restriction fragment analysis. Mbd4 inserts were digested from TOPO-TA using HindIII and NsiI, and pCDNA^{NSI*}-PGK-Cre was digested using HindIII and NsiI (fig. 10B). 10ng of each these digested plasmids were used for cohesive end ligation in a volume of 10ml for 1 hour at 25°. Amp^Rkan^S colonies were picked for each of the various Mbd4 constructs from the original amp plates after kan counterselection and positive clones containing pCDNA^{NSI*}-PGK-Mbd4 were identified via restriction fragment analysis. To make an empty vector control plasmid we took pCDNA^{NSI*}-PGK-Mbd4 and digested it using HindIII and ClaI and filled in the ends with Klenow (fig. 10C). DNA was recircularized overnight. The correct clones were identified by restriction fragment analysis. All plasmids identified herein are made into Maxipreps (Qiagen) and are verified by sequencing using a 3730 XL capillary ABI machine.

2.14.3 pMSCV transduction constructs

To re-introduce the Mbd4 cDNA and various constructs (see appendix D) or the cDNA for UNG and its derivatives via retroviral transduction, pMSCV-puro or pMCSV-gfp and TOPO-TA-Mbd4 or TOPO-TA-Ung plasmids, respectively, were digested with Eco RI and religated to

make pCDNA-Mbd4-puro or pCDNA-Ung-gfp. Steps in obtaining these plasmids are similar to section 2.14.

2.15 Transfection

WT and Mbd4 KO CH12.F3 cells were counted and viable cell density was verified at $1 - 2.5 \times 10^5$ cells/mL using trypan blue exclusion (Gibco). For each transfection 2×10^6 cells were spun down and resuspended in 100ml Nucleofactor^R buffer L (Lonza). Cells are gently pipetted up and down and then added to 2-4 μ g of various pCDNA^{NSI*}-PGK-Mbd4 constructs that had been linearized with PvuI (New England Biolabs), treated with phenol chloroform to remove restriction enzyme, ethanol precipitated, and resuspended in 12 μ l of water. The cell and DNA mixture was then added to cuvettes provided by the Nucleofactor Kit L and electroporated using the D-023 program on the Nucleofactor (Lonza VCA-1005). Dead cells were allowed to float to the top of the cuvette; the remaining viable cells were removed using a sterile pipette and gently added drop-wise to 2.5ml of pre-warmed CH12.F3 media without antibiotics and reduced glutamine (400nM). Transfected cells were allowed to recover for 24 hours, three individual transfections were pooled together, and cells were selected with G418 (750 μ g/ml) for six days, refreshing media every two days. Cells were then counted and viable cell density was verified at $1 - 2.5 \times 10^5$ cells/ml using trypan blue exclusion (Gibco). 1.25×10^6 cells were then seeded into 8ml of CH12.F3 media with or without the CIT cytokine cocktail for 24 hours and then checked for surface IgA by flow cytometry. Stimulated cultures were performed in duplicate. Mbd4 expression was verified by Western blot.

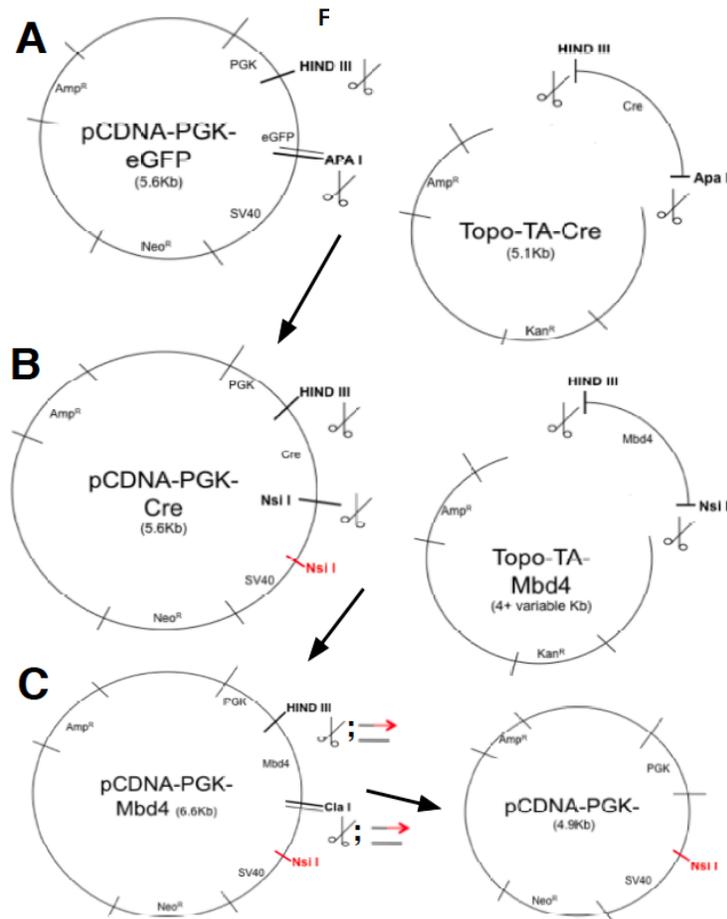


Figure 10. **Construction of pCDNA-PGK-Cre, pCDNA-PGK-Mbd4, and pCDNA-PGK.** Scissors symbol indicates sites of cutting. Thymidine kinase gene (TK), Ampicillin resistance gene, (Amp^R), Kanamycin resistance gene, (Kan^R). \Rightarrow symbol indicates fill-in with Klenow. **A)** pCDNA-PGK-eGFP and TOPO-TA-Cre plasmids were cut with HindIII and ApaI, and ligated to make pCDNA-PGK-Cre. **B)** To facilitate cloning an NsiI site (red) was destroyed in pCDNA-PGK-Cre plasmid, then this plasmid and TOPO-TA-Mbd4 were cut with NsiI and HindIII and ligated together to make pCDNA-PGK-Mbd4. **C)** pCDNA-PGK-Mbd4 was cut with HindIII and ClaI, ends were filled-in with Klenow, and cut plasmid was recircularized with T4 DNA ligase to make the empty vector, pCDNA-PGK.

2.16 Genetic Complementation and transduction

Mbd4 cDNA (exons 1-8), catalytic mutants (D534A,Y514F) and truncated constructs, (see appendix C) were cloned into pMSCV-puro (Clontech). Retrovirus was made by transfecting Mbd4-pMSCV plasmids (4 μ g) and a plasmid containing the glycoprotein from vesicular stomatitis virus (VSV-gp) (1 μ g) plasmid into the Platinum GP cell line (Cell Biolabs). Viral supernatants were harvested 48 hours post transfection, filtered, and used to infect Mbd4 KO CH12.F3 cells. Cells, 2.5×10^5 , in 100 μ l of CH12.F3 media were mixed with 400 μ l of viral supernatant and spinoculated at 1000 g for 2 hours in 24 –well plates (Falcon). Cells were then selected in puromycin (2ng/ml) (Sigma) for 2-5 days, depending on the construct, and subjected to the IgA CSR assay. A similar protocol was used to transduce activated primary WT and Ung deficient B cells. cDNAs encoding full-length Ung1 and Ung2 and their mutated active site counterparts (D147N), were cloned into pMSCV-eGFP, and transfected into the Platinum E cell line to make retrovirus. These spinoculations used 1×10^6 cells. In this experiment selection was not required, the B cells were subjected to IgG1 and IgG3 CSR flow cytometry assays, and infected cells were monitored by GFP expression.

CHAPTER 3

RESULTS:

Characterization of the Mbd4 Δ E2-5 mouse

3.1 CSR is not perturbed in the Δ E2-5 Mbd4 KO mouse

The question whether other DNA glycosylases participate during CSR has gained traction over the years as it has become apparent that DSBs and CSR still occur, albeit at a reduced frequency, in the absence of UNG (32, 55, 93). In SHM there seems to be even less a requirement for UNG in generating mutations at A:T basepairs, which are thought to proceed via recognition of U/G and error prone rectification via an Msh2 -dependent mechanism that may require additional nicking activity not provided by either UNG or APE1/APE2 (see section 1.4.10). There is little room for argument that UNG contributes to CSR via its uracil excision function, especially within areas of S DNA called tandem repeats that are peppered with the AID hot spot, where both strands of the DNA duplex are targeted by AID and UNG and the close proximity of SSBs can likely result in a DSB (see chapter 1.4.10) (70). However, in cases where SSBs are far away from each other, it is now generally accepted that MMR plays a crucial role in events that lead to DSB formation (108, 154, 155) (see chapter 1.4.10). In a model of CSR where not all SSBs are dependent on UNG and the BER associated APE1/2, loading of Exo1 still has to occur via a nearby SSB. As MMR does not possess an undisputable intrinsic endonucleolytic activity, an attractive hypothesis is that other DNA glycosylases or endonucleases supply the required chemistry. As stated previously a similar mechanism is probably necessary to generate A:T mutations during SHM, which cannot be dependent on the putative endonucleolytic function of PMS2, because *Pms2* null B cells do not have a perturbed A:T mutation spectra (reviewed in 1.4.9).

One DNA glycosylase that came to the forefront of our attention is Mbd4. Not only does it interact with Mlh1 (136), but it has also been linked to active DNA methylation in a process that involves the coupling of its mismatch glycosylase activity with AID deamination of ⁵meC

(133). A putative lyase function of Mbd4 is also intriguing with regard to additional activity required during SHM (and potentially CSR) not provided by UNG and APE1/2. Therefore, we sought out to investigate whether Mbd4 could participate in CSR alongside MMR based on these observations.

We scanned the literature for the availability of Mbd4 mutant mice and found two with different genetic manipulations (145, 150) that lead to Mbd4 ablation. We have obtained one of these mice (150) and have tested it for CSR deficiency. However, we find that like the other previously examined Mbd4^{Δ3/Δ3} mouse (146), the Mbd4^{Δ2-5/Δ2-5} mouse does not display any clear CSR phenotype.

When analyzing CSR to the IgG1 and IgG3 there is no obvious difference between Mbd4^{+/+} (WT) and Mbd4^{Δ2-5/Δ2-5} (KO) littermates. This is summarized in figure 11A in an experiment that used two WT control animals and two KO animals. There is no statistical difference between WT and KO mice for Ig1 CSR at either day 3, or day 4, p=0.4, and p=0.7, respectively. Although CSR to IgG3 was suboptimal in this particular experiment, we observe the same finding for IgG3 CSR, with no statistical difference between WT and KO mice at either day 3 or day 4, p=0.2 and p=0.9, respectively.

3.2 MMR proteins are substantially reduced the ΔE2-5 Mbd4 KO mouse

MMR contributes to both SHM and CSR and deficiencies in this genetic pathway lead to significant changes in the outcomes of both processes. A more in depth review of MMR function in antibody diversification can be found in the chapter 1.4.6, however it is useful to reiterate the importance of Msh2 and Mlh1 in CSR. The effect of loss of either Msh2 or Mlh1 on *in vitro*

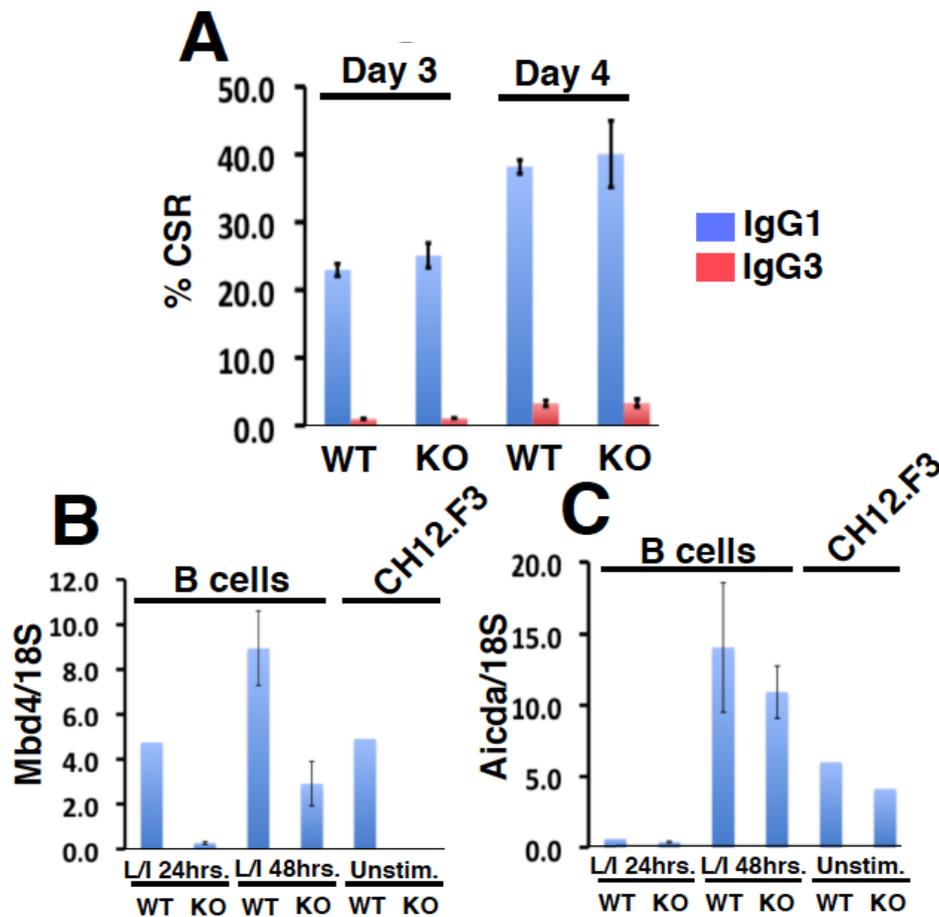


Figure 11. **E Δ 2-5 deletion of Mbd4 gene does not influence IgG1 or IgG3 CSR. Mbd4 $+/+$ (WT) and Mbd4 Δ 2-5/ Δ 2-5 (KO).** To determine if CSR was affected by loss of Mbd4 resting primary B cells were purified from Mbd4 KO mice and monitored for surface expression of switched Ig and compared to WT littermates. Also shown are Mbd4 and Aicda transcripts at 2 days post activation with LPS and IL4. **A)** Flow cytometry was used to check for surface IgG1 or IgG3 on primary B cells that were isolated and stimulated with LPS with or without IL4 to induce IgG3 or IgG1 CSR, respectively, for 3 or 4 days. Histogram plots the average with SEMs of two independent mice for each condition. No statistical significance was reported using two-tailed Student's t-test. **B)** qRT-PCR used Mbd4 exons 6-7 primers to amplify cDNA from RNA isolated from B cells stimulated LPS+IL4 for 24 or 48 hours. cDNA prepared from unstimulated CH12.F3 WT or KO RNA served as positive and negative controls for the Mbd4 qRT-PCR signal. Histograms plot the average with SEMs from one experiment (N=3), except for 24 hour WT B cell and CH12.F3 controls (N=1). **C)** Same as fig. 1B, except that Aicda primers were used instead.

CSR ranges between around a 40-80% reduction in switched Ig, depending on the isotype that is being interrogated, with IgG3, IgG1, and IgA being less affected than IgG2a and IgG2b (92, 96, 98, 104). Additionally, even 50% reductions of MMR proteins, which occurs in mice that are heterozygous for Msh2 and Mlh1, causes a measurable deficit in CSR, ranging from 20-50% (Personal communication, Janet Stavnezer). Curiously, combined deficiency of Msh2 and Mlh1 does not result in a further reduction in CSR, hinting at an epistatic relationship between these genes (95). Mbd4 was originally identified as an Mlh1 binding partner (136), and unexpectedly, Mbd4 ablation led to a significant alteration of the steady state level of Mlh1 and other MMR proteins (150). In this report Mlh1 and Pms2 proteins were attenuated ~80% in mouse embryonic fibroblasts (MEF), while Msh2 and Msh6 were less affected (~2-3 fold), and the reduction of all proteins was controlled by a post-transcriptional event. We have extended the characterization of MMR proteins in MEF to primary B cells from the same mouse Mbd4 KO (Mbd4^{Δ2-5/Δ2-5}).

Immunoblot analysis of nuclear extracts from B cells activated with LPS and IL4 shows a strong reduction of MLH1, around 90%, and a similar decrease in MSH2 (fig. 12A) in Mbd4^{Δ2-5/Δ2-5} cells compared to WT littermate controls. We also confirmed results from Cortellino et al. (149), that the reduction of MMR proteins is not preceded by a reduction of Msh2 or Mlh1 transcripts in Mbd4^{Δ2-5/Δ2-5} cells suggesting that the Mbd4 protein is involved in stabilizing MMR proteins (fig. 12B). The lack of reduction of CSR in B cells from the Mbd4^{Δ2-5/Δ2-5} mouse is a surprising result considering that normally just a 50% reduction of MMR proteins causes measurable deficits in CSR. We hypothesize that in order to achieve WT levels of CSR despite the markedly reduced levels of MMR proteins found in cells from an Mbd4^{Δ2-5/Δ2-5} background that Mbd4 must possess a modulating function during CSR. Once this function is not present, as

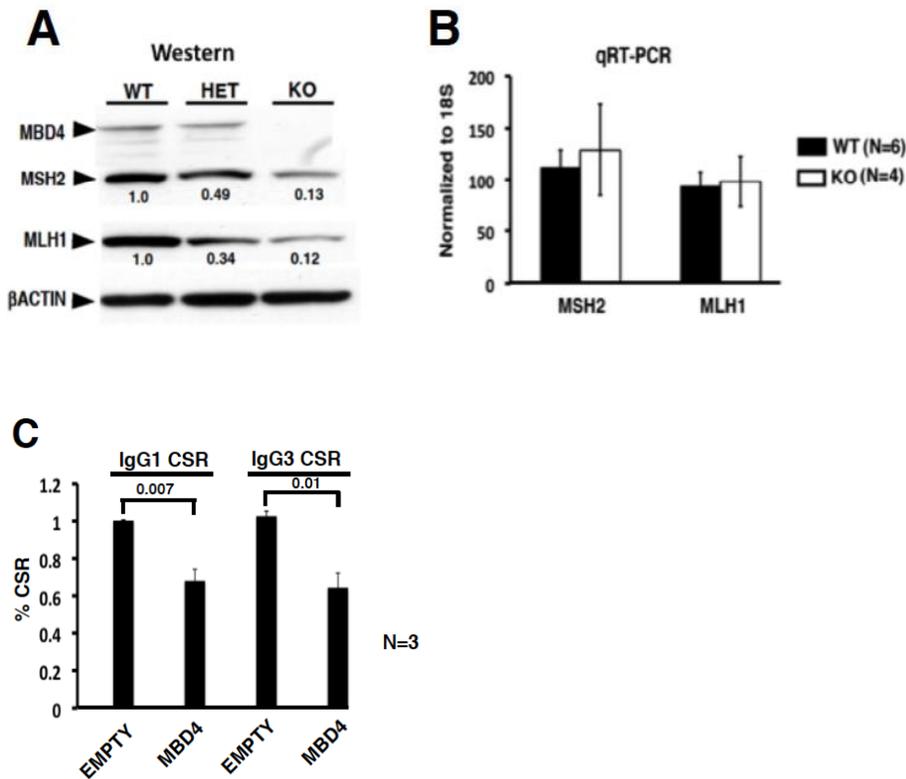


Figure 12. Loss of Mbd4 compromises stability of MMR proteins and this occurs post-transcriptionally. MMR proteins MSH2 and MLH1 are probed in the $\Delta E2-5$ mouse in Western analysis and transcripts for Msh2 and Mlh1 from the same mouse are scored by qRT-PCR. Additionally, an experiment summarizing the effect of Mbd4 overexpression in WT primary B cells is shown. Where appropriate Student's two tailed *t* test was used to determine statistical significance. **A**) Immunoblot analysis of nuclear extracts prepared from wild type (WT), heterozygous (Het), and null (KO) Mbd4 primary B cells stimulated LPS (50 μ g/ml) and IL4 (10ng/ml) for 48 hours. Full length MBD4 (69kD) was probed with a rabbit anti-MBD4 antibody (Ab) followed by a goat anti-rabbit HRP-conjugated Ab. MSH2 (107 kD) was probed with a rabbit anti-MSH2 antibody and a goat anti-rabbit HRP-conjugated antibody. MLH1 (95kD) was probed with a rabbit anti-MLH1 antibody and a goat anti-rabbit HRP-conjugated antibody. β -Actin (43kD) was used as a loading control and probed with mouse anti- β ACTIN antibody and a goat anti-mouse HRP-conjugated antibody. A chemiluminescent substrate was added and visualized with chemiluminescent films. **B**) Msh2 and Mlh1 transcripts were analyzed by qRT-PCR using cDNA prepared from RNA isolated from WT (solid bar) and null (KO) Mbd4 (open bar) primary B cells stimulated for 48 hours as in **A**. All transcript signals were normalized to the control gene, 18S. Histogram depicts the average with SEMs (N=6 for WT; N=4 for KO). **C**) Empty or Mbd4 retroviral vectors were used to transduce primary WT B cells stimulated for 24 hours as in **A**, or with LPS alone. 48 hours post infection cells were harvested and subjected to IgG1 or IgG3 CSR flow cytometry assays. Histogram depicts the average of triplicate transductions with SEMs (N=3) normalized to empty vector and set to 1.

in an $Mbd4^{\Delta 2-5/\Delta 2-5}$ background, WT levels of CSR can still occur, even though loss of Mbd4 causes substantial reduction of MMR proteins. It follows that overexpression of MBD4 in a WT background will cause a reduction of CSR. I will now show some preliminary data that supports this hypothesis.

3.3 Mbd4 modulates CSR in primary B cells

To investigate the effect that overexpression of Mbd4 would have on CSR we transduced WT primary B cells that had been stimulated to undergo CSR to either IgG1 or IgG3 with a retrovirus encoding the cDNA for full-length Mbd4. A total of three mice were used in two separate experiments. Compared to the empty vector control, overexpression of MBD4 causes a modest (20-50%) but statistically significant inhibition of both IgG1 and IgG3 CSR (fig. 12C). This result is congruent with the hypothesis that Mbd4 modulates CSR physiologically and explains why B cells from the $Mbd4^{\Delta 2-5/\Delta 2-5}$ background, which are expected to show a substantial reduction of CSR based on diminished MMR proteins, in fact do not manifest a reduced CSR phenotype.

There is a parallel between overexpression of MBD4 and UNG, namely that both cause inhibition of CSR. The next section describes experiments using enforced UNG expression in primary B cells. Although the following section may seem like a departure from the focus on Mbd4, I have included it here because by virtue of its U/G mismatch excision properties Mbd4 could act very similarly to Ung during CSR. Moreover, I infer that the balance that exists in B cells between faithful and error-prone repair, which is normally skewed towards more error-prone repair, can be tipped in favor of faithful repair. This is evidenced by overexpression of both proteins, which causes an inhibition of CSR.

3.4 Ectopic expression of UNG2, but not UNG1, inhibits CSR to IgG1

Class switch recombination (CSR) proceeds via formation of double stranded breaks (DSBs) initiated by AID. The lesion left in the wake of AID deamination of cytosine creates a uracil within the DNA that must be removed. Failure to achieve removal and correction of uracil could lead to deleterious mutations that would wreak havoc on the genome. Uracil DNA glycosylase (Ung) deficiency leads to pronounced defects in CSR in both human and mouse (32, 81, 93). In the Deamination model of CSR, UNG initiates formation of DSBs by introducing an abasic site, which can then be recognized by APE1/2 endonucleases and processed into SSBs. If two SSBs are found nearby enough to one other on opposite strands of DNA, a DSB will be generated (see chapter 1.4.10) (108).

This model is complicated by later experiments which demonstrated that retrovirally introduced catalytically inactive forms of Ung (D147N, N206V, and H270L) that retain less than 0.6% of WT UNG activity, but are fully competent for DNA binding, were able to rescue the CSR deficiency associated with loss of Ung in murine primary B cells (34). The authors of this report cited a likely scaffolding role for UNG in recruiting various proteins associated with DNA repair and CSR. However, it was proposed shortly after, and later demonstrated unequivocally, that the residual glycosylase activity left in these active site mutants was indeed sufficient to rescue CSR (83). The reduced catalytic activity was counterbalanced by the high expression of achieved by retroviral reconstitution, which is up to 100 times that of endogenous levels (83). Still somewhat perplexing, however, was the finding in this report that catalytically deficient constructs of UNG were better than WT UNG at rescuing CSR in *Ung* deficient B cells.

