Mechanisms of transversion mutation are dependent on sequence context and nucleotide paucity during antibody somatic hypermutation

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Eddy Sanchai Thientosapol BSc (Medicine) MB BS MMED FRACP
Centenary Institute of Cancer Medicine and Cell Biology
The University of Sydney
January 2019

Table of Contents

Declaration

The work described in this thesis was performed by the author in the DNA Repair Laboratory at the Centenary Institute, Camperdown, NSW, between March 2013 and June 2018 as part of the requirements of the degree of Doctor of Philosophy. This work is original and has not been previously submitted for the purpose of obtaining any other degree.

Eddy Sanchai Thientosapol

BSc (Medicine) MB BS FRACP

January 2019

Acknowledgements

I would especially like to thank Chris Jolly, for without Chris, this PhD would not have been possible.

I would also like to thank the other members of the DNA Repair Laboratory including Adam Cook, Daniel Bosnjak, Tim Durack and Igor Stevanovski for all the assistance provided to me. I would like to thank the members of the Sydney Cytometry Facility including Frank Kao, Suat Dervish, Steven Allen, and Caroline Royle for flow cytometry and cell sorting assistance. I would also like to thank other members of the Centenary Institute including Christine Yee, Aimei Lee, Thomas Tu, Goldie Lui, Michelle Simmons and Julie Feng for all their help.

I would also like to thank my wonderful partner Alexandra Newton, who proofread this manuscript at short notice.

I would like to acknowledge my funding sources including a Chris O'Brien Lifehouse Fellowship, University of Sydney Australian Postgraduate Award, National Health and Medical Research Council Postgraduate Clinical Scholarship, Australian American Fulbright Commission Postgraduate Scholarship and Rockefeller University for additional research stipend.

Abbreviations

1-βTRCP	1-β-transducing-repeat-containing protein
53BP1	p53-binding protein 1
AEJ	alternative end-joining
AGS	Aicardi-Goutieres syndrome
AID	activation induced cytidine deaminase
AIM2	absent in melanoma-2
ANOVA	analysis of variance
APC	allophycocyanin
APE1	apurinic/apyrimidinic endonuclease-1
APOBEC	apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like
Arg	arginine
Asp	aspartic acid
ATM	ataxia telangiectasia-mutated kinase
ATR	ataxia telangiectasia-mutated and Rad3 related
Bcl-6	B-cell lymphoma 6
BER	base excision repair
BLAST	basic local alignment search tool
BSA	bovine serum albumin
BST2	bone marrow stromal cell antigen 2
CD4	cluster of differentiation 4
CD40	cluster of differentiation 40

CDK	cyclin dependent kinase
cDNA	complementary DNA
ChIP	chromatin Immunoprecipitation
CLL	chronic lymphocytic leukaemia
CRISPR/Cas9	clustered regulatory interspaced short palindromic repeats
CRL4	cullin-ring finger ligase-4
CSR	class switch recombination
CT	computer tomography
DAMP	damage-associated molecular patterns
DAPI	4',6-diamidino-2-phenylindole
DCAF1	Ddb1–cullin4-associated-factor 1
DCIP	dendritic cell-derived IFN- γ
DMEM	Dulbecco's Modified Eagle Medium
DNA	deoxyribonucleic acid
$DNAPK_{cs}$	DNA-dependent protein kinase, catalytic subunit
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
dGTP	deoxyguanosine triphosphate
dNDP	deoxynucleoside diphosphate
dNTP	deoxyribonucleotide triphosphate
dTTP	deoxythymidine triphosphate
DSB	double stranded break
EDCI	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide

EDTA	ethylenediaminetetraacetic acid
EXO1	exonuclease 1
FACS	flow assisted cytometric sorting
FAM	fluorescein
FCS	foetal calf serum
FITC	fluorescein isothiocyanate
GFP	green fluorescent protein
Gln	glutamine
GTP	guanosine-5'-triphosphate
HD	(His---His-Asp---Asp) domain
HEL	hen egg lysozyme
HMG1/2	high-mobility group protein 1/2
HIGM2	hyper-IgM syndrome 2
HIV-1	human immunodeficiency virus 1
Hox C4	homeobox protein Hox-C4
HR	homologous recombination
IBFQ	Iowa Black FQ quencher
IFN-α	interferon alpha
IL-4	Interleukin 4
IMDM	Iscove's Modified Dulbecco's Medium
IRES	internal ribosome entry site
LB	lysogeny broth
MES	2-(*N*-morpholino)ethanesulfonic acid

mKO	monomeric Kusabira Orange
MLH1	MutL homolog 1, colon cancer, nonpolyposis type 2
MMR	mismatch repair
MRN	Mre11-Nbs1-Rad50
MSH2	DNA mismatch repair protein Msh2
MSH6	DNA mismatch repair protein Msh6
NDP	ribonucleoside diphosphate
NF-κB	nuclear factor -κB
NHEJ	non-homologous end joining
NOD	nucleotide oligomerization domain
NP-40	nonyl phenoxypolyethoxylethanol 40
PAMP	pathogen-associated molecular patterns
PARP1	poly-ADP-ribose polymerase 1
Pax5	paired box protein Pax-5
PBA	PBS with 2% BSA
PBS	phosphate buffered saline
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
PMS2	mismatch repair endonuclease PMS2
Pol	polymerase
PRR	pattern recognition receptors
qRTPCR	quantitative reverse transcription polymerase chain reaction
R1	ribonucleotide redcutase subunit R1