Interestingly, in a doubly deficient, *Msh2/ Ung* background the opposite was observed: now WT UNG was better than catalytically attenuated UNG at rescuing CSR. This raised the

intriguing possibility that BER and MMR are actually competing for access to the U/G mismatch that is created by AID. Importantly, it suggested that during CSR (and SHM) UNG acts independently of the MMR pathway (such as Msh2, and Mlh1), which is different than what was previously proposed (93, 156). Perhaps Ung is even separated spatially and temporally from MMR during CSR, as has been previously suggested (107), although some degree of overlap between the two pathways during CSR cannot be ruled out.

Another insight came from using *E.coli* Ung in reconstitution studies of CSR in *Ung* deficient B cells, which concerns the proper timing of UNG activity during CSR. Owing to the presence of distinct, alternating promoters, the Ung locus in both human and mouse gives rise to two different forms of Ung: Ung1, which is the mitochondrial form, and Ung2, which is the nuclear form (80). It is assumed that Ung2 is physiologically relevant for CSR (32, 34, 81, 93). While this is likely to be the case, it was stunning to find out that *E. coli* ung is perfectly capable of rescuing the CSR defect associated with *Ung* deficiency in primary B cells (35). *E.Coli* does not have either of the complex mitochondrial or nuclear import signals found in Ung1 and Ung2, respectively. However, the eD54N mutation (the equivalent of mD147N) abolished the rescue, which is not the case for catalytically attenuated forms of UNG2. This suggested to us that when using the retroviral reconstitution system there is an excessive amount of protein produced so enough eUng makes it into the nucleus to rescue the CSR defect. It also suggested that very little UNG protein is required in the nucleus to promote CSR. This is also supported by the finding that CSR is not perturbed in mice that are heterozygous with respect to *Ung* (our unpublished data). Based on the limited ability of *E.coli* UNG to enter the nucleus also suggested that the activity of UNG that is required for promoting CSR might be restricted to a particular phase of the cell cycle.

We have extended these findings and now show that like *e.coli* Ung WT UNG1 (the mitochondrial isoform) is perfectly competent at rescuing the CSR defect in Ung deficient primary B cells, but the catalytically attenuated form cannot (D147N) (fig. 13A, 13B). Interestingly, UNG1 is 2-5 fold better at rescuing the CSR defect in *Ung* deficient primary B cells than UNG2 (nuclear isoform), and this difference is statistically significant ($p=0.011$) (fig. 13B). We have also reproduced the finding that a catalytically attenuated form of UNG2 (D147N) is approximately 2-fold better than WT UNG2 at rescuing the CSR defect in primary B cells deficient in *Ung* (fig. 13B), which is also statistically significant ($p=0.019$). Additionally, we are the first to demonstrate that overexpressing UNG2 in an *Ung* sufficient background actually decreases CSR to IgG1 by 25-50%, $p=0.016$. This lends support to the notion that excessive BER activity favors a straightforward repair of U/G mismatch versus mutagenic repair. Further support for this idea is evidenced by the fact that primary B cells deficient in polymerase β (an important component of BER) show increased CSR to IgG2a, IgG2b, and IgG3 isotypes (86). The highly processive nature of AID ensures that many U/G mismatches are created. Not all U/G mismatches can be processed by MMR. Similarly, not all of the U/G mismatches left over by MMR can be rectified by BER. However, if the system is flooded with excessive DNA proteins (whether it's BER or MMR) the balance that was once tipped in favor of mutagenic repair probably shifts towards a straightforward repair mechanism. This would explain why both MMR and BER are required for CSR, and yet are probably not entirely overlapping pathways, since in the absence of MMR UNG-dependent DSBs still occur and can elicit CSR. The type of DSB which completely relies on MMR and BER (108, 154, 155), probably arises from an interruption of MMR by BER as cells begin to enter S phase as proposed in (106).

Lastly, the inability of catalytically mutated UNG1 (D147N) to rescue CSR in *ung* deficient B cells mirrors what was found for the mutant of *E.Coli* Ung, suggesting that the catalytic attenuation of UNG proteins following the D147N mutation can only rescue CSR if enough of it enters the nucleus (as is the case with the catalytic mutant of UNG2). Inhibiting CSR in the WT primary B cells requires a substantially elevated level of enforced UNG2 expression. D147N UNG2 cannot achieve this feat, suggesting that the inhibition of CSR by overexpression of UNG2 is related to the rectification of U/G mismatches by a uracil excision mechanism that favors faithful repair. Unfortunately, the lack of available catalytic mutants of MBD4 (see section 4.7) has prevented me from addressing whether in a situation analogous to the one documented with UNG2 overexpression the inhibition of CSR by MBD4 overexpression is catalytically dependent. Therefore we cannot be sure whether Mbd4 directly impacts CSR by edification of U/G mismatches via uracil excision. The inhibition of CSR that is caused by overexpression of Mbd4 could occur by another mechanism that is currently not understood.

3.5 Exons 6-8 of Mbd4 are expressed in a residual transcript in $\Delta E2-5$ Mbd4 KO mouse

CSR is not affected in B cells from the Mbd4 ^{$\Delta 2-5/\Delta 2-5$} mouse (see section 3.1) even though MMR proteins are reduced to a level that should have produced a deficit in CSR (see section 3.2). Our investigation of CSR in the Mbd4 ^{$\Delta 2-5/\Delta 2-5$} mouse precisely matches what was documented for CSR in another Mbd4 KO mouse, the Mbd4 ^{$\Delta 3/\Delta 3$} mouse. The targeted deletion in Mbd4 ^{$\Delta 2-5/\Delta 2-5$} mouse is more comprehensive, including all of exons 2-5. However, similar to the Mbd4 ^{$\Delta 3/\Delta 3$} mouse the promoter, exon 1, and exons 6-8 are still intact in this mouse. We decided to check whether any residual transcripts were still made in the Mbd4 ^{$\Delta 2-5/\Delta 2-5$} mouse.

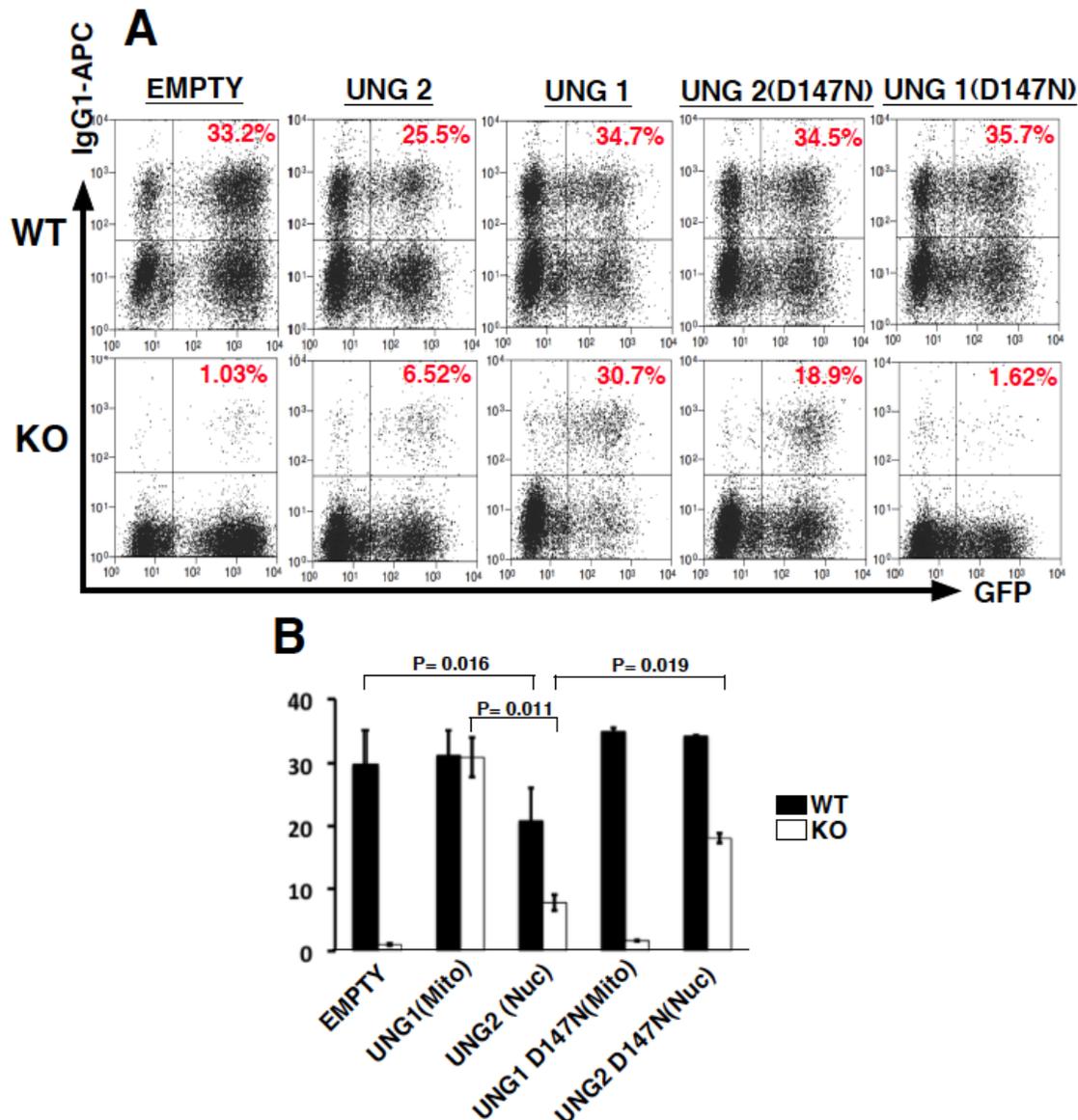


Figure 13. **Nuclear UNG (UNG2) inhibits IgG1 CSR.** WT and null (KO) Ung primary B cells were activated and transduced as in fig. 12C with various constructs of UNG: nuclear UNG (UNG2), mitochondrial UNG (UNG1), and catalytic mutants (D147N). Cells were harvested two days after infection and subjected to the IgG1 flow cytometry assay. **A)** Representative flow cytometry histograms from one experiment are shown. WT (top panel) and KO (bottom panel). GFP (+) cells are plotted on the x-axis and the IgA (APC) (+) cells on the y-axis. The percent of GFP (+) cells that stained with IgA-APC is shown in the upper right corner of each histogram. **B)** Histogram of fig. 13A plots the average of several experiments with SEMs (N=3, except for catalytic mutants, where N=2). Student's *t* test was used to determine statistical significance. (Solid bar=WT; open bar=KO)

Using qRT-PCR we find that there is a transcript (see Appendix A, fig. 1I, 1J). Amplifying cDNA prepared from activated B cells, using primers that span the junction at exons 7-8, one can note that a transcript is expressed in the Mbd4 KO at roughly 30% of the WT transcript after 48 hours of cytokine stimulation (fig. 11B). It is also on at a time when AID expression is elevated (fig. 11C). Another similar residual transcript is found in the Mbd4^{Δ3/Δ3} mouse (145). This raises the possibility that a truncated version of Mbd4 that includes a large portion of the catalytic core of the glycosylase domain is still present in both of the Mbd4 KO mouse systems that have been studied to date (145, 149). There is an internal Kosak sequence found in exon 6 and we have previously shown that expressing exons 6-8 from this kosak yielded a stable peptide. Therefore we are not ready to dismiss a role for Mbd4 in CSR and in the next chapter I will detail attempts to address a role for Mbd4 in CSR.

CHAPTER 4

Results: Characterization

of the Mbd4 Δ E6-8 KO CH12.F3 cell line

4.1 Targeting the catalytic exons of Mbd4 for deletion

To address the role of Mbd4 in CSR, we designed a targeting strategy that would remove catalytic exons 6-8 from the Mbd4 locus. We used a gene-targeted homologous recombination approach to replace exons 6-8 with a neomycin resistance selectable marker gene (neo^f) (fig. 14A). The negative selection thymidine kinase (TK) cassette confers sensitivity to gancyclovir, thereby eliminating clones that randomly integrate the construct. We chose as a model system the B cell line, CH12-F3, which can be induced to undergo robust CSR to IgA (157, 158). To build our targeting construct, 5' and 3' Mbd4 homology arms, 2.85KB and 5.34KB, respectively, were cloned into the pLNTK vector. The neo^f gene is flanked by 34 bp *loxP* sequences and is subsequently deleted by exposure to Cre recombinase to facilitate targeting of the second Mbd4 endogenous allele. The size and integrity of the gene-targeting event was confirmed by Southern blot analyses with two independent probes (fig. 14B). For the endogenous allele, 7.35Kb and 7.7Kb BamH1 restriction fragments are found when analyzed with the 5' and 3' probes, respectively. By contrast, 6.0Kb and 6.9Kb fragments appear when examining the targeted allele using the same probes. Cre-mediated deletion results in removal of the neo^f gene, along with a BamH1 site, leaving a single *loxP* at the site. Following deletion by Cre, both 6.0Kb and 6.9 fragments become a 11kb fragment that can be analyzed with either probe.

Combining three independent Southern blot screens we have identified 3 clones out of 173 (M19 $N\Delta/+$, 1A-12 $N\Delta/+$, and 4-18 $N\Delta/+$) that are correctly targeted at the Mbd4 locus. However, in 4-18 $N\Delta/+$ the targeting event evidently occurred in a clone that had already switched to IgA, hence this clone was eliminated from further analysis. We chose to use 1A-12 $N\Delta/+$ for targeting the second allele of Mbd4. To create 1A-12 $\Delta/+$, the Neo^R cassette was removed by introducing pCDNA^{NSI*}-PGK-Cre via electroporation (Amaxa Nucleofactor). After

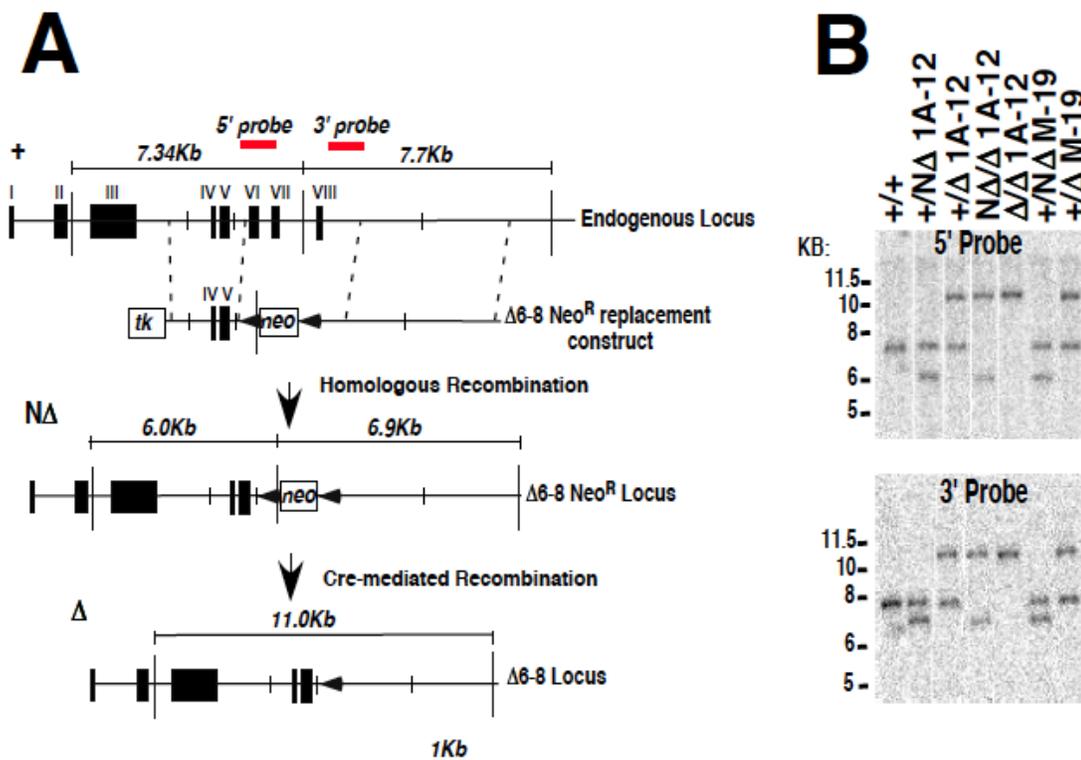


Figure 14. **Deleting exons 6-8 of Mbd4 in the CH12.F3 cell line.** Homologous recombination was used to target exons 6-8 of the Mbd4 gene for deletion. **A**) Map of Mbd4 locus depicting exons 1-8. The endogenous Mbd4 Locus (+) is shown in the top panel with the targeting vector. Middle panel shows the targeted locus (**N** Δ) with replacement of exons 6-8 with the neo^R cassette, after successful homologous recombination. Following Cre expression, the neo^R cassette is floxed from the targeted locus, leaving behind one LoxP site (**Δ**). Bam HI sites are shown as long vertical bars and expected sizes of BamHI fragments are indicated. Probe positions are shown as red horizontal bars. LoxP sites are drawn as horizontal arrowheads. **B**) Southern blot of Bam HI- digested genomic DNA from correctly targeted clones using an Mbd4 5' probe (Top panel) or an Mbd4 3' probe (Bottom panel). Both Mbd4 alleles were targeted in clone 1A-12. In clone M-19 only one allele was targeted. (**N**) denotes presence of neo^R gene. Molecular marker sizes listed in kilobases (**KB**).

24 hours cells were plating via limiting dilution into 96-well plates and individual colonies arising from single cells were picked 7 days later and placed into media with or without G418. Neo^S colonies were identified by their sensitivity to G418 and confirmed by Southern blot using the aforementioned probes. The entire targeting process was reiterated to generate an Mbd4 -/- clone. 1 clone out of 112 was positively identified via Southern blot. In order to rule out any influence of the neo resistance marker in future experiments, 1A-12 NΔ/Δ was subjected to Cre to create 1A-12 Δ/Δ. This is the Mbd4 KO clone used for all further experiments.

4.2 1A-12 Δ/Δ is an Mbd4 KO clone and has a diminished CSR phenotype

The murine cell lymphoma, CH12.F3 was used to identify AID as the paramount B cell specific immunoglobulin diversification enzyme, and it can be induced to switch exclusively from IgM to IgA (29). CH12.F3 undergoes robust CSR and can be relatively easily genetically manipulated (116, 157), thus making it an ideal system for creating a knock out clone of Mbd4. In our hands, administering the CIT cocktail, composed of Il-4 (5ng/ml), TGF-β (0.2ng/ml), and CD-40L (25% v/v) for 24 hours induces CH12.F3 to switch its surface IgM to roughly 18-25% IgA, as assessed by flow cytometry using an anti-IgA-APC conjugated antibody. When comparing to WT, there is a statistically significant, 3 fold reduction ($p < 0.01$) of IgA CSR in the Mbd4 KO (fig. 15A). Two independent Heterozygous Mbd4 clones, M19 and 1A-12, both show a 2 fold reduction compared to WT ($P < 0.01$) (fig. 15A). The ability to score a similar IgA CSR deficit in two completely unrelated clones strengthens the hypothesis that the targeted deletion of Mbd4 translates into a reduction of IgA CSR. Furthermore, comparing KO to heterozygous clones shows about a 30% decrease in IgA CSR, ($P < 0.05$) (fig. 15A), again highlighting a link between loss of Mbd4 and decreased CSR. We discovered sometime later that

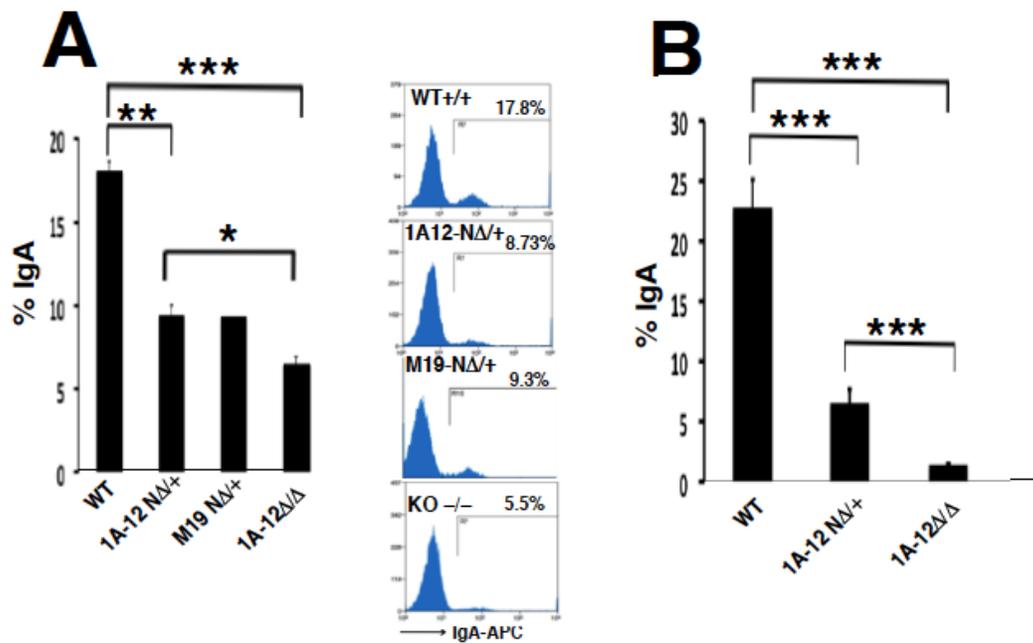


Figure 15. IgA CSR is diminished as a result of Mbd4 ablation in CH12.F3 cell line. To determine whether loss of Mbd4 affects IgA CSR in the CH12.F3 cell line various targeted clones of Mbd4 were subjected to CSR assays and surface expression of IgA was monitored by flow cytometry. Where appropriate Student's two tailed *t* test was used to determine statistical significance (**p* < 0.05, ***p* < 0.01, ****p* < 0.001). **A)** Flow cytometry used to detect surface IgA 24 hours after stimulation with CIT (CD40L (25%v/v), IL4 (5ng/ml), and TGF- β (0.2ng/ml)) in wild type (WT), heterozygous (1A-12 N Δ /+, M19 N Δ /+), and null (1A-12 Δ / Δ) Mbd4 CH12.F3 cells. Antibody used was anti-IgA conjugated to Allophycocyanin (APC). Histogram plots the average from a single experiment with SEMs (N=4, except for M19N Δ /+, where N=1). Single representative histograms for each genotype from flow cytometry results are shown in the right panel. **B)** Similar to fig. 1A, except that a different antibody was used to detect surface IgA. Here a biotinylated anti-IgA antibody was used first followed by staining with an APC-conjugated streptavidin secondary antibody. Histogram displays averages with standard error of mean (SEMs) from duplicates from two independent experiments (N=4). The M19 N Δ /+ clone was not included in this experiment. Flow cytometry results represent 5000-7500 events gated on viable cells as determined by forward and side scatter.

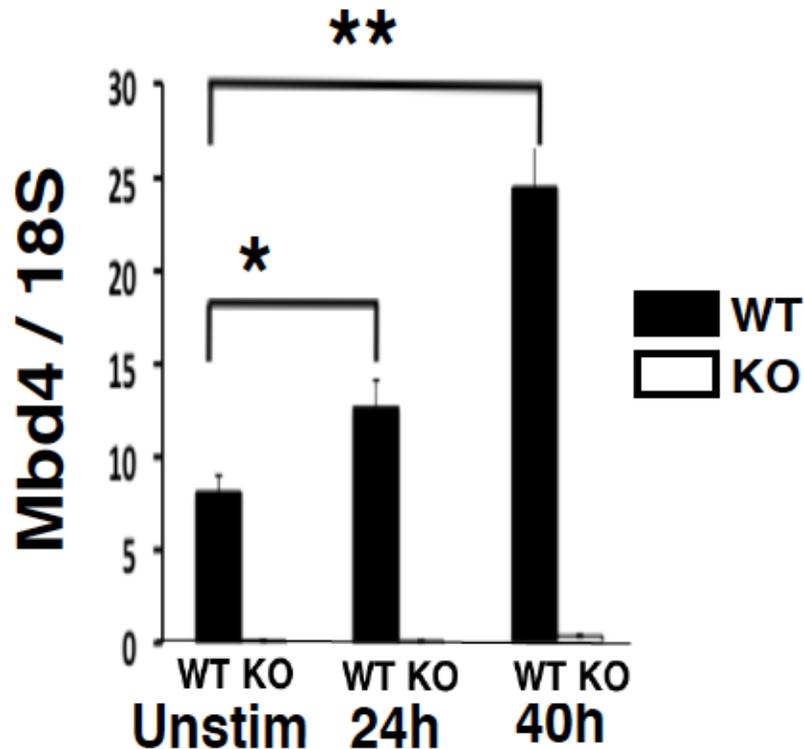


Figure 16. **qRT-PCR analysis confirms lack of Mbd4 mRNA and induction in wild type.** Mbd4 transcripts were analyzed by qRT-PCR using cDNA prepared from RNA isolated from CH12.F3 WT (solid bar) and null (KO) Mbd4 (open bar). Cells were left unstimulated for 24 hours or given CIT (CD40L (25%v/v), IL4 (5ng/ml), and TGF β (0.2ng/ml)) for 24 or 40 hours. All transcript signals were normalized to the control gene, 18S. There was no detectable signal for the Mbd4 transcript in Mbd4 KO cells. WT CH12.F3 cells show induction of the Mbd4 mRNA over time. Histogram depicts the average with SEMs (N=4) of duplicate samples from two independent experiments. Student's two tailed *t* test was used to determine that the induction is statistically significant (**p* < 0.05, ***p* < 0.001).

this particular IgA antibody had a high background. We learned that what had appeared to be a relatively moderate deficit in IgA CSR was actually far more pronounced after correcting for the background by using a different two-step staining method for detecting surface IgA (fig. 15B). In summary, the effect that this particular Mbd4 ablation in CH12.F3 has had on CSR is drastic, approaching a nearly 95% deficit in class switch to IgA.

Lastly, to evaluate whether either the Mbd4 transcript or the Mbd4 protein can be detected in the CH12.F3 Mbd4 KO clone we performed quantitative (q) RT-PCR and Immunoblot analysis, respectively. q RT-PCR demonstrates that the Mbd4 transcript is absent in the Mbd4 KO as compared to WT in unstimulated and CIT-induced cells (fig. 16).

However, in WT cells Mbd4 expression is detectable in unstimulated cells and induced upon exposure to CIT, and this is statistically significant (fig. 16). Western analysis indicates the presence or absence of the 69KD MBD4 protein in WT or Mbd4 KO cells, with or without CIT stimulation, respectively (fig. 17). Furthermore, MBD4 levels were intermediate in the heterozygous clone, 1A-12 $\Delta/+$, in which one Mbd4 allele has been deleted (fig. 17). Collectively, these findings confirm the targeted deletion of Mbd4 in CH12.F3.

4.3 Appropriate proliferation, GLT, and AID expression

CSR can only occur in cells that have undergone several rounds of cell division. In fact the first appearance of switched immunoglobulin *in vitro* does not happen until around the third cell division. Therefore, in order to rule out a proliferation defect associated with Mbd4 deficiency that if present would mar any efforts to analyze CSR in the Mbd4 KO clone, we show the equivalent dilution overtime of an incorporated dye, CFSE (fig. 18). Thus, proliferation is not perturbed in the KO clone. AID is the single most important factor in initiating CSR, and its

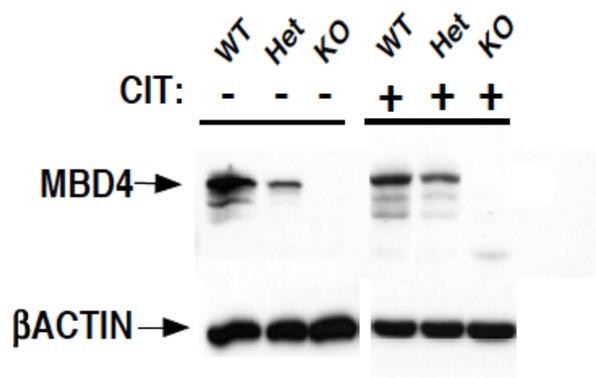


Figure 17. **MBD4 is reduced in an Mbd4 heterozygous clone and abolished in an Mbd4 null clone.** Immunoblot analysis of nuclear extracts prepared from wild type (WT), heterozygous (Het), or null (KO) Mbd4 CH12.F3 cells that were left unstimulated for 24 hours or stimulated with CIT for 24 hours. Full length MBD4 (69kD) was probed with a rabbit anti-MBD4 antibody (Ab) directed against a C-terminal epitope of MBD4 in combination with goat anti-rabbit HRP-conjugated Ab. β -Actin (43kD) was used as a loading control and probed using a mouse anti- β ACTIN Ab followed by a goat anti-mouse HRP-conjugated Ab. A chemilluminiscent substrate was added and visualized with chemilluminiscent films.

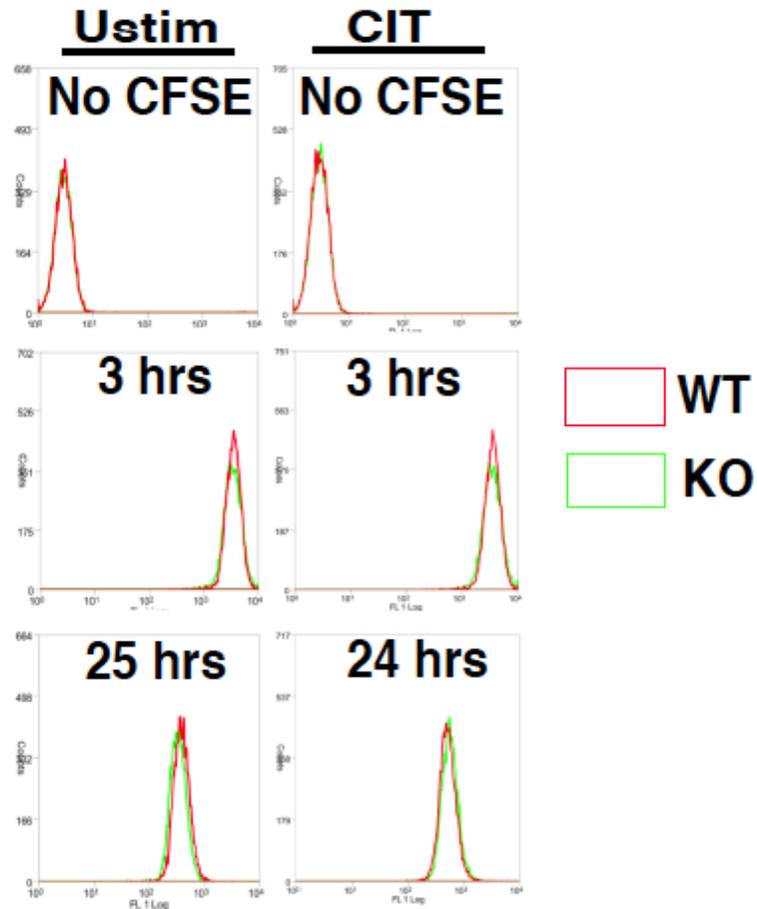


Figure 18. **Mbd4 Null CH12.F3 cells proliferate normally.** To determine whether loss of Mbd4 impacts proliferation CH12.F3 WT (red) or null Mbd4 (green) cells were either left unstimulated or treated with CIT and were stained with the permanent dye, CFSE (15 μ M). Cells were subjected to flow cytometry to monitor dilution of the CFSE dye over time at 3 and 24 hours post staining. Unstained controls (NO CFSE) show the boundary of the dye uptake. This flow cytometry analysis represents 10,000 events that were gated on the viable population of cells as determined by forward and side scatter. A 488nm laser was used to excite the CFSE dye.

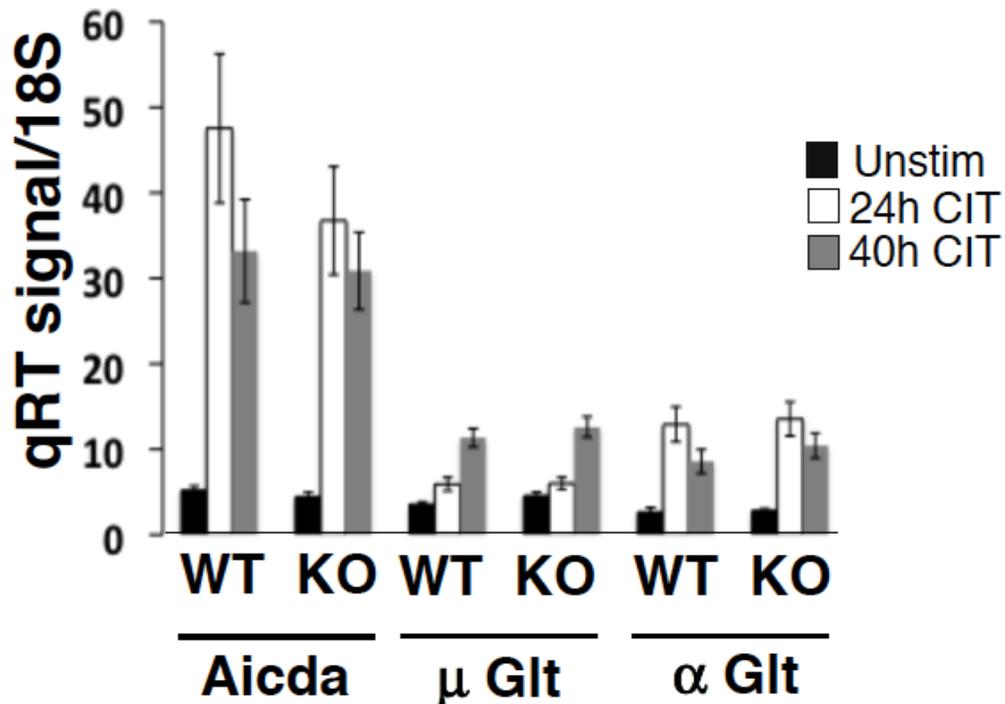


Figure 19. **Aicda** gene transcript, μ GLT, and α GLT are induced normally after CIT stimulation in null (KO) Mbd4 CH12.F3 cells. To ascertain whether loss of Mbd4 impacts expression of AID, μ GLT, and α GLT qRT-PCR was used to measure transcripts for AID, μ GLT, and α GLT from cDNA that was prepared from RNA isolated from wild type (WT) and KO Mbd4 CH12.F3 cells that were left unstimulated (black box) or given CIT for 24 (white box) or 40 (grey box) hours. All transcript signals were normalized to the control gene, 18S. Histogram plots the average with SEMs (N=4) from duplicates from two independent experiments. Student's *t* test was used to determine if statistical significant differences were found between WT and KO Mbd4 CH12.F3 cells. None were found.

ability to deaminate cytosine within DNA is tightly correlated with the induction of germline transcripts (GLTs). Accordingly, we also show by qRT-PCR that inducing CSR in the CH12.F3 Mbd4 KO clone gives rise to μ GLT, α GLT, and AID transcripts at levels that are comparable to WT (fig. 19). It was crucial to show that none of these essential components were altered during making of the Mbd4 KO clone. Therefore, the IgA CSR defect appears to be genuinely related to the loss of Mbd4 in the Mbd4 KO CH12.F3 clone.

4.4 The MMR proteins, MSH2 and MLH1, are reduced in Mbd4 KO CH12.F3

To determine whether MMR proteins were affected by Mbd4 deficiency in CH12.F3 as they were by loss of Mbd4 in the Mbd4 ^{Δ 2-5/ Δ 2-5} mouse model we performed Immunoblot assay to probe for MSH2 and MLH1 using extracts from WT or null Mbd4 cells that were either left unstimulated or given CIT. We see a consistent 40-75% reduction in the steady state level of Msh2 in the Mbd4 KO cell line in both cytoplasmic and nuclear extracts (fig. 20). It also appears that stimulation causes a decrease in MSH2 (fig. 20), but currently we are not certain how this would impact the role of Msh2 in CSR. MLH1 levels are also reduced in nuclear extracts, and we see similar reduction to that of MSH2, around 40% (fig. 20). Clearly, creating a new knock out of Mbd4 in CH12.F3 represents a challenge for studying a direct role of Mbd4 in CSR. If MMR proteins are affected by the loss of Mbd4 then the reduction of CSR in the Mbd4 KO cell line may actually be indirectly linked to the Mbd4 ablation. In summary, our Mbd4 KO in CH12.F3 recapitulates what was found previously in Mbd4 ablation in a mouse system (136), namely that MMR proteins seem to be destabilized by loss of MBD4. Our own micro array data in CH12.F3 also confirms that this is not occurring at the level of transcription. However, the key difference between the cell line and the mouse system is that deletion of Mbd4 in the CH12.F3

eliminates exons 6-8 and causes a reduction of CSR whereas the deletion of exons 2-5 in the mouse does not.

4.5 Analysis of DSBs in the Mbd4 KO CH12.F3 Cell line

Double strand breaks (DSBs) are the obligatory intermediate of the CSR reaction. Consequently deletion of genes involved in the formation of DSBs leads to a decrease in the efficiency of CSR. The contribution that AID and UNG make to the formation of DSBs has been well characterized (55). Loss of Ung leads to a moderate decrease in DSBs, while the loss of AID severely impacts DSB formation. Members of the MMR pathway, including Msh2, Mlh1, and Pms2, also all seem to varyingly contribute to DSB formation. (106). Interestingly, crossing MMR deficiencies with the S μ tandem repeat (TR) KO (106) mouse further exacerbates loss of DSBs (106). The latter finding demonstrates that DSB formation in the absence of the S μ TR region relies at least to some extent on MMR processing (reviewed in chapter 1.4.9).

4.5.1 Reduced double strand breaks (DSBs) in the Mbd4 deficient cells

We sought to determine whether the significant decrease of IgA CSR in the Mbd4 KO CH12.F3 cells (see fig. 15A) is related to inability to form or process DSBs. To extend our characterization of the Mbd4 KO clone, we have used an LM-PCR procedure to detect the incidence of DSBs (9). In this assay an asymmetric, double-stranded oligo linker is unidirectionally ligated to DSBs that are initiated by AID activity. Next the ligated LM-PCR products are amplified using a sequence-specific 5' S μ primer paired with a 3' linker primer. LM-PCR products can then be analyzed by Southern hybridization with an S μ specific probe and associated ³²P counts are quantified. In two independent experiments we find a significant 4-fold decrease (p<0.006) in the number of breaks in the Mbd4 KO clone compared to WT (fig. 21A,

21B). This demonstrates that the decrease of IgA CSR that results from Mbd4 deficiency is linked to a deficit in DSB formation.

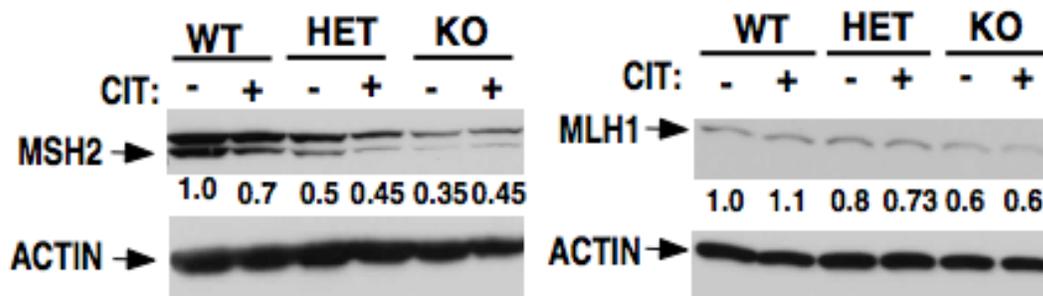


Figure 20. **MSH2 and MLH1 proteins are diminished in the absence of MBD4.** To determine whether MMR proteins were affected by Mbd4 deficiency in CH12.F3 cytoplasmic or nuclear extracts were prepared from WT, heterozygous, and null ($\Delta 6-8$) CH12.F3 cells that were activated with CIT for 24 hours and probed with various antibodies in Western analysis. A chemiluminescent substrate was added and visualized with chemiluminescent films. Spot densitometry was used to quantify intensity of bands. Samples were normalized to β -Actin and WT was set to 1. MSH2 (107 kD) (left panel) was detected in cytoplasmic extracts using a rabbit anti-MSH2 antibody and a goat anti-rabbit HRP-conjugated antibody. MLH1 (95kD) (right panel) was detected using nuclear extracts using a rabbit anti-MLH1 antibody and a goat anti-rabbit HRP-conjugated antibody. β -Actin (43kD) was detected with mouse anti- β ACTIN antibody and a goat anti-mouse HRP conjugated antibody.

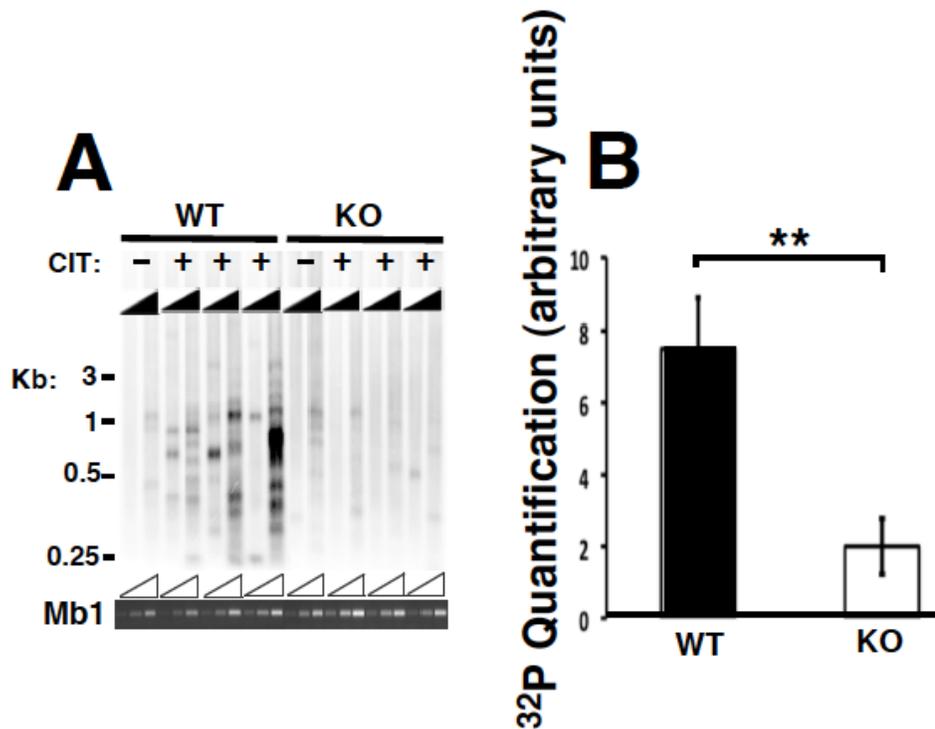


Figure 21. **Mbd4 ablation leads to a substantial reduction of DSBs.** In order to determine whether loss of Mbd4 in the CH12.F3 cell line impacts the formation of double strand breaks (DSBs) Ligation mediated PCR in a Southern analysis was used to quantify DSBs. **A)** LM-PCR products were amplified using agarose plug DNA from WT or KO Mbd4 CH12.F3 that were left unstimulated or given CIT for 12 hours. LM-PCR products were separated on a 1.5% agarose gel, transferred to nylon membranes, and probed with an S_{μ} - specific probe in Southern hybridization. Semi-quantitative PCR of the Mb1 gene was used to normalize input amount from each sample. MB1 PCR was harvested at 29, 31, and 33 cycles. **B)** Quantification of ^{32}P counts from LM-PCR products was done using Imagequant software. Histogram plots the average with SEMs from samples stimulated with CIT (N=3). Student's two-tailed t test was used to determine statistical significance is shown (** $p < 0.01$).

4.5.2 Skewing of DSBs to AG in the Mbd4 KO cell line

The next question to ask was whether residual DSBs in the Mbd4 KO resembled or differed from those in WT. To this end, LM-PCR products were cloned and their DNA sequence was analyzed to locate where the breaks had occurred within S μ . Because this kind of analysis had never been used in the CH12.F3 cell line, it was first important to ask whether the profile of S μ DSBs from WT CH12.F3 matched what was previously shown for WT primary B cells.

In previous primary B cell studies 40-50% of S μ breaks were found in the AID hotspot DGYW/WRCH (where D=G, A, or T, Y=C or T, W= A or T, R=G or A, and H= C, T, or A) consensus, and additionally, almost all breaks (80-95%) occur at the underlined G or C positions (55, 106). The DGYW/WRCH consensus is used here because it covers a wider range of AID hotspots found in S regions, which makes it appropriate when analyzing DSBs. The WRC/GYW consensus that has been described before (see chapter 1.2.3) is used mainly in SHM. Our DGYW/WRCH analysis in CH12.F3 WT cells conforms to data obtained from WT B cells, with 43% of DSBs falling in AID hotspots and 87.0% of DSBs were in Gs or Cs, thus establishing CH12.F3 as a model system for investigating DSB formation.

AGCT, followed GGCT and TGCT, are the most common versions of the DGYW consensus in S μ . Because the breakpoint can occur at any position within the consensus I asked whether there was a preference for nearest neighbor nucleotides preceding the G/C breakpoint and whether this preference reflected the AID hotspot consensus. For example, if we scored a break at AGCT, then it would go into the GC breaks category. We found nearly identical profiles when comparing our CH12.F3 DSBs with those from WT B cells where AG is most common category, followed by TG and GG, which are distributed equally, and then GC. The

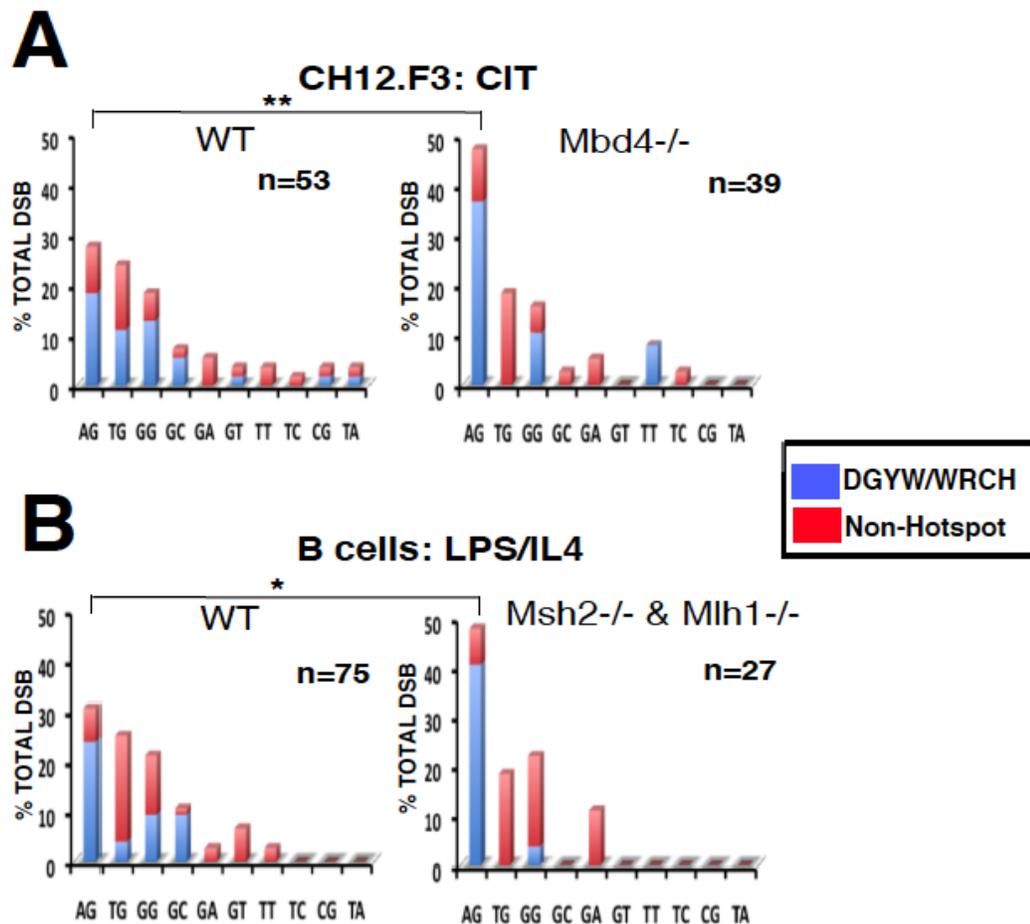


Figure 22. **Breaks are skewed towards AG in both CH12.F3 Mbd4- deficient cells and Msh2- and Mlh1- deficient primary B cells.** To determine whether DSBs from fig. 21 skewed to a particular category, LM-PCR products were sequenced to locate the position of the DSB and placed into various categories in the format NX, where the preceding nucleotide (N) can be any nucleobase, and X marks the nucleotide. χ^2 analysis was performed where appropriate to determine statistical significance (* $p < 0.05$, ** $p < 0.01$) **A**) LM-PCR products were cloned from WT or Mbd4 null (-/-) CH12.F3 cells that were stimulated for 12 hours with CIT. Histograms plot the different categories of breaks versus percent of total DSBs. For WT CH12.F3 (left panel) 53 DSBs were analyzed. For Mbd4 KO, (right panel), 39 DSBs were analyzed. The fraction of DSBs that fall into either DGYW/WRCH AID hotspots (blue) or non-AID hotspots (red) is shown for each category. **B**) Similar to fig. 22A, except that LM-PCR products were cloned from primary B cells that were stimulated with LPS + IL4 for 48 hours. 75 DSBs from WT primary B cells (right panel) are compared to 27 combined DSBs from Msh2- and Mlh1- deficient primary B cells (left panel).

few remaining categories together comprise a minor fraction of all DSBs (see WT CH12.F3 and WT B cells: fig. 22A,22B).