R2	ribonucleotide redcutase subunit R2
RAG	recombination activating gene
Rev1	DNA repair protein REV1
RIG-1	retinoic-acid-inducible gene-1
RNA	ribonucleic acid
RNR	ribonucleotide reductase
RPA	replication protein A
RPMI	Roswell Park Memorial Institute medium
RRM2	ribonucleotide reductase subunit M2
RT	room temperature
SAM	sterile alpha motif
SAMHD1	sterile alpha motif domain and histidine-aspartic domain containg protein 1
SEC-MALLS	size exclusion chromatography via multi-angle light scattering
sgRNA	single guide ribonucleic acid
shRNA	single hairpin ribonucleic acid
siRNA	single inhibitory ribonucleic acid
SKP-cullin	Skp, Cullin, F-box containing complex
SLE	systemic lupus erythomatosisis
SRBC	sheep red blood cells
ssDNA	single stranded DNA
TALEN	transcription activator-like effector nucleases
Taq	*Thermus aquaticus*

TBST	Tris-Buffered Saline and Tween 20
TLR	toll-like recptor
TLS	Translesion DNA
TREX1	three prime repair exonuclease 1
TRIM5a	tripatite Motif 5 alpha
ugi	uracil Glycosylase Inhibitor
UNG	U N-glycosylase
UV	ultraviolet
VDJ	variable (V), diversity (D), joining (J)
WMI	Westmead Millenium Institute
XLF	XRCC4-like factor
XPV	xeroderma pigmentosum variant
XRCC1	X-ray repair cross-complementing protein 1
XRCC4	X-ray repair cross-complementing protein 4
ZFN	zinc finger nucleases

List of Tables:

List of Figures:

Abstract

Somatic hypermutation of antibodies during humoral immune responses depends on expression of Activation Induced Deaminase (AID) in antibody-producing B cells. AID initiates somatic hypermutation by converting cytosine (C) residues in antibody genes into uracil (U) residues, by deamination. Alone, conversion of cytosine into uracil can only produce C:G to T:A transition mutations, by replication across U (phase 1A mutation). Processing of C deaminations by base excision repair (BER) or mismatch repair (MMR) diversifies mutation, predominantly at C:G (phase 1B mutation) and A:T (phase 2 mutation), respectively.

Mutations at C along the *Ig* variable region are not equally distributed. AID deaminates C more often if they occur as part of WRCY motif (A/T,A/G,C,C/T). WRCY sequences are concentrated in hypervariable regions of *Ig* genes, where nucleotide substitutions are likely to be effective at generating useful amino acid substitutions to optimize affinity maturation. Of all WRCY motifs, AGCT and AACT are the most mutated hotspots. AGCT is also enriched in switch regions and facilitates CSR.

In Chapter three, using large datasets of a transgenic mouse model, I compared *Igh* hypermutation between SW_{HEL} B cells, SW_{HEL} B cells deficient for UNG2 *via* retroviral expression of the uracil glycosylase inhibitor (ugi), SW_{HEL} B cells deficient for MutSα by crossing $Msh2^{ko}$ alleles into SW_{HEL} mice and SW_{HEL} B cells deficient for both UNG2 and MutSα.

I found that phase 1B mutations occur by distinct MMR-independent or MMR dependent pathways. At or in proximity to AGCW motifs, phase 1B mutations were driven by UNG2 without requirement for mismatch repair. Deaminations in AGCW were refractive both to processing by UNG2 and to high-fidelity base excision repair (BER) downstream of UNG2, regardless of mismatch repair activity. Outside AGCW motifs, transversions at C:G are co-dependent on UNG2 and MMR.

Classically, MMR mediates high fidelity repair of mismatches introduced during replication. The reasons for the profound differences in repair accuracy between classical and AID-induced MMR have not been elicited. During S-phase of the cell replication cycle, when classical post-replication MMR occurs, nucleotide triphosphate (dNTP) levels are optimal for DNA replication, while in G1-phase dNTP levels are lower. Since there is evidence that AID is active in G1-phase, we hypothesized that low dNTP levels may be the cause of low fidelity MMR.

Two enzymes are the major determinant of dNTP pools: ribonucleotide reductase (RNR), which converts ribonucleotides into deoxyribonucleotides predominantly during S-phase, and SAMHD1, which degrades dNTPs predominantly outside of S-phase.

In Chapters four and five, I quantified antibody hypermutation in B cells lacking SAMHD1 and/or over-expressing RNR. I observed a 2-fold decrease in mutations at A:T bases in cells lacking SAMHD1. This decrease was comparable to the decrease induced by RNR over-expression and was consistent with our hypothesis.