We could now ask whether the residual DSBs that are found the Mbd4 KO cells differ from the WT profile. In terms of G/C breakpoints and AID hotspot usage we did not see obvious differences between WT and KO, 87.0% vs. 84.6% (G/C), and 43% vs. 53% (AID hotspot), respectively. This is an important result because it demonstrates that loss of DSB formation associated with Mbd4 deficiency is not due to a loss of AID or UNG targeting in these cells, although the reduced efficiency of targeting cannot be ruled out. DSBs that have been studied in B cells from mice that lacked either AID or UNG show a strong drop in breaks that occur at G/C basepairs and AID hotspot usage (55). Thus, there is a qualitative difference between the reduction of DSBs that arise from Mbd4 deficiency from that of AID or UNG deficiencies. This suggests that the reduction of DSBs that result from absence of Mbd4 must arise from another mechanism.

Interestingly, deficiencies in MMR proteins lead to a decrease of DSBs while not changing their frequency at G/C basepairs or the DGYW/WRCH consensus. Coupling this fact together with the finding that incidence of S μ DSBs were similar between Mbd4 and Mlh1 deficiencies (106) prompted us compare a DGYW DSB category analysis of our Mbd4 KO with data from B cells that were deficient in MMR. We have found a statistically significant elevated number of AG breaks comparing the WT to Mbd4 KO CH12.F3, 28.3% vs. 46.1%, respectively, ($p < 0.01$) (fig. 22A), which strikingly matches the profile found comparing WT to Msh2 and Mlh1 deficient mice, 30.6% vs. 48.1%, respectively ($p < 0.05$) (fig. 22B) (J.Stavnezer, unpublished data). In both Mbd4 KO CH12.F3 and MMR KO B cells there is also a trend showing a paucity of breaks occurring at TG and GG compared to WT. To summarize, the

pattern of DSBs from Mbd4 resembles that from Msh2 or Mlh1 deficiencies. This suggests that in the absence of Mbd4, and perhaps more in general, that in the absence of MMR, AG breaks are elevated, and breaks at GG and TG are likely diminished. A possible explanation for this phenomenon will be considered in the following section.

4.5.3 Skewing of S μ DBSs in the Mbd4 KO

By analyzing CSR in WT primary B cells one can find a correlation between the propensity of CSR to occur within a certain region of S μ and the density of the AID hotspot consensus within that region (154). Thus, CSR occurs more frequently in the S μ tandem repeat (TR) region, where the DGYW consensus comprises 53% of the sequence in this area, than in the region 5' of S μ located upstream of the TRs, where the DGYW consensus comprises only 26% of the sequence (fig. 23B). The AGCT motif, which is the most common iteration of the DGYW AID hotspot motif in S μ makes up only 2% of the 5' S μ region sequence. Conversely, the AGCT motif is relatively overrepresented in the core S μ TR region representing 40% of sequence (fig. 23B). I wondered whether the DSBs in WT B cells that are skewed to the S μ TRs and frequently associated with AID hotspots are also more frequently located in AGCT motifs.

This is in fact what is found when looking at the distribution CSR in WT B cells, with 17% of CSR occurring in the 5' S μ region versus 73% in the S μ TR region. The assay used in this report uses digestion circularization PCR (DC-PCR) to score for CSR events (154), which is not the same as my DSBs assay, which scores an intermediate in the CSR reaction, but broadly speaking the two assays are similar. I decided to perform an analysis of the location of DSBs in S μ in order to ask whether DSBs in WT CH12.F3 cells were also distributed according to prevalence of the AGCT motif.

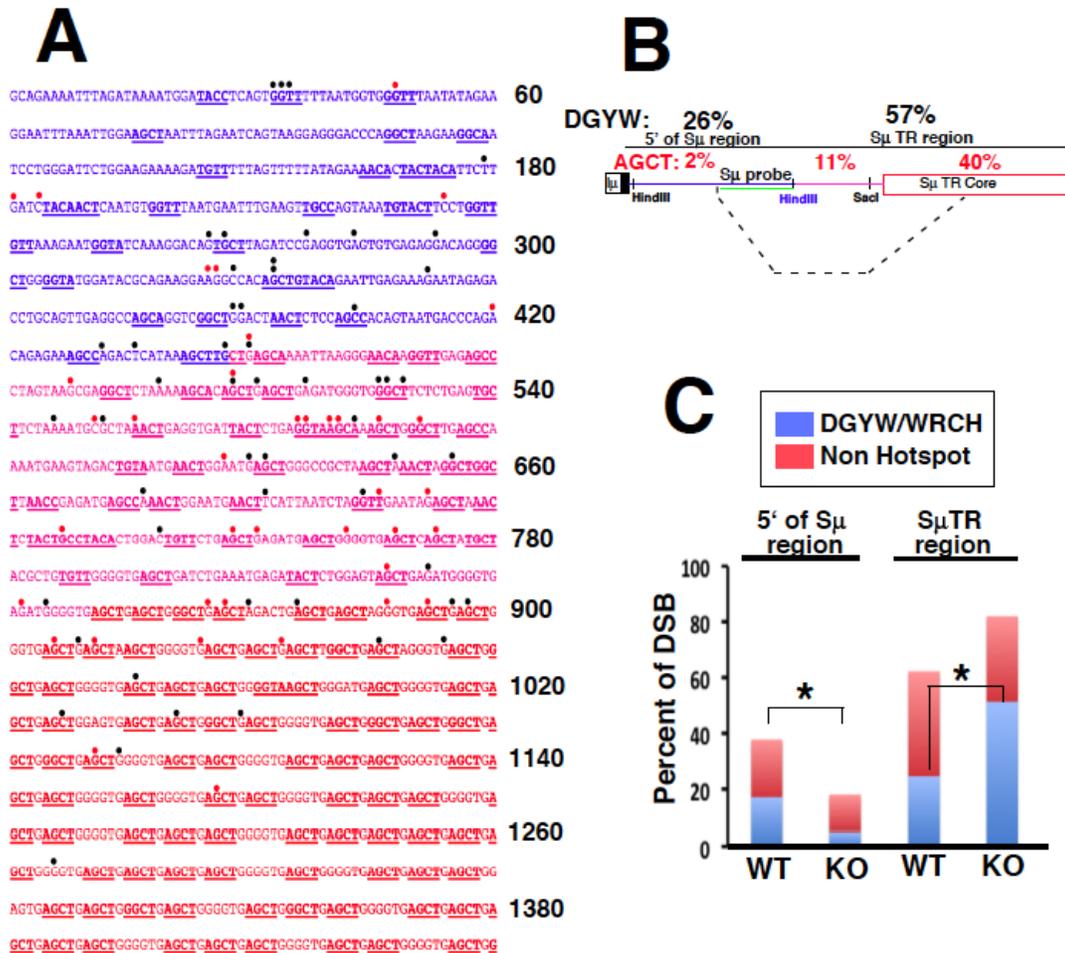


Figure 23. **DSBs skew toward S_μ TR region in the absence of Mbd4.** This figure displays all of the DSBs that were tallied from fig. 22A as they occur in the S_μ sequence of the *Igh* locus. **A)** Positions of DSBs are depicted for WT (closed black circle) and Mbd4 KO (closed red circle). All known DGYW/WRCH AID hotspots in S_μ are underlined and boldfaced. **B)** A map of fig. 23A is shown which demarcates the 5' of S_μ region (blue) and the S_μ TR region (in pink and red) by a Hind III site (blue). The S_μ TRs become more frequent in the S_μ TR core region (red). The location of the S_μ Probe is shown (green). Cloned LM-PCR products are limited to a subsection of S_μ (dashed line). The percent of DGYW in each region is listed in black. The percent of the AGCT motif in each region is listed in red. **C)** Histogram summarizes data from fig. 23A and shows total number of DSBs that occur in 5' of S_μ versus S_μ TR regions for both WT and KO Mbd4. The fraction of DSBs that fall into either DGYW/WRCH AID hot spot (blue) or non- AID hotspot is plotted. Where appropriate χ^2 analysis was used to determine statistical significance (* $p < 0.05$).

The analysis of DSB positioning in WT CH12.F3 shows a reasonably similar distribution, with 39% of DSBs occurring in the 5' S μ region versus 59% in the S μ TR region. Therefore, broadly speaking, the distribution of DSBs in WT CH12.F3 matches what was found for an analysis of CSR in WT B cells, namely that both CSR and DSBs occur more frequently in a region of S μ that has a higher density of the DGYW AID hotspot motif, AGCT.

Studies of B cells that were deficient in Msh2 demonstrated an even greater amount of skewing towards the S μ TR region. CSR was reduced by 3 fold in the region 5' of S μ (154). The increase of CSR within the S μ TR region suggested that Msh2 was especially important for effective CSR in regions where there are few DGYW AID hotspot motifs. (155). Relatedly, in B cells where the S μ TR region is deleted, residual CSR occurs in the 5' S μ region almost exclusively via an MMR-dependent mechanism, because combined deficiency of MMR proteins and S μ TR results in near ablation of CSR. This suggests that when AID induced SSBs occur in the upstream CSR region they must be processed by MMR to result in DSBs (154, 155) (reviewed in chapter 1.4.10). Importantly, it also helps explain why one sees more AG breaks in MMR knockouts, since the AGCT motif becomes increasingly prevalent (40% of all sequence) in the S μ TR region (fig. 23A, 23B).

The next question to ask was to what extent would the DSBs that we have collected in the Mbd4 KO CH12.F3 be distributed unequally in the 5' of S μ region versus S μ TR region (154). Strikingly, the percent of DSBs in the 5' of S μ region drops to less than 18% in Mbd4 KO CH12.F3 compared with 40% in the WT and this 2-fold drop is statistically significant (fig. 23C), demonstrating increased skewing towards the S μ TR region. Overall, this pattern is similar to what has been previously shown for Msh2 deficiency (154). Lastly, the remaining DSBs found in the S μ TR region from the Mbd4 KO cluster more heavily towards AID hotspots compared to

WT, 62.5% vs. 39.4%, which is statistically significant ($p < 0.05$) (fig. 23C). Elevated AID hotspot incidence is probably a consequence DSBs not being channeled away from the S_{μ} TR region, where AID hotspots (especially AGCT) are more densely packed (see fig. 23A, 23B). This indicates that when Mbd4 is absent DSBs are largely confined to or forced into the S_{μ} TR region, and the increase of AG breaks in the Mbd4 KO (see fig.22A) can also be explained by the fact that the AGCT is more prevalent in S_{μ} TR region. This phenomenon closely mirrors what was found for Msh2 deficiency, and is explained by lack of processing by MMR that does not allow for DSBs to be channeled away from the S_{μ} TR region. This establishes yet another criterion suggesting that Mbd4 possibly functions in the MMR pathway during CSR.

4.6 S_{μ} - S_{α} microhomology length is increased in Mbd4 KO clone

The striking similarity of the Mbd4 KO DSB phenotype to that of various MMR proteins prompted us to ask whether microhomology based joining in the absence of Mbd4 differs from WT. In WT primary B cells, typically, 20-50% of CSR-mediated joins are direct, while the remaining joins show a degree of overlap, typically 1-3 nucleotides of identity, between S_{μ} donor and S_{α} acceptor germline sequences, which is defined as microhomology. It is rare to find junctions with more than five nucleotides of microhomology (97). Those joins that are direct or that involve shorter 1-3bp of microhomologies are indicative of classical non-homologous end joining (C-NHEJ), while joins with microhomologies greater than 4bp are considered to arise by an alternative end joining pathway (A-EJ) (115). Interestingly, various MMR proteins have been assayed for their contribution to end joining. In general, MSH2/MSH6/EXO1 deficiencies produce joins with shorter microhomologies relative to WT junctions, while Mlh1/PMS2 deficiencies give rise to longer microhomologies (94, 96, 98). In view of the similar effect that

ablation of MMR proteins has on CSR, it is not currently understood why a difference should exist in the quality of joins in an Msh2 vs. and Mlh1 KO, although one model has been proposed (98). It should be noted, however, that when any of the MMR deficiencies are combined with the S μ TR KO, then all microhomologies become longer, like that of Mlh1 or Pms2 deficiency alone (98).

The close resemblance of the Mbd4 KO DSB analysis with MMR deficiency begs the question, what do junctions look like the absence of Mbd4? We had no a priori knowledge of whether Mbd4 KO junctions would mirror the shorter based microhomologies that were found with ablation of Msh2, or the longer microhomologies that are associated with the loss of Mlh1. Based on the similar incidence of DSBs between our Mbd4 KO and previous Mlh1 ablation reports, it seemed reasonable to hypothesize that microhomologies at S μ -S α junctions would be longer in the absence of Mbd4. However, the altered positioning of DSBs in the Mbd4 KO also mimicked what was described previously for Msh2 deficiency. Therefore, Mbd4 ablation could resemble either Msh2 or Mlh1 loss, or perhaps neither, when concerning microhomologies at S μ -S α junctions.

We find differences when comparing S μ -S α junctions from WT versus Mbd4 KO CH12.F3, and curiously, they are similar to what has been described for Mlh1 and Pms2 deficiencies. Mbd4 ablation leads to a 33% increase in the average microhomology length compared to WT, 2.82 vs. 1.94, $p=0.06$ (fig. 24A), a result which is trending towards statistical significance. A compendium of all the S μ -S α junctions identified in this report can be found in appendix B. Dissecting the data further revealed more obvious differences, namely that the frequency of direct joins is substantially diminished by more than 3 fold in the Mbd4 KO

compared to WT, 8.3% versus 29.4%, respectively, which is statistically significant ($p < 0.025$) (fig. 24B). S μ -S α junctions with greater than 4bp of microhomology were conversely increased

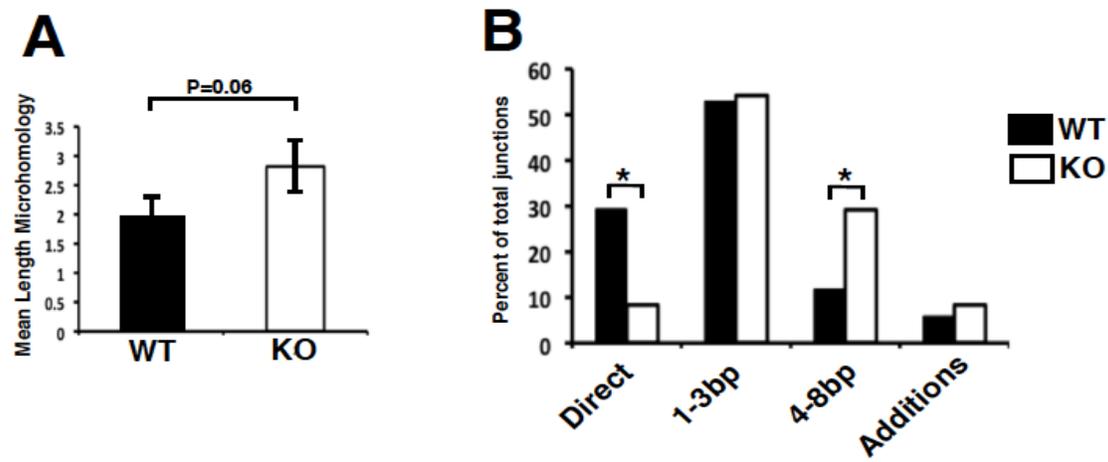


Figure 24. **S μ -S α junctions from Mbd4 deficient CH12.F3 cells have longer microhomologies.** In order to determine whether loss of Mbd4 impacts the extent of microhomology found at S μ -S α junctions CH12.F3 cells were stimulated to undergo IgA CSR for 24 hours and genomic DNA was prepared for PCR using primers that amplify the S μ -S α junctions. WT CH12.F3 (solid squares); Null (KO) Mbd4 (open squares). **A**) The average length of microhomology from 34 WT and 24 KO S μ -S α junctions is shown with SEMs. Student's single tailed *t* test was used to determine statistical significance. **B**) Histogram places S μ -S α junctions into different categories of 0 bp of microhomology (direct), 1-3bp of microhomology, 4-8bp of microhomology, or insertions at the junction (additions). The percent of total junctions that each category makes up is displayed. Where possible χ^2 analysis was performed to determine statistical significance (**p* < 0.05).

by nearly 3 fold in the Mbd4 KO compared to WT, 29.2% versus 11.8%, respectively, ($p < 0.025$) (fig. 24B). Both of these findings duplicate what has been documented for S μ -S γ junctions arising from Mlh1 ablation in primary B cells (98). This suggests that Mbd4 is potentially involved in MMR-based processing of DSBs and that it is especially important for generating blunt DSBs, which are the favored substrate of C-NHEJ. The striking similarity of S μ -S α junctions in our Mbd4 knock out to those found in Mlh1 and Pms2 deficiencies reiterates that the effect Mbd4 deficiency has on CSR is probably more akin to loss of Mlh1, then it is to loss of Msh2.

4.7 Retroviral transduction and genetic complementation in Mbd4 KO cell line

To test whether genetic complementation with Mbd4 rescues the IgA CSR defect in the Mbd4 KO CH12.F3 clone, full length Mbd4, various versions of it, including catalytically inactive forms, and truncated variations that do not contain the N-terminal exons of the protein, were introduced via stable retroviral transduction (see appendix C for various Mbd4 constructs).

Surprisingly, catalytically inactive forms, including a lyase (Y514F), or a glycosylase mutant (D534A), and additionally a shorter construct that was cloned from the Mbd4 ^{Δ 2-5/ Δ 2-5} mouse that only contained exons 6-8, all seemed to rescue to the same extent as full length WT MBD4. The differences when compared to empty vector were small (~20%) in this experiment for all constructs that were interrogated (fig. 25A). Western analysis was performed to confirm that full length WT protein was expressed,. WT MBD4 protein could be detected in the reconstituted Mbd4 KO CH12.F3 and the level of expression was comparable to that of

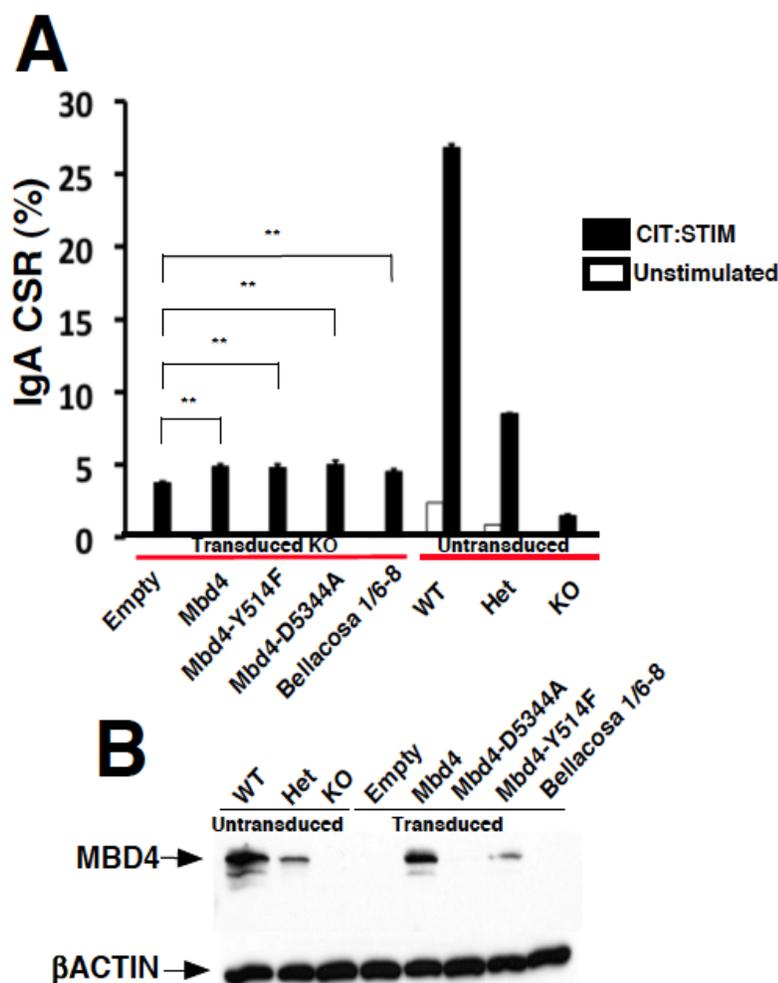


Figure 25. **Genetic complementation with MBD4 in IgA rescue does not coincide with expression of protein.** To determine whether reconstitution of MBD4 rescues the IgA CSR defect associated with the Mbd4 KO CH12.F3 cell line the cDNA for Mbd4 and various iterations of it were introduced by retroviral transduction. Untransduced WT, heterozygous, and null (KO) Mbd4 cells were included as controls. **A**) KO Mbd4 CH12.F3 cells were transduced with retrovirus encoding WT MBD4, catalytic mutants (Y514F, D534A), or a cDNA from the Mbd4 KO (Δ 2-5) mouse (Mbd4 E1/6-8). Cells were selected in puromycin (2ng/ml) for 2-4 days. Cells were then taken out of drug, left unstimulated or given CIT for 24 hours, and subjected to the IgA CSR assay. Histogram plots the average percent of IgA CSR with SEMs. For stimulated samples, N=3; for unstimulated samples, N=1. Where possible student's *t* test was used to determine statistic significance (** $p < 0.01$). **B**) Cells (2×10^6) were prepared for Western analysis from unstimulated samples in 25A and probed for MBD4 (69kDal) (indicated by a horizontal arrow) using anti-MBD4 rabbit antibody. β ACTIN (43kDal) (indicated by horizontal arrow), probed with anti- β ACTIN antibody was used as a loading control.

endogenous WT and heterozygous Mbd4 CH12.F3 cells (fig. 25B). The lyase mutant was expressed at a reduced amount, and the glycosylase was barely detectable. This raised doubt as to whether the differences that arose in the IgA CSR assay when comparing Mbd4 constructs to the empty vector control could be attributed to authentic Mbd4 expression. Clearly, the near absent level of MBD4 protein of the glycosylase mutant (D534A) should not show a comparable amount of IgA CSR rescue as WT MBD4. Based on this finding it seems likely that another factor, unrelated to Mbd4 expression, was responsible for the small differences that were measured in comparing empty vector - to MBD4- transductions.

One crucial difference between the empty vector control and the other constructs is the size of the plasmid used to make retrovirus. The smaller empty vector plasmid leads to a higher virus titer, which translates to a larger infected pool that will grow out faster in puromycin selection. Therefore, because of this difference we had to harvest the empty vector virus-transduced population first, storing them in liquid nitrogen, which allowed the other cultures to catch up, before also freezing them in liquid nitrogen. On average the empty vector virus-infected cultures would grow out two to three days faster. Unfortunately CH12.F3 spontaneously undergoes CSR in culture in the absence of inducers, and the shorter time that the empty vector virus- infected cells spent in culture could account for the lower rate of CSR. For the following experiment, the cells that were transduced with empty vector virus were kept in culture for the same amount of time as the other constructs. And indeed, in a conservative experiment comparing just empty to full length WT Mbd4, we found that the difference in IgA CSR became indistinguishable (fig. 26A). Keeping the cells in puromycin during stimulation with CIT did not lead to any detectable difference in the amount of CSR between empty vector- and Mbd4-

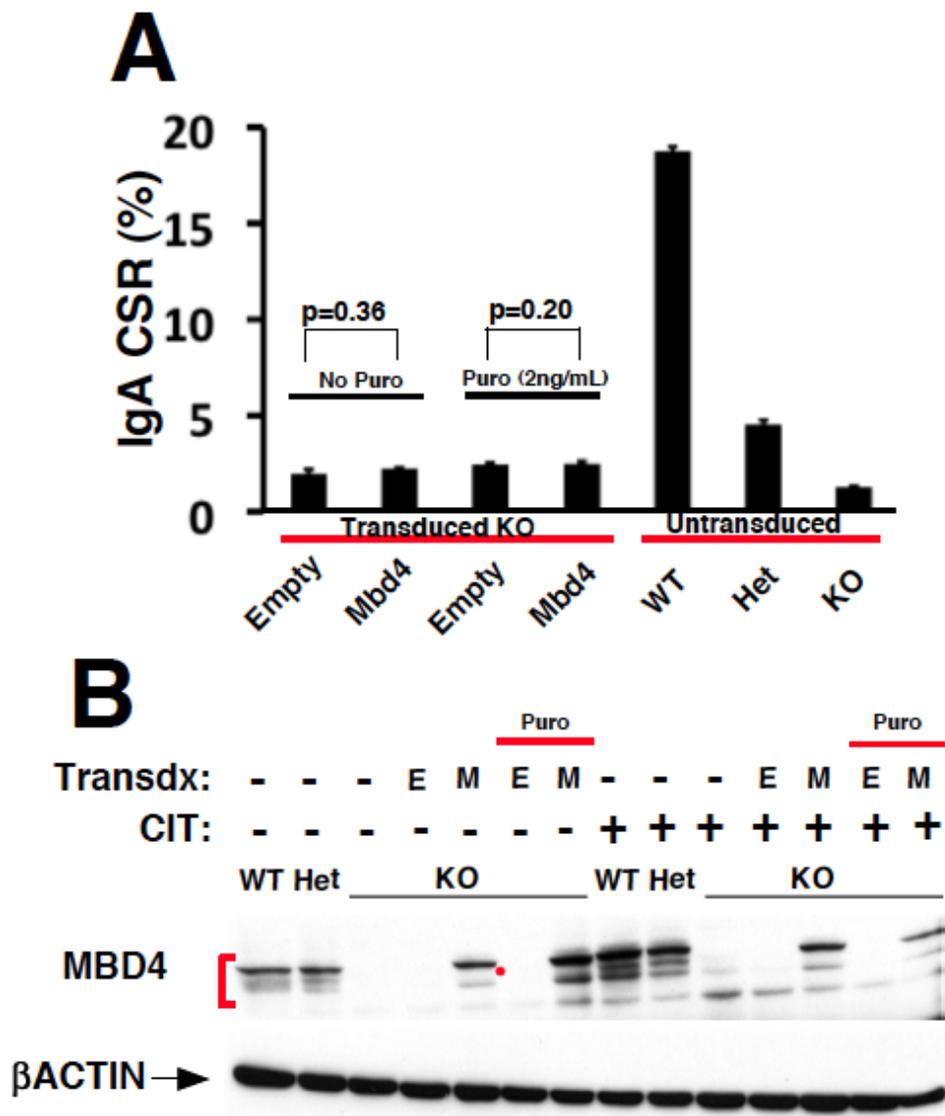


Figure 26. **Genetic complementation with MBD4 in Mbd4 deficient cells fails to rescue IgA CSR.** Repeat experiment of fig. 25, except that cells were kept in selection for the entire duration of experiment and empty vector populations were grown for the same number of days. **A)** Null Mbd4 CH12.F3 cells were transduced with retrovirus encoding WT MBD4 or empty vector, selected in puromycin (2ng/ml) for 2-4 days, kept in or removed from drug during unstimulated or CIT conditions for 24 hours, and subjected to the IgA CSR flow cytometry assay. Untransduced WT, heterozygous, and null Mbd4 cells were included as controls. Histogram plots the average with SEMs (N=3). Student's t-test was performed and statistics are shown. **B)** Cells (2×10^6) were prepared for Western blot for validating expression of retrovirally introduced MBD4 samples from fig. 26A. Red bracket denotes endogenous pattern of MBD4. The middle isoform (red dot) is conspicuously absent in Mbd4 KO CH12.F3 transduced with MBD4 cDNA. β ACTIN (43kDa), indicated by horizontal arrow, was used as a loading control.

transduced populations. Interestingly, it appeared that stimulated cultures were suffering from a lack of MBD4 induction (fig. 26B). The amount of protein provided by the retroviral system during IgA CSR stimulation appears to be sub-optimal and perhaps below a certain threshold (i.e. less than the level of the endogenous Het) no rescue can be attained. Therefore, we decided to move away from this system.

One other key finding from this experiment is that transduction with full length Mbd4 cDNA (exons 1-8) did not recapitulate the endogenous profile of Mbd4 expression. Clearly, one can note that the regular triplet pattern of isoforms from endogenous Mbd4 expression is not re-established after transduction of the cDNA encoding exons 1-8 of Mbd4 (fig. 26B) The conspicuous absence of the middle isoform is consistent throughout transduced samples. The lack of the middle isoform may be important for CSR and will have to be investigated further (see chapter 5.3 and appendix A).

The poor expression of Mbd4 in transduced Mbd4 KO cells stimulated to undergo CSR prompted us to develop another strategy for re-introducing Mbd4 into the Mbd4 KO CH12.F3 clone. We opted to make stably-transfected pools that could then be stimulated to undergo CSR to IgA. This approach would rely on electroporation of pCDNA-based Mbd4 plasmids. To minimize the amount of work that might go into a fully exhaustible interrogation of all the available different Mbd4 constructs we limited ourselves to empty vector, WT full-length MBD4, and a construct containing exons 4-8 of Mbd4 (glycosylase domain, see appendix A). Added crucial controls included parallel transfections in WT CH12.F3 cells.

Once again, however, we were not able to detect any difference between the empty vector versus the MBD4 constructs in the IgA CSR assay (fig. 27A). It is noteworthy that in this experiment the time in culture after introduction of plasmid DNA is identical for all constructs.

Hence, transfection of the pCDNA vector did not discriminate between the larger sized Mbd4 constructs and the smaller empty vector. Additionally, Western analysis showed that stimulation did not affect Mbd4 expression in samples induced to switch to IgA (fig. 27B).

However, there was a statistically significant ~ 30% decrease of IgA CSR in WT CH12.F3 cells that were transduced with full length WT MBD4 ($p= 0.006$) (fig. 27A). This was unexpected and perhaps it points to the reason why rescue in the KO Mbd4 CH12.F3 has failed. To wit upon further introspection it became obvious that this situation mirrored what was found when we overexpressed Mbd4 in WT primary B cells (see fig. 12C). Moreover, the modulatory function of Mbd4 on CSR could directly conflict with any attempt to rescue CSR in the Mbd4 KO cell line. In fact there is a similar reduction of IgA CSR in the Mbd4 KO CH12.F3 that was transfected with MBD4 ($p =0.005$), but the decrease here (18%) is small. A construct expressing exons 4-8 also failed to rescue IgA, but we have not validated expression of the truncated protein.

In summary, two different strategies have been used in an attempt to rescue the IgA CSR defect in the Mbd4 KO CH12.F3 cells. Neither strategy has been successful, but both have advanced the thinking that has gone into future attempts. Suboptimal levels of expression, the failure to reproduce the endogenous pattern of Mbd4 expression, and the finding that overexpression actually seems to decrease IgA CSR are all possible reasons why the experiment thus far has not succeeded. The implications of this are discussed in chapter 5.

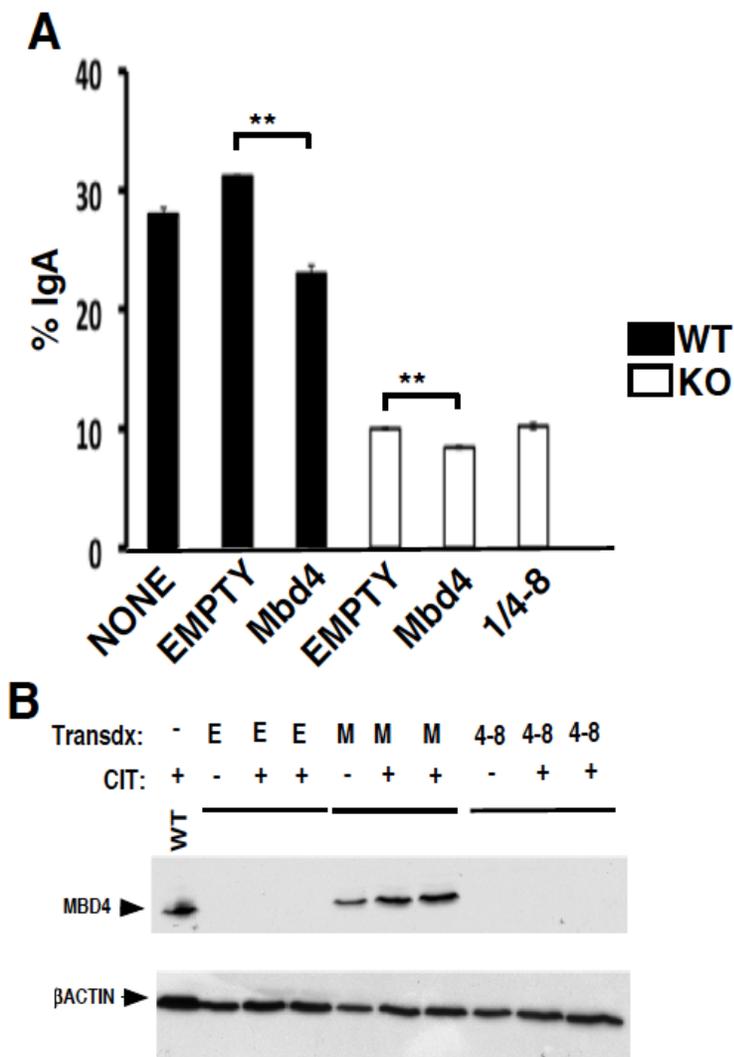


Figure 27. pCDNA expression system failed to rescue IgA CSR defect in Mbd4 KO cells. To determine whether a pCDNA vector based strategy would work to rescue the IgA CSR defect associated with the Mbd4 KO CH12.F3 cell line, WT or Mbd4 null (KO) CH12.F3 cells were transfected with either Empty (E), Mbd4 (M), or 1/4-8 (4-8) pCDNA constructs and selected in G418 (750 μ M) for 5 days. Cells were maintained in the presence of drug throughout the duration of the experiment. **A)** Summary of IgA CSR assay using WT (solid bars) and KO (open bars) transfected cells. Histograms show the average percent of IgA CSR with SEMs. (N=2) Student's *t* test was used to determine statistical significance (** $p < 0.01$). **B)** Western analysis probed for MBD4 (69kD) with anti-Mbd4 antibody from nuclear extracts prepared from WT or KO cells that were unstimulated (-) or given CIT (+) for 24 hours. β Actin (43kD) was used as a loading control.

CHAPTER 5
DISCUSSION

5.1 Summary

The aim of this study was to determine whether the mismatch DNA glycosylase, Mbd4, participates in CSR. To test this hypothesis gene targeted homologous recombination was used to create a novel deletion of the Mbd4 gene, which specifically removed the exons encoding the catalytic properties of the protein (Mbd4^{Δ6-8/Δ6-8}). The initial characterization of the Mbd4 KO CH12.F3 cell line revealed a substantially diminished CSR phenotype. This phenotype was not associated with impairments to other requisite CSR properties, such as GLTs, AID, and proliferation, suggesting that the CSR phenotype in the Mbd4 KO cell line was genuinely related to the loss of Mbd4. Furthermore we have seen in three independent instances of targeting Mbd4 for deletion that there is a concomitant decrease in IgA CSR (see fig 15A, 15B). Finding a corresponding reduction of IgA CSR in three separate clones that have lost Mbd4 expression seems unlikely to be attributed to chance. Moreover, we have also isolated multiple clones from the WT and KO populations by limiting dilution and found that the reduced CSR phenotype was preserved in KO clones. This result strongly indicated that the reduced CSR phenotype found in the Mbd4 KO clones is not a spurious result. It should also be noted that previous gene targeting attempts in the CH12.F3 cell line have not reported this kind of phenomenon (116, 157)

Moreover a deeper examination of Mbd4 in CSR by way of the Mbd4^{Δ6-8/Δ6-8} cell line has demonstrated that Mbd4 contributes to DSB formation and shares phenotypes with the MMR proteins, MSH2, and MLH1. These proteins have a documented role in CSR. Our unique analysis of AG DSBs in the Mbd4^{Δ6-8/Δ6-8} cell line and Msh2 and Mlh1 deficient B cells has revealed a contrasting profile with what happens in a WT background. The increase of AG DSBs found in the Mbd4 KO relative to WT is correlated with the distribution of DSBs skewing to a region where the AGCT AID hotspot motif is more frequent. Broadly speaking, this DSB result

is congruent with a similar account of *Msh2* deficiency, which demonstrated a reduced ability to channel DSB formation away from the TR region of S μ . The analysis of altered S μ -S α junctions in the Mbd4 ^{Δ 6-8/ Δ 6-8} cell line unequivocally matches the phenotype of longer microhomologies at junctions associated with Mlh1 deficiency. In view of the similarity to other MMR phenotypes in CSR perhaps it is not surprising to find that MMR repair proteins are reduced in the Mbd4 ^{Δ 6-8/ Δ 6-8} cell line. The destabilization of MMR proteins in the Mbd4 ^{Δ 6-8/ Δ 6-8} cell line has made it difficult to conclude whether Mbd4 has a direct or indirect role in CSR, but regardless of its role in CSR, the connection of Mbd4 to MMR merits further examination.

5.2 **Failure to rescue the IgA CSR defect with genetic complementation**

The inability to rescue the IgA CSR defect associated with Mbd4 ^{Δ 6-8/ Δ 6-8} cell line has not deterred me from advancing the hypothesis that Mbd4 is involved in CSR. There are several reasons that can account for the failure to rescue.

A straightforward explanation is that the appropriate level of Mbd4 expression has not been achieved. Certainly, the finding that overexpression of MBD4 reduced IgA CSR in WT CH12.F3 cells gives a hint that this system cannot be used as is for successful genetic complementation (fig. 27A). This result is reminiscent of enforced UNG2 expression, which we have demonstrated also inhibits CSR in WT primary B cells (see chapter 3.4). In light of the UNG finding, clearly too much of the protein seems to negatively impact CSR. If like UNG MBD4 is involved in repairing the U/G mismatch, then flooding the system with too much protein may force a straightforward repair mechanism over the mutagenic program that is required for CSR. Retrovirally introduced SMUG1, but not TDG, can rescue the CSR defect associated with *Ung* deficiency, which shows that not all glycosylases fit into the framework of

the CSR reaction (83). It's possible that enforced Mbd4 expression might rescue the CSR defect in the UNG system as well, which would give another clue that Mbd4 fits into the CSR apparatus. In total these data suggest that a carefully titrated amount of Mbd4 is required to promote efficient CSR. Lastly, we suspect that based on our findings from the Mbd4^{Δ6-8/Δ6-8} cell line Mbd4 is involved at some level in MMR, suggesting that re-stabilizing MMR proteins will also likely play a role in successful genetic complementation. For these reasons we have considered moving to a different expression system that can be carefully titrated. Fusing AID to the estrogen receptor (AID-ER) has been widely utilized in B cells. In this system one can control the amount of AID that enters the nucleus in response to tamoxifen (159). A similar approach with Mbd4 will be pursued.

We are also curious to know if the endogenous expression pattern of Mbd4 that was not recapitulated with transduction of the full-length cDNA of Mbd4 is important for rescue. The middle isoform is clearly missing in transduced samples. We have identified other Mbd4 transcripts from WT B cells and CH12.F3 that based on their sizes are predicted to possibly give rise to the middle isoform (see appendix A, fig. 1C,1D). The top isoform corresponds to full length Mbd4 and the middle isoform is smaller by about 4-6 kD, which would translate to a transcript that was around 50-75 basepairs smaller than the full length mRNA. Exon 4 is precisely 75 basepairs, making the Δ4 Mbd4 transcript a promising candidate. We are currently testing this possibility.

Another reason that may account for the inability to rescue the IgA CSR defect is the presence of a residual transcript in the Mbd4^{Δ6-8/Δ6-8} cell line encoding exons 1-5 that could potentially act as a dominant negative. I reject this as a possibility for the following reasons. In an Affymatrix gene expression array experiment (available on request) we found that this

transcript is expressed at a level that is at least 100 fold reduced compared to WT. A semi-quantitative PCR result confirms this finding (see appendix D, fig.30B). To follow-up this finding we are currently using 3' rapid amplification of cDNA ends (3'RACE) to clone this residual mRNA. This would allow us to determine whether it can be translated. We next plan to overexpress this hypothetical dominant negative transcript in a WT background to see if CSR is inhibited. Although I do not expect to see this effect, it would provide further evidence for the failure to rescue the CSR IgA defect in the Mbd4^{Δ6-8/Δ6-8} cell line, namely that a dominant negative effect stemming from a residual transcript precludes any attempts to rescue the CSR defect.

A large portion of my time has been devoted to attempting to rescue the IgA CSR defect associated with the Mbd4^{Δ6-8/Δ6-8} cell line. However, this effort has failed. Along the way, we have acquired knowledge that will hopefully lead to the objective at hand. I have cited several possible explanations here for the failure to rescue the IgA CSR defect of the Mbd4^{Δ6-8/Δ6-8} cell line. These possibilities will be tested soon.

At about the same time of completing the Mbd4^{Δ6-8/Δ6-8} KO in CH12.F3, we began to characterize primary B cells from the Mbd4^{Δ2-5/Δ2-5} mouse (150). We proceeded testing for CSR deficiencies to IgG1 and IgG3 in these B cells and surprisingly found that there were none. The observations that were made from examining the Mbd4^{Δ6-8/Δ6-8} cell line suggested a role for Mbd4 in CSR, but the lack of a CSR phenotype in the Mbd4^{Δ2-5/Δ2-5} mouse gave us pause. However, I will provide a model that reconciles the striking difference in phenotypes between the Mbd4^{Δ2-5/Δ2-5} mouse and the Mbd4^{Δ6-8/Δ6-8} cell line.

5.3 Deletion of exons 2-5 of Mbd4 did not produce a CSR phenotype

This brings the discussion to the perplexity of the $Mbd4^{\Delta 2-5/\Delta 2-5}$ mouse lacking a phenotype in CSR. To begin, it is important to reiterate the significant finding that MMR proteins are affected by the loss of Mbd4 in the $Mbd4^{\Delta 2-5/\Delta 2-5}$ mouse. Assaying for MMR proteins in this Mbd4 deficient background showed that all MMR proteins were substantially reduced but not entirely ablated (between 2.3- 5.8 fold) (150). Our investigation of B cells from the same $Mbd4^{\Delta 2-5/\Delta 2-5}$ mouse revealed a profound defect in the stability of MSH2 and MLH1 proteins (fig. 12A). As stated earlier, several reports have demonstrated the importance of the MMR proteins for promoting CSR (see chapter 1.4.6, 1.4.7). **In $Msh2^{+/-}$ and $Mlh1^{+/-}$ mice, which have a 50% reduction of MSH2 and MLH1 proteins, respectively, there is already a measurable CSR defect (J. Stavnezer, personal communication). Therefore it is difficult to comprehend the $Mbd4^{\Delta 2-5/\Delta 2-5}$ mouse, which has a near 90% reduction of these proteins and yet does not have a reduced CSR phenotype.** This sets up the obvious comparison between the $Mbd4^{\Delta 2-5/\Delta 2-5}$ mouse and the $Mbd4^{\Delta 6-8/\Delta 6-8}$ cell line.

The question becomes why does the $Mbd4^{\Delta 2-5/\Delta 2-5}$ mouse, which has severely reduced levels of MMR proteins not display a defective CSR phenotype? Even if one ignores Mbd4 in this picture they are still faced with the fact that MMR proteins are reduced to a point in the $Mbd4^{\Delta 2-5/\Delta 2-5}$ mouse where a CSR phenotype should exist. In fact I was convinced prior to obtaining these mice that the MMR deficit associated with these mice would guarantee a CSR phenotype. As mentioned previously, certain isotypes are more affected by MMR deficiency than others (see chapter 1.4.6, 1.4.7). It's possible that if the isotypes that are more affected by MMR deficiency (i.e., IgG2a and IgG2b) had been studied we would have seen a CSR defect. I would argue, however, that if the IgG1 and IgG3 CSR phenotypes were subtle in the $Mbd4^{\Delta 2-5/\Delta 2-5}$ mouse, then the fact that we checked CSR at day 3 (where reductions in CSR due to genetic

alterations are more pronounced) as well as day 4 should have revealed a difference (see fig. 22A).

There is a profound difference in the CSR phenotype of the Mbd4^{Δ2-5/Δ2-5} mouse and the Mbd4^{Δ6-8/Δ6-8} cell line. The Mbd4^{Δ2-5/Δ2-5} mouse displays surprisingly normal CSR, while the Mbd4^{Δ6-8/Δ6-8} cell line is severely comprised in CSR, and this is in spite of the fact that in both systems MMR proteins are reduced. What could account for this puzzling result? It is important to note that the Mbd4 gene expresses a number of alternatively spliced transcripts, and that exon 6 has a strong in frame Kozak which could give rise to a truncated protein that includes most of the glycosylase domain (see appendix A). If one puts this observation together with the results from the contrasting phenotypes of the Mbd4^{Δ2-5/Δ2-5} mouse and the Mbd4^{Δ6-8/Δ6-8} cell line then a model can be formulated. I suspect the Mbd4^{Δ2-5/Δ2-5} mouse does not have a CSR phenotype because it expresses a truncated MBD4 protein that retains glycosylase and/or scaffolding functions related to its interactions with MLH1. We can extrapolate from the Mbd4^{Δ2-5/Δ2-5} mouse phenotype that in WT cells MBD4 is not just involved in MMR protein stability, but that is also competes with or modulates MMR activity during CSR. For example, MBD4 may normally direct AID induced DNA lesions toward high fidelity repair rather than CSR. Our results from overexpressing MBD4 in a WT background support this hypothesis. We can also extrapolate from the reduced CSR phenotype Mbd4^{Δ6-8/Δ6-8} cell line that in WT cells the C-terminus of Mbd4 is required for CSR. This potentially explains why the Mbd4^{Δ2-5/Δ2-5} mouse does not have a CSR defect associated with the Mbd4^{Δ2-5/Δ2-5} deletion. It appears that low levels of MMR proteins are sufficient to support CSR the Mbd4^{Δ2-5/Δ2-5} mouse, because the MMR-modulating function of MBD4 is lost in the deletion, but the function associated with the C-terminus that is essential for CSR is still present.

Clearly, finding a truncated MBD4 protein moiety in the Mbd4^{Δ2-5/Δ2-5} mouse is crucial to understanding the differences between the Mbd4^{Δ2-5/Δ2-5} mouse and the Mbd4^{Δ6-8/Δ6-8} cell line. It should be noted that all the crystal structures of the glycosylase domain of Mbd4 that have been reported to date show potentially crucial deoxythymidine (dT from the T/G mismatch) binding residues, Q423, and V422, that would be missing in a hypothetical truncated isoform from the Mbd4^{Δ2-5/Δ2-5} mouse (137, 140). Key residues involved in catalysis, including Y514, D534, K536, and L480 would still be present, but since this truncated construct has not been assayed for enzymatic function, there is no way to know whether or not it retains catalytic function.

On the other hand, it's possible that Mbd4 (and I infer that it's the C-terminus that is important) does not participate in CSR catalytically, and that it instead serves a scaffolding function. The lack of residues that may be important for catalysis and that are missing in the hypothetical truncated isoform would therefore not rule out a scaffolding hypothesis for Mbd4 in CSR. We have not been successful in identifying a truncated species of Mbd4 in the Mbd4^{Δ2-5/Δ2-5} mouse, but it could be difficult to detect using our Mbd4 antibody. In the past we have had trouble seeing the full-length species of MBD4, so if the truncated species is less abundant then it may not be possible to see it. The easiest way to address this issue would be to add an Immuno-tag to the hypothetical Mbd4^{Δ2-5/Δ2-5} isoform, such as, c-myc, or V5 that could then be probed with validated antibodies.

Another perplexity surrounding the Mbd4^{Δ2-5/Δ2-5} mouse is that it does not develop tumors. The etiology of HNPCC has direct genetic links to defective MMR genes, with tumorigenesis occurring somatically, often associated with loss of heterozygosity (LOH) (142, 143) (see chapter 1.6.4). Sporadic cancers of various tissues also show defective MMR genes. These results have been reproduced in a controlled lab setting using mouse models that

demonstrate loss of Msh2, Mlh1, and Pms2 triggers tumorigenesis (147, 148). It seems puzzling that a mouse with nearly 10 fold reductions in MSH2 and MLH1 would not develop tumors (fig. 12A). I suppose it can be argued that since Msh2^{+/-} mice do not develop tumors in the controlled lab setting (147), then the residual amount of MSH2 and MLH1 in the Mbd4^{Δ2-5/Δ2-5} mouse **(albeit much less than the 50% of MSH2 found in an Msh2^{+/-} mouse)** is enough to suppress the initiation of tumors. In light of the contrasting CSR phenotypes of the Mbd4^{Δ2-5/Δ2-5} mouse and the Mbd4^{Δ6-8/Δ6-8} cell line it is tempting to hypothesize that a mouse constructed to contain our Δ6-8 deletion will develop tumors where the Δ2-5 deletion has not. However, it may not be possible to make comparisons between CSR and cancer phenotypes.

In conclusion, a model has now been introduced that explains the lack of CSR phenotype in the Mbd4^{Δ2-5/Δ2-5} mouse. From the data that I have generated there is one outstanding inconsistency between the mouse and cell line results. Loss of Mbd4 in both models leads to destabilized MMR protein levels (see fig. 12A, 20), and yet CSR was only affected in the cell line. In addition, the extent to which MSH2 and MLH1 are reduced in the Mbd4^{Δ2-5/Δ2-5} mouse should have produced a CSR phenotype. It is not possible to have an 8 fold reduction of either MSH2 or MLH1 and not display a reduced CSR phenotype. However, if the modulating effect that Mbd4 possesses over MMR is removed from the equation, then MMR proteins may still promote WT levels of CSR, despite their reduced stability. This model also stipulates that the function provided by the C-terminus of MBD4, which is essential for CSR, as evidenced by the Mbd4^{Δ6-8/Δ6-8} cell line, is retained in the Mbd4^{Δ2-5/Δ2-5} mouse in the form of a truncated species. Relatedly, we have attempted to rescue the IgA CSR defect in the Mbd4^{Δ6-8/Δ6-8} cell line using the residual cDNA from the Mbd4^{Δ2-5/Δ2-5}, but it does not work. I will now provide a specific example of why this particular cDNA does not work.

Regarding the apparent inability to genetically complement the $Mbd4^{\Delta 6-8/\Delta 6-8}$ cell line with the residual transcript that was found in the $Mbd4^{\Delta 2-5/\Delta 2-5}$ mouse we have found an anti-sense transcript (CN781668) that has been deleted in our $Mbd4$ KO cell line but is still present in the $Mbd4^{\Delta 2-5/\Delta 2-5}$ mouse (see appendix A, fig. 1G). We now recognize a long non-coding RNA (lncRNA) is transcribed from a divergent promoter shared with *CD3ORF25* in the opposite orientation of the *Mbd4* transcript and that partially overlaps with exon 8 was deleted during construction of the $Mbd4$ KO. Transcription initiation of these transcripts has not been affected, as scored by RT-PCR indicating that the promoter is still functioning (our unpublished data). It's possible that lncRNAs could influence expression of *Mbd4* (reviewed in (160)). The recent finding that the HOTAIR lncRNA interacts with the Polycomb complex (PRC2) and remodels chromatin architecture shows that lncRNAs can have trans-acting functions (161). Importantly, lncRNAs can also bind protein and act as scaffolds. Perhaps the CN781668 lncRNA actually participates in conjunction with *Mbd4* in promoting CSR. We have recently cloned the CN781668 transcript and will be attempting to re-introduce it into the $Mbd4^{\Delta 6-8/\Delta 6-8}$ cell line. It could be that for IgA CSR rescue, both protein and lncRNA need to be present. Therefore this possibly explains why the rescue in the cell line does not work with the residual transcript from the $Mbd4^{\Delta 2-5/\Delta 2-5}$ mouse, **to wit the CN781668 transcript is not altered in the $Mbd4^{\Delta 2-5/\Delta 2-5}$ mouse, but is missing in the $Mbd4^{\Delta 6-8/\Delta 6-8}$ cell line.**

5.4 Concluding remarks

The original discovery of *Mbd4* as an *Mlh1*-binding partner coupled with our recent DSB and $S\mu$ - $S\alpha$ junction data from the $Mbd4^{\Delta 6-8/\Delta 6-8}$ cell line provides new tantalizing evidence for an epistatic relationship between the two genes. I must stress that it is not currently possible to

deduce from these data whether Mbd4 is directly or indirectly involved in CSR. It is possible that Mbd4 simply serves to stabilize Mlh1, and potentially Msh2, as well, via a scaffolding role. In this regard the lack of availability of stable catalytic mutants (Y514F, D534A) of Mbd4 has stymied any attempts to test for a *bona fide* enzymatic dependence of CSR on Mbd4 *in vivo* (see fig. 23B). A recently characterized glycosylase catalytic mutant (D534N) may be put to good use in that regard (140).

The study of the involvement of Mbd4 in CSR opens a new area of research that could potentially lead to new and exciting paradigms that move the field forward. The analysis of DSBs and S μ -S α junctions in cells that are deficient in Mbd4 is of key interest because it suggests that Mbd4 is linked to MMR. The observation that absence of Mbd4 causes DSBs to skew towards a region of S μ where the AID hotspot is frequent is in agreement with data that implicated a similar phenomenon for Msh2 deficiency. Furthermore, this observation is congruent with the model that predicts DSB formation in the absence of MMR can only occur when SSBs are in close proximity (see chapter 1.4.9). The finding that Mbd4 affects the stability of MMR proteins has now been documented in two independent systems, which creates an impetus to study the mechanism of interaction between Mbd4 and MMR proteins. At the same it has complicated the objective at hand, which was to assess a direct role of Mbd4 in Ab diversification. Indeed, it may be difficult to separate Mbd4 from MMR, as we have seen from the characterization of DSBs and S μ -S α junctions in the Mbd4 KO, which is nearly identical to what has been already described for MMR deficiency.

Our investigation was somewhat derailed initially after finding out that a Mbd4 ^{$\Delta 3/\Delta 3$} mouse did not manifest any clear phenotypes in CSR or SHM (162). Interestingly the genetic manipulation of this mouse only targeted exon 3, which does not contain any of the catalytic

domains of Mbd4. Moreover, a residual transcript is still made in this mouse (146), setting up a possible scenario where a truncated protein retaining the catalytic and/or scaffolding properties of Mbd4 as they relate to Mlh1 could continue to support CSR. This report mirrors the analysis of Mbd4^{Δ2-5/Δ2-5} mouse that was tested for CSR in this thesis. Finally, it is not clear whether the targeted deletions of Mbd4 in these either of these mice truly ablate expression of the protein. Our data in CH12.F3 features a knock-out of catalytic domains of the DNA glycosylase/lyase (exons 6-8) of Mbd4, and this negatively impacted CSR, leaving open the possibility that Mbd4 is involved in CSR.

APPENDICES

Appendix A

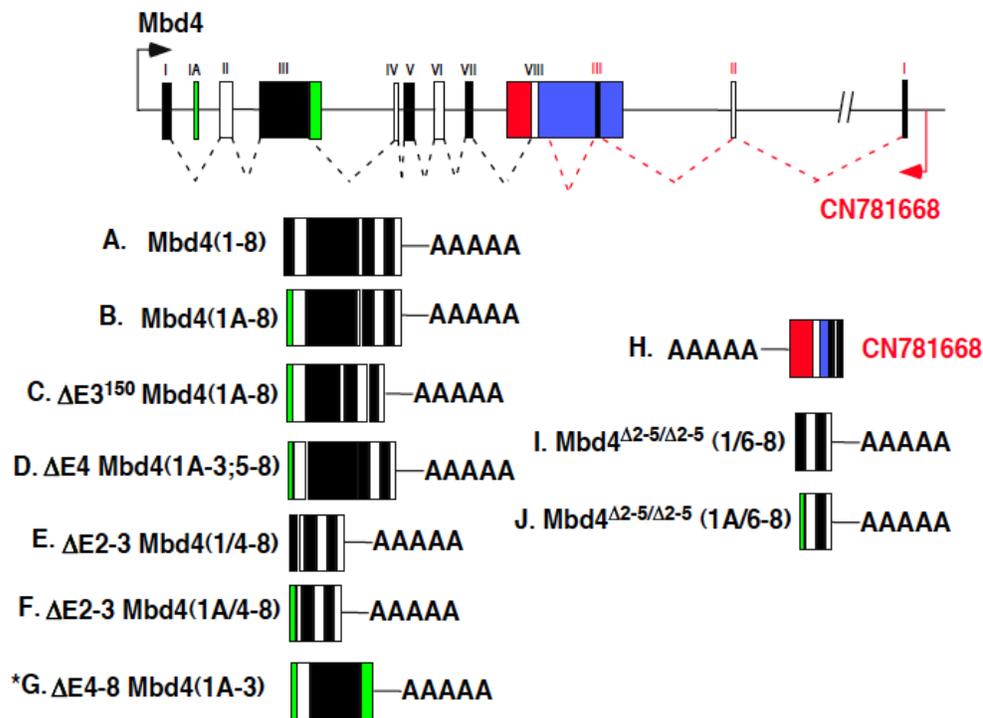


Figure 28. **Transcription in the Mbd4 Locus is complex and overlaps with CN781668.** Exons 1-8 of the Mbd4 locus and Exons 1-3 of CN781668 are depicted in alternating black and white colors and labeled with roman numerals (black, Mbd4; red, CN781668). The prototypical splicing pattern of full length Mbd4 (1-8) is demarcated by dashed lines (black). Intronic sequences of Mbd4 (green) can be spliced into mRNA. Transcription of CN781668 is anti-sense to Mbd4 and the splicing pattern is shown by dashed lines (red). The 3' UTR of Mbd4 (blue) is shown. Exon 8 from Mbd4 and parts of the 3' UTR overlap with CN781668. Intronic sequence (red) from Mbd4 is part of the CN781668 mRNA. **A)** Full length Mbd4 mRNA (exons 1-8). **B)** Alternate full length Mbd4 uses an alternate first exon (1A-8). **C)** Similar to **B**, except that the c-terminal 150 amino acids are spliced out of exon 3 ($E3\Delta 150$). **D)** Similar to **B**, except that exon 4 is spliced out ($\Delta E4$). **E)** Similar to **A** except that exons 2 and 3 are spliced out ($\Delta E2-3$). **F)** Similar to **B**, except that exons 2 and 3 are spliced out ($\Delta E2-3$). **G)** Similar to **B** except that mRNA incorporates intronic sequence immediately following Exon 3 and terminates prior to exon 4. **H)** CN781668. **J)** A residual Mbd4 transcript is found in the Mbd4 $\Delta 2-5/\Delta 2-5$ KO mouse (1/6-8). **I)** Similar to **J**, except that the alternate first exon is used (1A/6-8). (*indicates that this message uses a different polyadenylation sequence from other Mbd4 transcripts).

WT #19 ($S\mu$ 250/ $S\alpha$ 59)

TGAGCAAATTAAGGGAACAAGGTTGAGAGCCCTAGTAAGCGAGGCTCTAAAAAGCACAGCTGAGCTGAGATGGGTGGGCTTCTCTGAGTGCTTCTAAAA
 |||
 TGAGCAAATTAAGGGAACAAGGTTGAGAGCCCTAGTAAGCGAGGCTCTAAATAGGGTGGAAATGGGCTGAACAAGGCTGAGCTTACCTAGACCGGGCAGA
 |||
 CTGAACTAGTATAAACTTGGCTAGGCTACAATGGATTGAGCTGAGCTAGACTTAGGGTGGAAATGGGCTGAACAAGGCTGAGCTTACCTAGACCGGGCAGA
 |||

KO #36 ($S\mu$ 511/ $S\alpha$ 169)

AGCTAAACTCTACTGCCTACTGACTGTTCTGAGCTGAGATGAGCTGGGGTGGCTCAGCTATGCTACGCTGTGTTGGGGTGGCTGATCTGAAATGA
 |||
 AGCTAAACTCTACTGCCTACTGACTGTTCTGAGCTGAGATGAGCTGTGGAAGGGCTGGCTGAGCTAGACTAGGCTGGGCTGAGCTGGAATGAGCTG
 |||
 CTAGGATAAACTAAGCTGGGATGAGACAGGCTGGACTGCAGGAGGAAGACTGGAAGGGCTGGCTGAGCTAGACTAGGCTGGGCTGAGCTGGAATGAGCTG
 |||

KO #48 ($S\mu$ 132/ $S\alpha$ 247)

TGTACAGAATTGAGAAAGAATAGAGACCTGCAGTTGAGGCCAGCAGGTCGGCTGGACTAACTCTCCAGCCACAGTAATGACCCAGACAGAGAAAGCCAGA
 |||
 TGTACAGAATTGAGAAAGAATAGAGACCTGCAGTTGAGGCCAGCAGGCTTGGCTGAACTGGGCTAAGCTGGGATGGACTAGGATAAACTAAGCTGGGA
 |||
 ACAAAGCTAGGCTACTGCCTACTGTCTGGCTAGGCTGTACTGGAATGAGCTGAGCTGAACTGGGCTAAGCTGGGATGGACTAGGATAAACTAAGCTGGGA
 |||

WT #13 ($S\mu$ 498/ $S\alpha$ 144)

ATCTAGGTTGAATAGAGCTAACTCTACTGCCTACTGACTGTTCTGAGCTGAGATGAGCTGGGGTGGCTCAGCTATGCTACGCTGTGTTGGGGTGA
 |||
 ATCTAGGTTGAATAGAGCTAACTCTACTGCCTACTGACTGTTCTGAGCTGAGATGAGCTGGGGTGGCTCAGCTATGCTACGCTGTGTTGGGGTGA
 |||
 ACAGGCTGGACTGCAGGAGGAAGACTGGAAGGGCTGGCTGAGCTAGACTAGGCTGGGCTGAGCTGGAATGAGCTGGGTTGAGCTGAACTAGTATAAACTT
 |||

WT #16** ($S\mu$ 74/ $S\alpha$ 438)

CGAGGTGAGTGTGAGAGGACAGGGGCTGGGTTATGGATACGCAGAGGAAGGCCACAGCTGTACAGAATTGAGAAAGAATAGAGACCTGCAGTTGAGGCC
 |||
 CGAGGTGAGTGTGAGAGGACAGGGGCTGGGTTATGGATACGCAGAGGACTGGGCTGGTGCAGGCTGAGTTAGGCTGAGCTGAGCTGGAATGAGATGGAA
 |||
 GGAGAGGAGAGGAGAGGAGAGGAGGAGGAGCTAGGCTGGAATAGGTTGGACTGGGCTGGTGCAGCTGAGTTAGGCTGAGCTGAGCTGGAATGAGATGGAA
 |||

WT #7 ($S\mu$ 318/ $S\alpha$ 222)

GCTGAGATGGGTGGGCTTCTCTGAGTGCTTCTAAAATGCGCTAACTGAGGTGATTACTCTGAGGTAAGCAAAGCTGGGCTTGGCCAAAATGAAGTAGA
 |||
 GCTGAGATGGGTGGGCTTCTCTGAGTGCTTCCAAAATGCGCTAAA-TGAGGTGATTTGTTGGACTAGGATAAACTAAGCTGGGATGAGACAGGCTGGAC
 |||
 ACTGCACTGTCTGGCTAGGCTGTACTGGAATGAGCTGAGCTGAACTGGGCTAAGCTGGGATGGACTAGGATAAACTAAGCTGGGATGAGACAGGCTGGAC
 |||

KO #44 ($S\mu$ 505/ $S\alpha$ 95)

TGAATAGAGCTAACTCTACTGCCTACTGACTGTTCTGAGCTGAGATGAGCTGGGGTGGCTCAGCTATGCTACGCTGTGTTGGGGTGGCTGATCT
 |||
 TGAATAGAGCTAACTCTACTGCCTACTGACTGTTCTGAACTGAGATTGGCTAGGCTACAATGGATTGAGCTGAGCTAGACTTAGGGTGGAAATGGGC
 |||
 AGGCTGGGCTGAGCTGGAATGAGCTGGGTTGAGCTGAACTAGTATAAACTTGGCTAGGCTACAATGGATTGAGCTGAGCTAGACTTAGGGTGGAAATGGGC
 |||

KO #45 ($S\mu$ 377/ $S\alpha$ 80) -14bpdeletion in $S\alpha$

AAGCAAAGCTGGGCTTGGCCAAAATGAAGTAGACTGTAATGAACTGGAATGAGCTGGGCCGCTAAGCTAACTAGGCTGGCTTAACCGAGATGAGCCAA
 |||
 AAGCAAAGCTGGGCTTGGCCAAAATGAAGTAGACTGTAATGAACTGGATTGAGCTGAGCTAGACTTAGGGTGAATG-----AGCTTAC
 |||
 TGAGCTGGGTTGAGCTGAACTAGTATAAACTTGGCTAGGCTACAATGGATTGAGCTGAGCTAGACTTAGGGTGGAAATGGGCTGAACAAGGCTGAGCTTAC
 |||

Figure 29. Compendium of $S\mu$ - $S\alpha$ junction sequences. (Continued)

KO #42 (S μ 259/S α 147)

AAGGGAACAAGGTTGAGAGCCCTAGTAAGCGAGGCTCTAAAAAGCACAGCTGAGCTGAGATGGGTGGGCTTCTCTGAGTGCTTCTAAAATGCGCTAAACT
 |||
 AAGGGAACAAGGTTGAGAGCCCTAGTAAGCGAGGCTCTAAAAACACAGACTAGGCTGGGCTGAGCTGGAATGAGCTGGGTTGAGCTGAACTAGTAT
 |||
 GATGAGACAGGCTGGACTGCAGGAGGAAGACTGGAAGGGCTGGCTGAGCTAGACTAGGCTGGGCTGAGCTGGAATGAGCTGGGTTGAGCTGAACTAGTAT

KO #41 (S μ 481/S α 291)

TGGAATGAACTTCATTAATCTAGGTTGAATAGAGCTAAACTCTACTGCCTACACTGGACTGTTCTGAGCTGAGATGAGCTGGGGTGGAGCTCAGCTATGCT
 |||
 TGGAATGAACTTCATTAATCTAGGTTGAATAGAGCTAAATTTCTACTGCTAGGCTACACTGCACTGTCTGGCTAGGCTGTACTGGAATGAGCTGAGCTGA
 |||
 AAAGTGAAGCTAAGCTAGGCTGAAATGGGCTGAGCAGAGCTGGACAAAGCTAGGCTACACTGCACTGTCTGGCTAGGCTGTACTGGAATGAGCTGAGCTGA

KO #39 (S μ 503/S α 199)

TTGAATAGAGCTAAACTCTACTGCCTACACTGGACTGTTCTGAGCTGAGATGAGCTGGGGTGGAGCTCAGCTATGCTACGCTGTGTTGGGGTGGAGCTGATC
 |||
 TCGAATAGAGCTAAATTTCTACTGCCTACACTGGACTGTTCTGAGCTGAGGATGAGACAGGCTGGACTGCAGGAGGAAGACTGGAAGGGCTGGCTGAGCT
 |||
 GCTGAGCTGAACTGGGCTAAGCTGGGATGGACTAGGATAAACTAAGCTGGGATGAGACAGGCTGGACTGCAGGAGGAAGACTGGAAGGGCTGGCTGAGCT

KO #38 (S μ 309/S α 338)

CTGAGCTGAGATGGGTGGGCTTCTCTGAGTGCTTCTAAAATGCGCTAAACTGAGGTGATTACTCTGAGGTAAGCAAAGCTGGGCTTGAGCCAAAATGAAG
 |||
 CTGAGCTGAGATGGGTGGGCTTCTCTGAGTGCTTCTAAAATGCGCTAAACTGGAAGTGAAGTGGCTGAAATGGGCTGAGCAGAGCTGGACAAAG
 |||
 GATGGAATAGGCTGGGCTGGCTGGTGTGAGCTGAGCTAGGCTGAGCTGCGCTAAACTGAGCTAAGCTAGGCTGAAATGGGCTGAGCAGAGCTGGACAAAG

KO #37 (S μ 710/S α 135)

TGAGCTAAGCTGGGGTGGAGCTGAGCTGAGCTTGGCTGAGCTAGGGTGGAGCTGGGCTGAGCTGGGGTGGAGCTGAGCTGAGCTGGGGTAAAGCTGGGATGAGC
 |||
 TGAGCTAAGCTGGGGTGGAGCTGAGCTGAGCTTGGCTGAACTAGGGTGGAGCTGGGCTGGGTTGAGCTGAACTAGTATAAACTGGCTAGGCTACAA
 |||
 CAGGAGGAAGACTGGAAGGGCTGGCTGAGCTAGACTAGGCTGGGCTGAGCTGGAATGAGCTGGGTTGAGCTGAACTAGTATAAACTGGCTAGGCTACAA

KO #34 (S μ 665/S α 210)

CTGAGCTAGACTGAGCTGAGCTAGGGTGGAGCTGAGCTGGGTGGAGCTGAGCTAAGCTGGGGTGGAGCTGAGCTGAGCTTGGCTGAGCTAGGGTGGAGCTGGGC
 |||
 CTGAGCTAGACTGAGCTGAACTAGGGTGGAGCTGAGCTGGGTGGAGCTGAGCTAAGCTAAGCTGGGATGAGACAGGCTGGACTGCAGGAGGAAGACTGGAAG
 |||
 TGTACTGGAATGAGCTGAGCTGAACTGGGCTAAGCTGGGATGGACTAGGATAAACTAAGCTGGGATGAGACAGGCTGGACTGCAGGAGGAAGACTGGAAG

KO #25 (S μ 75/S α 53)

AGTGCTTAGATCCGAGGTGAGTGTGAGAGGACAGGGGCTGGGGTATGGATACGCAGAAGGAAGGCCACAGCTGTACAGAATTGAGAAAGAATAGAGACCT
 |||
 AGTGCTTAGATCCGAGGTGAGTGTGAGAGGACAGGGGCTGGGGTATGGTACGCAGAAAGGAAATGGGCTGAACAAGGCTGAGCTTACCTAGACCGGGCAGA
 |||
 CTGAACTAGTATAAACTGGCTAGGCTACAATGGATTGAGCTGAGCTAGACTTAGGTTGGAATGGGCTGAACAAGGCTGAGCTTACCTAGACCGGGCAGA

WT #14 (S μ 789/S α 323)

TGGGGTAAAGCTGGGATGAGCTGGGGTGGAGCTGAGCTGAGCTGGAGTGGAGCTGAGCTGGGCTGAGCTGGGGTGGAGCTGGGCTGAGCTGGGCTGAGCTGGGC
 |||
 TGGGGTAAAGCTGGGATGAGCTGGGGTGGAGCTGAGCTGGAGTGGAGCTGAGCTGGGCTGAGCTGGGCTGAGCTGGGCTGAGCTGGGCTGAGCTGGGC
 |||
 TGGCTGGTGTGAGCTGAGCTAGGCTGAGCTGCGCTAAACTGAGCTAAGCTAGGCTGAAATGGGCTGAGCAGAGCTGGACAAAGCTAGGCTACACTGCCT

WT #18 (S μ 388/S α 286) - 49bp deletion in S α

GTAGACTGTAATGAACTGGAATGAGCTGGGCCGCTAAGCTAAACTAGGCTGGCTTAACCGAGATGAGCCAACTGGAATGAACTTCATTAATCTAGGTTG
 |||
 GTAGACTGTAATGAACTGGAATGAGCTGGCTACACTGCTGTGTGG-----
 |||
 ATGGGCTGAGCAGAGCTGGACAAAGCTAGGCTACACTGCACTGTCTGGCTAGGCTGTACTGGAATGAGCTGAGCTGAACTGGGCTAAGCTGGGATGGACT

Figure 29. Compendium of S μ -S α junction sequences. (Continued)

NWT #5 ($S\mu 350/S\alpha 422$)

CAAAGCTGGGCTTGAGCCAAAATGAAGTAGACTGTAATGAACTGGAATGAGCTGGGCCCTAAGCTAACTAGGCTGGCTTAACCGAGATGAGCCAACT
 |||||
 CAAAGCTGGGCTTGAGCCAAAATGAAGTAGACTGTAATGAACTGGAATGACGAGCTGAGTTAGGCTGAGCTGAGCTGGAATGAGATGGAATAGGCTGGGC
 |||||
 GGAGAGGAGAGGAGAGGAGCTAGGCTGGAATAGGTTGGACTGGGCTGGTGCGAGCTGAGTTAGGCTGAGCTGAGCTGGAATGAGATGGAATAGGCTGGGC

NWT #7 ($S\mu 161/S\alpha 56$)

GACCTGCAGTTGAGGCCAGCAGGTCGGCTGGACTAACTCTCCAGCCACAGTAATGACCCAGACAGAGAAAGCCAGACTCATAAAGCTTGCTGAGCAAAAT
 |||||
 GACCTGCAGTTGAGGCCAGCAGGTCGGCTGGACTAACTCTCCAGCCACAGTAAGTTGGAATGGGCTGAACAAGGCTGAGCTTACCTAGACCGGGCAGACA
 |||||
 GAACTAGTATAAACTTGCTAGGCTACAATGGATTGAGCTGAGCTAGACTTAGGGTGAATGGGCTGAACAAGGCTGAGCTTACCTAGACCGGGCAGACA

NWT #8 ($S\mu 680/S\alpha 289$)

GCTGAGCTAAGCTGGGGTGGCTGAGCTGAGCTTGGCTGAGCTAGGGTGGCTGAGCTGGGGTGGCTGAGCTGAGCTGAGCTGAGCTGGGGTAACTGGGATGA
 |||||
 GCTGAGCTAAGCTGGGGTGAAGCTAGGCTACACTGCCTGTGTGG-----TGGACTAGG
 |||||
 GGCTGAGCAGAGCTGGACAAAGCTAGGCTACACTGCCTGTCTGGCTAGGCTGTACTGGAATGAGCTGAGCTGAACTGGGCTAAGCTGGGATGGACTAGG

NWT #9 ($S\mu 96/S\alpha 345$)

GGGCTGGGGTATGGATACGCAGAAGGAAGGCCACAGCTGTACAGAATTGAGAAAGAATAGAGACCTGCAGTTGAGGCCAGCAGGTCGGCTGGACTAACT
 |||||
 GGGCTGGGGTATGGATACGCAGAAGGAAGGCCACAGTTGTACAGAATTGTGCGCTAACTGAGCTAAGTTAGGCTGAAATGGGCTGAGCAGAGGTTGGAC
 |||||
 GAGAGATGGAATAGGCTGGGCTGGCTGGTGTGAGCTGAGCTAGGCTGAGCTGCGCTAACTGAGCTAAGCTAGGCTGAAATGGGCTGAGCAGAGCTGGAC

NWT #10 ($S\mu 201/S\alpha 244$)

TCTCCAGCCACAGTAATGACCCAGACAGAGAAAGCCAGACTCATAAAGCTTGCTGAGCAAAATTAAGGGAACAAGGTTGAGAGCCCTAGTAAGCGAGGCT
 |||||
 TCTCCAGCCACAGTAATGACCCAGACAGAGAAAGCCAGACTCATAAAGCTTGCTGAAATGGGCTAAGCTGGGATGGGCTAGGATAAACTAAGCTGGGATG
 |||||
 AAAGCTAGGCTACACTGCCTGTCTGGCTAGGCTGTACTGGAATGAGCTGAGCTGAACTGGGCTAAGCTGGGATGGACTAGGATAAACTAAGCTGGGATG

NWT #11 ($S\mu 487/S\alpha 143$)

TGAACTTCATTAATCTAGGTTGAATAGAGCTAACTCTACTGCCTACACTGGACTGTCTGAGCTGAGATGAGCTGGGGTGGCTCAGCTATGCTACGCT
 |||||
 TGAACTTCATTAATCTAGGTTGAATAGAGCTAAATCTACTGCCTACACTGGCTGGGCTGAGCTGGAATGAGCTGGGCTGAGCTGAACTAGTATAAACTT
 |||||
 ACAGGCTGGACTGCAGGAGGAAGACTGGAAGGGCTGGCTGAGCTAGACTAGGCTGGGCTGAGCTGGAATGAGCTGGGTTGAGCTGAACTAGTATAAACTT

NWT #12 ($S\mu 268/S\alpha 130$)

CAAGGTTGAGAGCCCTAGTAAGCGAGGCTCTAAAAGCACAGCTGAGCTGAGATGGGTGGGCTTCTCTGAGTGCTTCTAAAATGCGCTAAAAGTGGGTTGA
 |||||
 CAAGGTTGAGAGCCCTAGTAAGCGAGGCTCTAAAAGCACAGTTGAACTGAGTGGAAATGAGCTGGGTTGAGCCGAACTAGTATAAACTTGGCTAGGCTAC
 |||||
 TGCAGGAGGAAGACTGGAAGGGCTGGCTGAGCTAGACTAGGCTGGGCTGAGCTGGAATGAGCTGGGTTGAGCTGAACTAGTATAAACTTGGCTAGGCTAC

NWT #13 ($S\mu 203/S\alpha 409$)

CACAGTAATGACCCAGACAGAGAAAGCCAGACTCATAAAGCTTGCTGAGCAAAATTAAGGGAACAAGGTTGAGAGCCCTAGTAAGCGAGGCTCTAAAAG
 |||||
 CACAGTAATGACCCAGACAGAGAAAGCCAGACTCATAAAGCTTGCTGAGCGCTGAGCTGAGCTGGAATGAGATGGAATAGGCTGGGCTGGCTGGTGTGAG
 |||||
 GAGGAGCTAGGCTGGAATAGGTTGGACTGGGCTGGTGGAGCTGAGTTAGGCTGAGCTGAGCTGGAATGAGATGGAATAGGCTGGGCTGGCTGGTGTGAG

NWT #15 ($S\mu 500/S\alpha 37$)

TAGGTTGAATAGAGCTAACTCTACTGCCTACACTGGACTGTCTGAGCTGAGATGAGCTGGGGTGGCTCAGCTATGCTACGCTGTGTGGGGTGGCT
 |||||
 TAGGTTGAATAGAGCTAAATCTACTGCCTACACTGGACTGTCTGAGCTAGACTTAGGGTGAATGGGCTGAACAAGGCTGAGCTTACCTAGACCGGGC
 |||||
 GAGCTGAACTAGTATAAACTTGCTAGGCTACAATGGATTGAGCTGAGCTAGACTTAGGGTGAATGGGCTGAACAAGGCTGAGCTTACCTAGACCGGGC

Figure 29. Compendium of $S\mu$ - $S\alpha$ junction sequences. (Continued)

Appendix C

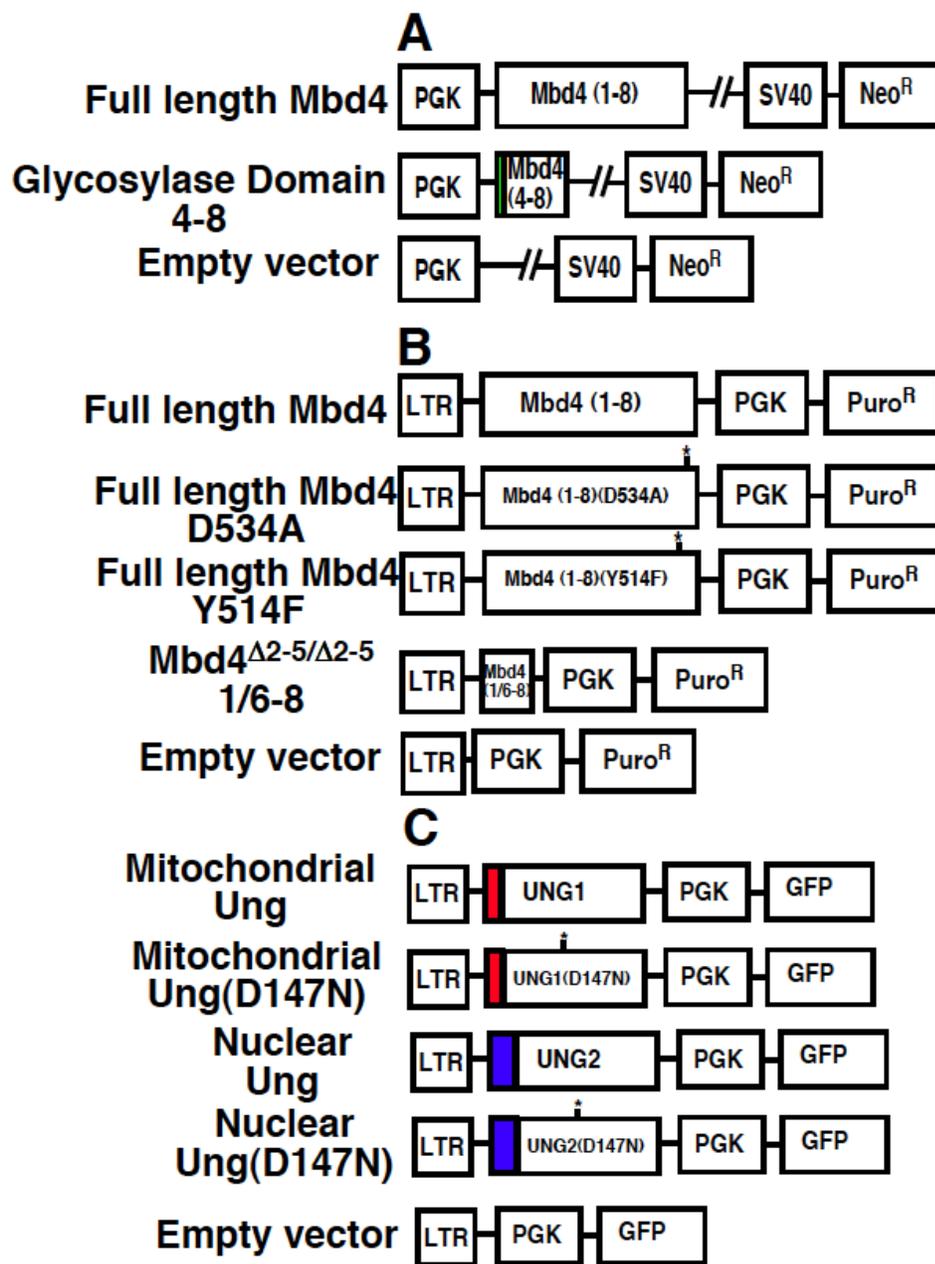


Figure 30. **Plasmid constructs used for transfection or transduction experiments.** **A)** pCDNA based Mbd4 rescue constructs. The green block indicates that this construct has a synthetic Kosak **B)** pMSCV based Mbd4 rescue constructs. **C)** pMSCV based Ung rescue constructs. Mitochondrial (red block) and nuclear (blue block) import sequences are depicted for UNG1 and UNG2, respectively.

Appendix D

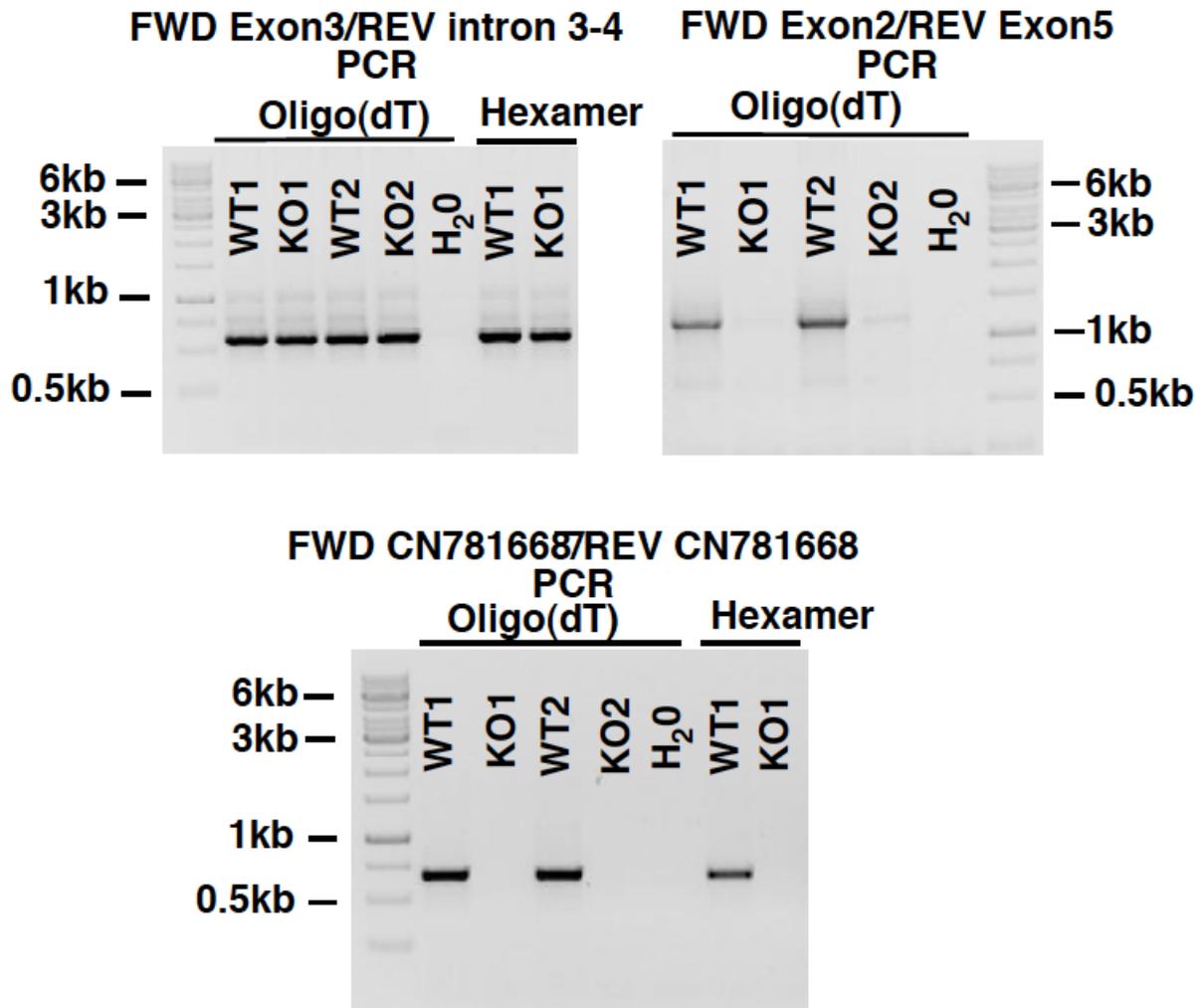


Figure 31. **PCR of Mbd4 and CN781668 transcripts.** First strand synthesis used both Oligo (dT) and random hexamer using cDNA from RNA isolated from WT or KO Mbd4 CH12.F3 activated with CIT for 48 hours. Two independent cDNAs from WT and KO were used for each PCR. **A)** PCR amplifies transcript from fig. 26G. This transcript is not affected by the Mbd4 Δ 6-8 deletion and uses a cryptic polyadenylation sequence. **B)** PCR amplifies full length Mbd4 (1-8) from WT and a residual Mbd4 transcript in the KO. **C)** PCR amplifies antisense CH781668 transcript .

REFERENCES

1. Hardy RR & Hayakawa K (2001) B cell development pathways. *Annu Rev Immunol* 19:595-621.
2. Tonegawa S (1983) Somatic Generation of Antibody Diversity. *Nature* 302(5909):575-581.
3. Difilippantonio MJ, McMahan CJ, Eastman QM, Spanopoulou E, & Schatz DG (1996) RAG1 mediates signal sequence recognition and recruitment of RAG2 in V(D)J recombination. *Cell* 87(2):253-262.
4. Schatz DG, Oettinger MA, & Baltimore D (1989) The V(D)J Recombination Activating Gene, Rag-1. *Cell* 59(6):1035-1048.
5. Oettinger MA, Schatz DG, Gorka C, & Baltimore D (1990) Rag-1 and Rag-2, Adjacent Genes That Synergistically Activate V(D)J Recombination. *Science* 248(4962):1517-1523.
6. Gorman JR & Alt FW (1998) Regulation of immunoglobulin light chain isotype expression. *Advances in Immunology, Vol 69* 69:113-181.
7. Li LQ & Boussiotis VA (2006) Physiologic regulation of central and peripheral T cell tolerance: lessons for therapeutic applications. *J Mol Med-Jmm* 84(11):887-899.
8. Viau M & Zouali M (2005) B-lymphocytes, innate immunity, and autoimmunity. *Clin Immunol* 114(1):17-26.
9. Purugganan MM, Shah S, Kearney JF, & Roth DB (2001) Ku80 is required for addition of N nucleotides to V(D)J recombination junctions by terminal deoxynucleotidyl transferase. *Nucleic Acids Res* 29(7):1638-1646.
10. Repasky JAE, Corbett E, Boboila C, & Schatz DG (2004) Mutational analysis of terminal deoxynucleotidyltransferase-mediated N-nucleotide addition in V(D)J recombination. *J Immunol* 172(9):5478-5488.
11. Kothapalli N & Fugmann SD (2011) Characterizing somatic hypermutation and gene conversion in the chicken DT40 cell system. *Methods Mol Biol* 748:255-271.
12. de Yebenes VG & Ramiro AR (2006) Activation-induced deaminase: light and dark sides. *Trends Mol Med* 12(9):432-439.
13. Tarlinton D (1998) Germinal centers: form and function. *Curr Opin Immunol* 10(3):245-251.
14. Peled JU, *et al.* (2008) The biochemistry of somatic hypermutation. *Annu Rev Immunol* 26:481-511.
15. Lebecque SG & Gearhart PJ (1990) Boundaries of Somatic Mutation in Rearranged Immunoglobulin Genes - 5'-Boundary Is near the Promoter, and 3'-Boundary Is Approximately-1 Kb from V(D)J Gene. *Journal of Experimental Medicine* 172(6):1717-1727.
16. Xue K, Rada C, & Neuberger MS (2006) The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in msh2^{-/-} ung^{-/-} mice. *J Exp Med* 203(9):2085-2094.
17. Rajewsky K, Forster I, & Cumano A (1987) Evolutionary and Somatic Selection of the Antibody Repertoire in the Mouse. *Science* 238(4830):1088-1094.
18. Di Noia JM & Neuberger MS (2007) Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem* 76:1-22.

19. Milstein C, Neuberger MS, & Staden R (1998) Both DNA strands of antibody genes are hypermutation targets. *Proc Natl Acad Sci U S A* 95(15):8791-8794.
20. Johnston CM, Wood AL, Bolland DJ, & Corcoran AE (2006) Complete sequence assembly and characterization of the C57BL/6 mouse Ig heavy chain V region. *J Immunol* 176(7):4221-4234.
21. Pettersson S, Cook GP, Bruggemann M, Williams GT, & Neuberger MS (1990) A second B cell-specific enhancer 3' of the immunoglobulin heavy-chain locus. *Nature* 344(6262):165-168.
22. Dunnick W, Hertz GZ, Scappino L, & Gritzmacher C (1993) DNA-Sequences at Immunoglobulin Switch Region Recombination Sites (Vol 21, Pg 365, 1993). *Nucleic Acids Res* 21(9):2285-2285.
23. Stavnezer J (2000) Immunology. A touch of antibody class. *Science* 288(5468):984-985.
24. Chaudhuri J & Alt FW (2004) Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat Rev Immunol* 4(7):541-552.
25. Stavnezer J (2000) Molecular processes that regulate class switching. *Curr Top Microbiol* 245:127-168.
26. Hein K, *et al.* (1998) Processing of switch transcripts is required for targeting of antibody class switch recombination. *Journal of Experimental Medicine* 188(12):2369-2374.
27. Storb U, *et al.* (1998) Somatic hypermutation of immunoglobulin genes is linked to transcription. *Somatic Diversification of Immune Responses* 229:11-19.
28. Muramatsu M, *et al.* (1999) Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J Biol Chem* 274(26):18470-18476.
29. Muramatsu M, *et al.* (2000) Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 102(5):553-563.
30. Revy P, *et al.* (2000) Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* 102(5):565-575.
31. Maul RW, *et al.* (2011) Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions. *Nat Immunol* 12(1):70-76.
32. Rada C, *et al.* (2002) Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr Biol* 12(20):1748-1755.
33. Ito S, *et al.* (2004) Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc Natl Acad Sci U S A* 101(7):1975-1980.
34. Begum NA, *et al.* (2004) Uracil DNA glycosylase activity is dispensable for immunoglobulin class switch. *Science* 305(5687):1160-1163.
35. Begum NA, *et al.* (2009) Further evidence for involvement of a noncanonical function of uracil DNA glycosylase in class switch recombination. *Proc Natl Acad Sci U S A* 106(8):2752-2757.
36. Conticello SG, Thomas CJ, Petersen-Mahrt SK, & Neuberger MS (2005) Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol* 22(2):367-377.

37. Severi F (2011) Analysis of Reptilian APOBEC suggests that RNA editing may not be its ancestral function. *Mol Biol Evol* 28(3):1125-1129.
38. Conticello SG, Langlois MA, Yang Z, & Neuberger MS (2007) DNA deamination in immunity: AID in the context of its APOBEC relatives. *Adv Immunol* 94:37-73.
39. Petersen-Mahrt S (2005) DNA deamination in immunity. *Immunol Rev* 203:80-97.
40. Kobayashi M, *et al.* (2009) AID-induced decrease in topoisomerase 1 induces DNA structural alteration and DNA cleavage for class switch recombination. *Proc Natl Acad Sci U S A* 106(52):22375-22380.
41. Kobayashi M, *et al.* (2011) Decrease in topoisomerase I is responsible for activation-induced cytidine deaminase (AID)-dependent somatic hypermutation. *Proc Natl Acad Sci U S A* 108(48):19305-19310.
42. Dickerson SK, Market E, Besmer E, & Papavasiliou FN (2003) AID mediates hypermutation by deaminating single stranded DNA. *J Exp Med* 197(10):1291-1296.
43. Chaudhuri J, *et al.* (2003) Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* 422(6933):726-730.
44. Stavnezer J (1996) Immunoglobulin class switching. *Curr Opin Immunol* 8(2):199-205.
45. Poltoratsky VP, Wilson SH, Kunkel TA, & Pavlov YI (2004) Recombinogenic phenotype of human activation-induced cytosine deaminase. *J Immunol* 172(7):4308-4313.
46. Petersen-Mahrt SK, Harris RS, & Neuberger MS (2002) AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature* 418(6893):99-103.
47. Mayorov VI, *et al.* (2005) Expression of human AID in yeast induces mutations in context similar to the context of somatic hypermutation at G-C pairs in immunoglobulin genes. *BMC Immunol* 6:10.
48. Roy D, Yu KF, & Lieber MR (2008) Mechanism of R-loop formation at immunoglobulin class switch sequences. *Mol Cell Biol* 28(1):50-60.
49. Huang FT, *et al.* (2007) Sequence dependence of chromosomal R-loops at the immunoglobulin heavy-chain S mu class switch region. *Mol Cell Biol* 27(16):5921-5932.
50. Zarrin AA, *et al.* (2004) An evolutionarily conserved target motif for immunoglobulin class-switch recombination. *Nat Immunol* 5(12):1275-1281.
51. Basu U, *et al.* (2005) The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature* 438(7067):508-511.
52. Perlot T, Li G, & Alt FW (2008) Antisense transcripts from immunoglobulin heavy-chain locus V(D)J and switch regions. *Proc Natl Acad Sci U S A* 105(10):3843-3848.
53. Shen HM, Ratnam S, & Storb U (2005) Targeting of the activation-induced cytosine deaminase is strongly influenced by the sequence and structure of the targeted DNA. *Mol Cell Biol* 25(24):10815-10821.
54. Li ZQ, Woo CJ, Iglesias-Ussel MD, Ronai D, & Scharff MD (2004) The generation of antibody diversity through somatic hypermutation and class switch recombination. *Gene Dev* 18(1):1-11.
55. Schrader CE, Linehan EK, Mochegova SN, Woodland RT, & Stavnezer J (2005) Inducible DNA breaks in Ig S regions are dependent on AID and UNG. *Journal of Experimental Medicine* 202(4):561-568.

56. Schrader CE, Guikema JE, Wu X, & Stavnezer J (2009) The roles of APE1, APE2, DNA polymerase beta and mismatch repair in creating S region DNA breaks during antibody class switch. *Philos Trans R Soc Lond B Biol Sci* 364(1517):645-652.
57. Barski A, *et al.* (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129(4):823-837.
58. Bernstein BE, *et al.* (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 120(2):169-181.
59. Wang L, Wuerffel R, Feldman S, Khamlichi AA, & Kenter AL (2009) S region sequence, RNA polymerase II, and histone modifications create chromatin accessibility during class switch recombination. *J Exp Med* 206(8):1817-1830.
60. Wang LL, Whang N, Wuerffel R, & Kenter AL (2006) AID-dependent histone acetylation is detected in immunoglobulin S regions. *Journal of Experimental Medicine* 203(1):215-226.
61. Pavri R, *et al.* (2010) Activation-Induced Cytidine Deaminase Targets DNA at Sites of RNA Polymerase II Stalling by Interaction with Spt5. *Cell* 143(1):122-133.
62. Okazaki IM, *et al.* (2003) Constitutive expression of AID leads to tumorigenesis. *J Exp Med* 197(9):1173-1181.
63. Jiang Y, Soong TD, Wang L, Melnick AM, & Elemento O (2012) Genome-wide detection of genes targeted by non-Ig somatic hypermutation in lymphoma. *PLoS One* 7(7):e40332.
64. Dedeoglu F, Horwitz B, Chaudhuri J, Alt FW, & Geha RS (2004) Induction of activation-induced cytidine deaminase gene expression by IL-4 and CD40 ligation is dependent on STAT6 and NF kappa B. *Int Immunol* 16(3):395-404.
65. Gonda H, *et al.* (2003) The balance between Pax5 and Id2 activities is the key to AID gene expression. *Journal of Experimental Medicine* 198(9):1427-1437.
66. Hauser J, *et al.* (2008) B-cell receptor activation inhibits AID expression through calmodulin inhibition of E-proteins. *Proc Natl Acad Sci U S A* 105(4):1267-1272.
67. Shinkura R, *et al.* (2004) Separate domains of AID are required for somatic hypermutation and class-switch recombination. *Nat Immunol* 5(7):707-712.
68. Ellyard JI, Benk AS, Taylor B, Rada C, & Neuberger MS (2011) The dependence of Ig class-switching on the nuclear export sequence of AID likely reflects interaction with factors additional to Crm1 exportin. *Eur J Immunol* 41(2):485-490.
69. Geisberger R, Rada C, & Neuberger MS (2009) 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 U S A* 106(16):6736-6741.
70. McBride KM, Barreto V, Ramiro AR, Stavropoulos P, & Nussenzweig MC (2004) Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. *J Exp Med* 199(9):1235-1244.
71. Cheng HL, *et al.* (2009) Integrity of the AID serine-38 phosphorylation site is critical for class switch recombination and somatic hypermutation in mice. *Proc Natl Acad Sci U S A* 106(8):2717-2722.
72. Chaudhuri J, Khuong C, & Alt FW (2004) Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature* 430(7003):992-998.
73. Teng G, *et al.* (2008) MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. *Immunity* 28(5):621-629.

74. Dorsett Y, *et al.* (2008) MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation. *Immunity* 28(5):630-638.
75. Ramiro AR, *et al.* (2004) AID is required for c-myc/IgH chromosome translocations in vivo. *Cell* 118(4):431-438.
76. Takizawa M, *et al.* (2008) AID expression levels determine the extent of cMyc oncogenic translocations and the incidence of B cell tumor development. *J Exp Med* 205(9):1949-1957.
77. Kotani A, *et al.* (2007) Activation-induced cytidine deaminase (AID) promotes B cell lymphomagenesis in Emu-cmyc transgenic mice. *Proc Natl Acad Sci U S A* 104(5):1616-1620.
78. Uchimura Y, Barton LF, Rada C, & Neuberger MS (2011) REG-gamma associates with and modulates the abundance of nuclear activation-induced deaminase. *Journal of Experimental Medicine* 208(12):2385-2391.
79. Wilson SH, *et al.* (2010) Base excision repair and design of small molecule inhibitors of human DNA polymerase beta. *Cell Mol Life Sci* 67(21):3633-3647.
80. Otterlei M, *et al.* (1998) Nuclear and mitochondrial splice forms of human uracil-DNA glycosylase contain a complex nuclear localisation signal and a strong classical mitochondrial localisation signal, respectively. *Nucleic Acids Res* 26(20):4611-4617.
81. Imai K, *et al.* (2003) Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nat Immunol* 4(10):1023-1028.
82. Kavli B, *et al.* (2005) B cells from hyper-IgM patients carrying UNG mutations lack ability to remove uracil from ssDNA and have elevated genomic uracil. *J Exp Med* 201(12):2011-2021.
83. Di Noia JM, *et al.* (2007) Dependence of antibody gene diversification on uracil excision. *J Exp Med* 204(13):3209-3219.
84. Guikema JE, *et al.* (2007) APE1- and APE2-dependent DNA breaks in immunoglobulin class switch recombination. *J Exp Med* 204(12):3017-3026.
85. Sabouri Z, *et al.* (2009) Apex2 is required for efficient somatic hypermutation but not for class switch recombination of immunoglobulin genes. *Int Immunol* 21(8):947-955.
86. Wu X & Stavnezer J (2007) DNA polymerase beta is able to repair breaks in switch regions and plays an inhibitory role during immunoglobulin class switch recombination. *J Exp Med* 204(7):1677-1689.
87. Lahue RS, Au KG, & Modrich P (1989) DNA Mismatch Correction in a Defined System. *Science* 245(4914):160-164.
88. Fukui K (2010) DNA mismatch repair in eukaryotes and bacteria. *J Nucleic Acids* 2010.
89. Modrich P (2006) Mechanisms in eukaryotic mismatch repair. *J Biol Chem* 281(41):30305-30309.
90. Kadyrov FA, Dzantiev L, Constantin N, & Modrich P (2006) Endonucleolytic function of MutLalpha in human mismatch repair. *Cell* 126(2):297-308.
91. Kadyrov FA, *et al.* (2007) *Saccharomyces cerevisiae* MutL alpha is a mismatch repair endonuclease. *Journal of Biological Chemistry* 282(51):37181-37190.

92. Chahwan R, *et al.* (2012) The ATPase activity of MLH1 is required to orchestrate DNA double-strand breaks and end processing during class switch recombination. *Journal of Experimental Medicine* 209(4):671-678.
93. Rada C, Di Noia JM, & Neuberger MS (2004) Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol Cell* 16(2):163-171.
94. Schrader CE, Vardo J, & Stavnezer J (2002) Role for mismatch repair proteins Msh2, Mlh1, and Pms2 in immunoglobulin class switching shown by sequence analysis of recombination junctions. *J Exp Med* 195(3):367-373.
95. Schrader CE, Vardo J, & Stavnezer J (2003) Mlh1 can function in antibody class switch recombination independently of Msh2. *J Exp Med* 197(10):1377-1383.
96. Schrader CE, Edelmann W, Kucherlapati R, & Stavnezer J (1999) Reduced isotype switching in splenic B cells from mice deficient in mismatch repair enzymes. *Journal of Experimental Medicine* 190(3):323-330.
97. Stavnezer J, Bjorkman A, Du L, Cagigi A, & Pan-Hammarstrom Q (2010) Mapping of switch recombination junctions, a tool for studying DNA repair pathways during immunoglobulin class switching. *Adv Immunol* 108:45-109.
98. Eccleston J, Schrader CE, Yuan K, Stavnezer J, & Selsing E (2009) 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* 183(2):1222-1228.
99. Saribasak H, Rajagopal D, Maul RW, & Gearhart PJ (2009) Hijacked DNA repair proteins and unchained DNA polymerases. *Philos Trans R Soc Lond B Biol Sci* 364(1517):605-611.
100. Wilson TM, *et al.* (2005) MSH2-MSH6 stimulates DNA polymerase eta, suggesting a role for A : T mutations in antibody genes. *Journal of Experimental Medicine* 201(4):637-645.
101. Zeng XM, *et al.* (2001) DNA polymerase eta is an A-T mutator in somatic hypermutation of immunoglobulin variable genes. *Nat Immunol* 2(6):537-541.
102. Faili A, *et al.* (2004) DNA polymerase eta is involved in hypermutation occurring during immunoglobulin class switch recombination. *Journal of Experimental Medicine* 199(2):265-270.
103. Mohd AB, *et al.* (2006) Truncation of the C-terminus of human MLH1 blocks intracellular stabilization of PMS2 and disrupts DNA mismatch repair. *DNA Repair (Amst)* 5(3):347-361.
104. van Oers JM, *et al.* (2010) PMS2 endonuclease activity has distinct biological functions and is essential for genome maintenance. *Proc Natl Acad Sci U S A* 107(30):13384-13389.
105. Mukherjee S, Ridgeway AD, & Lamb DJ (2010) DNA mismatch repair and infertility. *Curr Opin Urol* 20(6):525-532.
106. Schrader CE, Guikema JEJ, Linehan EK, Selsing E, & Stavnezer J (2007) Activation-induced cytidine deaminase-dependent DNA breaks in class switch recombination occur during G(1) phase of the cell cycle and depend upon mismatch repair. *J Immunol* 179(9):6064-6071.
107. Reynaud CA, *et al.* (2009) Competitive repair pathways in immunoglobulin gene hypermutation. *Philos Trans R Soc Lond B Biol Sci* 364(1517):613-619.

108. Stavnezer J & Schrader CE (2006) Mismatch repair converts AID-instigated nicks to double-strand breaks for antibody class-switch recombination. *Trends in Genetics* 22(1):23-28.
109. Stewart GS, *et al.* (2007) RIDDLE immunodeficiency syndrome is linked to defects in 53BP1-mediated DNA damage signaling. *Proc Natl Acad Sci U S A* 104(43):16910-16915.
110. Paull TT, *et al.* (2000) A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Curr Biol* 10(15):886-895.
111. Lee JH & Paull TT (2004) Direct activation of the ATM protein kinase by the Mre11/Rad50/Nbs1 complex. *Science* 304(5667):93-96.
112. Carney JP, *et al.* (1998) The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: Linkage of double-strand break repair to the cellular DNA damage response. *Cell* 93(3):477-486.
113. Varon R, *et al.* (1998) Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. *Cell* 93(3):467-476.
114. Petersen S, *et al.* (2001) AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. *Nature* 414(6864):660-665.
115. Yan CT, *et al.* (2007) IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature* 449(7161):478-U479.
116. Han L & Yu KF (2008) Altered kinetics of nonhomologous end joining and class switch recombination in ligase IV-deficient B cells (vol 205, pg 2745, 2008). *Journal of Experimental Medicine* 205(13):3215-3215.
117. Han L, Mao WF, & Yu KF (2012) X-ray repair cross-complementing protein 1 (XRCC1) deficiency enhances class switch recombination and is permissive for alternative end joining. *Proc Natl Acad Sci U S A* 109(12):4604-4608.
118. Boboila C, *et al.* (2012) Robust chromosomal DNA repair via alternative end-joining in the absence of X-ray repair cross-complementing protein 1 (XRCC1). *Proc Natl Acad Sci U S A* 109(7):2473-2478.
119. Casellas R, *et al.* (1998) Ku80 is required for immunoglobulin isotype switching. *Embo J* 17(8):2404-2411.
120. Manis JP, *et al.* (1998) Ku70 is required for late B cell development and immunoglobulin heavy chain class switching. *Journal of Experimental Medicine* 187(12):2081-2089.
121. Reina-San-Martin B, *et al.* (2003) H2AX is required for recombination between immunoglobulin switch regions but not for intra-switch region recombination or somatic hypermutation. *Journal of Experimental Medicine* 197(12):1767-1778.
122. Manis JP, Dudley D, Kaylor L, & Alt FW (2002) IgH class switch recombination to IgG1 in DNA-PKcs-deficient B cells. *Immunity* 16(4):607-617.
123. Callen E, *et al.* (2009) Essential Role for DNA-PKcs in DNA Double-Strand Break Repair and Apoptosis in ATM-Deficient Lymphocytes. *Mol Cell* 34(3):285-297.
124. Bosma GC, *et al.* (2002) DNA-dependent protein kinase activity is not required for immunoglobulin class switching. *Journal of Experimental Medicine* 196(11):1483-1495.
125. Moshous D, *et al.* (2001) Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell* 105(2):177-186.

126. Rooney S, Alt FW, Sekiguchi JA, & Manis JP (2005) Artemis-independent functions of DNA-dependent protein kinase in Ig heavy chain class switch recombination and development. *Proc Natl Acad Sci U S A* 102(7):2471-2475.
127. Buck D, *et al.* (2006) Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell* 124(2):287-299.
128. Li G, *et al.* (2008) Lymphocyte-specific compensation for XLF/Cernunnos end-joining functions in V(D)J recombination. *Mol Cell* 31(5):631-640.
129. Eccleston J, Yan C, Yuan K, Alt FW, & Selsing E (2011) Mismatch Repair Proteins MSH2, MLH1, and EXO1 Are Important for Class-Switch Recombination Events Occurring in B Cells That Lack Nonhomologous End Joining. *J Immunol* 186(4):2336-2343.
130. Ramiro A, *et al.* (2007) The role of activation-induced deaminase in antibody diversification and chromosome translocations. *Adv Immunol* 94:75-107.
131. Reina-San-Martin B, Chen JJ, Nussenzweig A, & Nussenzweig MC (2007) Enhanced intra-switch region recombination during immunoglobulin class switch recombination in 53BP(-/-) B cells. *Eur J Immunol* 37(1):235-239.
132. Bothmer A, *et al.* (2010) 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. *Journal of Experimental Medicine* 207(4):855-865.
133. Rai K, *et al.* (2008) DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* 135(7):1201-1212.
134. Bhutani N, *et al.* (2010) Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* 463(7284):1042-1047.
135. Popp C, *et al.* (2010) Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* 463(7284):1101-1105.
136. Bellacosa A, *et al.* (1999) MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. *Proc Natl Acad Sci U S A* 96(7):3969-3974.
137. Wu PY, *et al.* (2003) Mismatch repair in methylated DNA - Structure and activity of the mismatch-specific thymine glycosylase domain of methyl-CpG-binding protein MBD4. *Journal of Biological Chemistry* 278(7):5285-5291.
138. Hendrich B, Hardeland U, Ng HH, Jiricny J, & Bird A (1999) The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* 401(6750):301-304.
139. Yang HJ, *et al.* (2000) Characterization of a thermostable DNA glycosylase specific for U/G and T/G mismatches from the hyperthermophilic archaeon *Pyrobaculum aerophilum*. *Journal of Bacteriology* 182(5):1272-1279.
140. Hashimoto H, Zhang X, & Cheng X (2012) Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: structural basis and implications for active DNA demethylation. *Nucleic Acids Res.*
141. Petronzelli F, *et al.* (2000) Investigation of the substrate spectrum of the human mismatch-specific DNA N-glycosylase MED1 (MBD4): Fundamental role of the catalytic domain. *Journal of Cellular Physiology* 185(3):473-480.
142. Herman JG, *et al.* (1998) Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc Natl Acad Sci U S A* 95(12):6870-6875.

143. Liu B, *et al.* (1995) Mismatch repair gene defects in sporadic colorectal cancers with microsatellite instability. *Nat Genet* 9(1):48-55.
144. Bader S, *et al.* (1999) Somatic frameshift mutations in the MBD4 gene of sporadic colon cancers with mismatch repair deficiency. *Oncogene* 18(56):8044-8047.
145. Millar CB, *et al.* (2002) Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science* 297(5580):403-405.
146. Wong E, *et al.* (2002) Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation. *Proc Natl Acad Sci U S A* 99(23):14937-14942.
147. Reitmair AH, *et al.* (1996) Spontaneous intestinal carcinomas and skin neoplasms in Msh2-deficient mice. *Cancer Res* 56(16):3842-3849.
148. Prolla TA, *et al.* (1998) Tumour susceptibility and spontaneous mutation in mice deficient in Mlh1, Pms1 and Pms2 DNA mismatch repair. *Nat Genet* 18(3):276-279.
149. Kawate H, *et al.* (2000) A defect in a single allele of the Mlh1 gene causes dissociation of the killing and tumorigenic actions of an alkylating carcinogen in methyltransferase-deficient mice. *Carcinogenesis* 21(2):301-305.
150. Cortellino S, *et al.* (2003) The base excision repair enzyme MED1 mediates DNA damage response to antitumor drugs and is associated with mismatch repair system integrity. *Proc Natl Acad Sci U S A* 100(25):15071-15076.
151. Rhinn H, *et al.* (2008) Housekeeping while brain's storming Validation of normalizing factors for gene expression studies in a murine model of traumatic brain injury. *BMC Mol Biol* 9:62.
152. Wuerrffel R, *et al.* (2007) S-S synapsis during class switch recombination is promoted by distantly located transcriptional elements and activation-induced deaminase. *Immunity* 27(5):711-722.
153. Lin SC & Stavnezer J (1996) Activation of NF-kappaB/Rel by CD40 engagement induces the mouse germ line immunoglobulin Cgamma1 promoter. *Mol Cell Biol* 16(9):4591-4603.
154. Min IM, Rothlein LR, Schrader CE, Stavnezer J, & Selsing E (2005) Shifts in targeting of class switch recombination sites in mice that lack mu switch region tandem repeats or Msh2. *J Exp Med* 201(12):1885-1890.
155. Min IM, *et al.* (2003) The Smu tandem repeat region is critical for Ig isotype switching in the absence of Msh2. *Immunity* 19(4):515-524.
156. Frieder D, Larijani M, Collins C, Shulman M, & Martin A (2009) The Concerted Action of Msh2 and UNG Stimulates Somatic Hypermutation at A . T Base Pairs. *Mol Cell Biol* 29(18):5148-5157.
157. Han L, Masani S, & Yu KF (2010) Cutting Edge: CTNNB1 Is Dispensable for Ig Class Switch Recombination. *J Immunol* 185(3):1379-1381.
158. Han L & Yu K (2008) Altered kinetics of nonhomologous end joining and class switch recombination in ligase IV-deficient B cells. *J Exp Med* 205(12):2745-2753.
159. Nagaoka H, Ito S, Muramatsu M, Nakata M, & Honjo T (2005) DNA cleavage in immunoglobulin somatic hypermutation depends on de novo protein synthesis but not on uracil DNA glycosylase. *Proc Natl Acad Sci U S A* 102(6):2022-2027.
160. Hung T & Chang HY (2010) Long noncoding RNA in genome regulation: prospects and mechanisms. *RNA Biol* 7(5):582-585.

161. Kogo R, *et al.* (2011) Long Noncoding RNA HOTAIR Regulates Polycomb-Dependent Chromatin Modification and Is Associated with Poor Prognosis in Colorectal Cancers. *Cancer Res* 71(20):6320-6326.
162. Bardwell PD, *et al.* (2003) Cutting edge: The G-U mismatch glycosylase methyl-CpG binding domain 4 is dispensable for somatic hypermutation and class switch recombination. *J Immunol* 170(4):1620-1624.

BIOGRAPHICAL SKETCH

Provide the following information for the key personnel and other significant contributors in the order listed on Form Page 2.
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME Fernando Grigera		POSITION TITLE Ph.D.	
eRA COMMONS USER NAME			
EDUCATION/TRAINING <i>(Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.)</i>			
INSTITUTION AND LOCATION	DEGREE <i>(if applicable)</i>	YEAR(s)	FIELD OF STUDY
Oberlin College, Oberlin, OH	B.A.	4	Biology
University of Illinois, Chicago	Ph.D. candidate	6	Microbiology & Immunology

A. Positions and Honors.**Professional Experience:**

- 2009-Present Bridging the Gap Tutor, NSF program, Graduate Education in Medical Sciences, University of Illinois Chicago
- 2002-2006 Research Associate, Department of Microbiology and Immunology, University of Illinois Chicago
- 2001-2002 Research Assistant, Department of Biology, Oberlin College

Honors:

- 2009-Present NIH Diversity Supplement Award, Department of Microbiology and Immunology, University of Illinois Chicago
- 2001 Independent Research in Biology
Department of Biology, Oberlin College
- 1999 John Friedman Scholarship for Academic Excellence

B. Peer-reviewed publications.

- Wuerffel, R.; Wang, L.; **Grigera, F.**; Manis, J.; Selsing, E.; Perlot, T., Alt, F.W., Cogne, M., Pinaud, E., Kenter, A.L. (2007). S-S synopsis during class switch recombination is promoted by distantly located transcriptional elements and activation-induced deaminase. *Immunity* **27**:711-722.
- Battacharya, P.; **Grigera, F.**, Rogozin, I.B., McCarty, T.; Morse, H.C. 3rd, Kenter, A.L. (2008). Identification of murine B cell lines that undergo somatic hypermutation focused to A:T and G:C residues. *Eur. J. Immunol.* **38**:227-239.

C. Research Support

None

Curriculum Vitae

Fernando Grigera

531 West Deming Place #501
 Chicago, IL 60614
 (216) 978-3626 (cell)
 (312) 996-0946 (work)
 Email: fgrigera@gmail.com

Education

2003-Present

UNIVERSITY OF ILLINOIS Chicago, IL
 Graduate Student
 Finish: 2013; Overall GPA: 3.70

RELEVANT
 COURSES
 Biology,

Cell Biology, Biochemistry, Molecular Genetics, Molecular

Writing,

Integrative Biology, Advanced Immunology, Virology, Grant

Research Techniques Modules

UNIVERSITY OF ILLINOIS Chicago, IL
 Part time Undergraduate Student
 Finish: July 2006; Overall GPA: 4.0

RELEVANT
 COURSES

Biochemistry I, Genetics, Organic Chemistry II, Physics

1998-2002

OBERLIN COLLEGE Oberlin, OH
 Bachelor of Arts in Biology
 Graduation Date: May 2002; Overall GPA: 3.10

RELEVANT
 COURSES

Bioengineering, Virology, Organic Chemistry I (lab),
 Microbiology (lab), Inorganic Chemistry I and II (lab),
 Animal Physiology (lab), Vertebrate Structure (lab),
 Botany (lab), Physics I and II (lab), Calculus

1994-1998

CHAGRIN FALLS HIGH SCHOOL Chagrin Falls, OH
 Graduation Date: June 1998; overall GPA: 3.8

Awards and Distinctions

- 2001 McNair Scholarship for Independent Research in Immunology
 Summer project: Research Assistant
 Department of Biology
 Oberlin College Science Center
- 2004 Dean's List: University of Illinois
- 1998-2002 Dean's List: Two semesters at Oberlin College
- 1999 John Freedman Scholarship for Academic Excellence
 \$7,000 per year