Unexpectedly, loss of SAMHD1 also decreased transversion mutations at C:G by about 70%, and almost doubled transition mutations at C:G bases. RNR over-expression had no obvious impact on transversion mutations at C:G, but increased transition mutations at C:G bases similarly to loss of SAMHD1. Furthermore, loss of SAMHD1 decreased AID/BER-dependent antibody class switch recombination, while RNR over-expression did not.

These findings indicate that dNTPs play a role in MMR-mediated antibody mutation, as predicted by our hypothesis, but they also indicate a major role for SAMHD1 in AID-induced BER that was not predicted by our hypothesis or by current models of antibody hypermutation. This important finding warrants further investigation to identify the mechanism.

1. Introduction

In 1957, Frank Macfarlane Burnet published the "clonal selection hypothesis", a theory to account for the body's ability to produce antibodies to virtually all foreign antigens, with the absence of an immunological response to self (Burnet, 1957). He postulated that embryonic stage lymphocytes would "have a randomization of the coding responsible for part of the specification of gamma globulin molecules", and subsequently during antigen exposure, lymphocytes that bind antigen would proliferate. The clones with highest affinity to antigen would be selected and "inherited changes will occur as a result of somatic mutation". It would take another 19 years before molecular biology would identify the process of antigen receptor VDJ recombination and later somatic hypermutation of VDJ-rearranged immunoglobulin genes (Jung & Alt, 2004).

Overview of the immune system

It was Claude Bernard who, in 1865, pioneered the understanding of the human body as "simple phenomena, which could be associated together to form a common final purpose". It is these processes which allow for *la vie constant ou libre*: "where life is liberated from the external environment and succeeds in maintaining constant conditions within its own internal environment" (Cooper, 2008). It was later, in 1929, that Walter Cannon would coin the term homeostasis to refer to the "constancy of the internal environment after distortion by external stresses" (Cannon, 1929).

As such, the immune system can be conceptualised as a set of individual components to maintain homeostasis of the body in an environment of pathogenic microbes.

Components of the immune system

The immune system is complex. For simplicity, it can broadly be divided into three major components, which include constitutive innate immunity, reactive innate immunity and adaptive immunity.

As part of constitutive innate immunity, the skin and mucosal surfaces form physical barriers against infection. These surfaces prevent the passage of pathogens from the external environment into the internal tissues. The most superficial layer of the skin, the stratum corneum, mediates the barrier function of the epidermis (Baroni et al., 2012).

The strength of the physical barrier in the stratum corneum is provided by its cellular structure of flat, anucleated squamous cells with an internal structure of keratin filaments. These keratin filaments assemble into a web-like structure originating from a ring around the nucleus. This structure terminates at desmosomes, located on the cell membrane, which interconnect adjacent keratinocytes and provide intercellular cohesion. The keratinocyte is also surrounded by a cell envelope consisting of cross-linked proteins (Elias, 2007). The intercellular space is composed of lipids, which counteract the loss of water and salts from the skin and prevent the penetration of water soluble substances into the skin (Baroni et al., 2012).

Similarly, the cells forming the lining of the gastrointestinal tract, the intestinal epithelial cells, form a physical barrier. In contrast to the skin, however, the physical barrier is selectively permeable, allowing absorption of nutrients (Pitman & Blumberg, 2000). Intestinal epithelial cells maintain their structural integrity through a complex cytoskeletal network of microfilaments. These include actin filaments, which form rings at both the apical (luminal) and basolateral poles of the cell, and intermediate filaments which spread through the cytoplasm and, similarly to keratinocytes, terminate at the cell membrane. Tight junctions on the apical (luminal) surface separate the lumen of the gastrointestinal tract from the paracellular space.

Also contributing to physical barrier function is the secretion of mucins onto the luminal surface by goblet cells. Mucins, which are a group of glycoproteins, provide a hydrated viscous layer to prevent trauma to the epithelial surface. Bacteria also adhere to mucins, preventing bacterial contact to intestinal epithelial cells. Mucus is also propelled down the intestinal tract, facilitating removal of bound bacteria. Specifically, in the stomach, a low pH contributes to pathogen elimination (Pitman & Blumberg, 2000).

At these physical barriers, including skin and gastrointestinal tract, certain antimicrobial peptides are produced, including lysozyme and defensins. Lysozyme is a 1,4-*β*-N-acetylmuramidase that cleaves the glycosidic bond between the C-1 of N-acetylmuramic acid and the C-4 of N-acetylglucosamine. This bond occurs in peptidoglycan, which is an important component of the gram-positive bacterial cell

wall. Lysozyme in humans is most abundant in tears, gastric fluid and also breastmilk (Jolles & Jolles, 1984).

Defensins are antimicrobial peptides with a characteristic β-sheet-rich fold and a framework of six disulfide-linked cysteines (Ganz, 2003). Defensins permeabilise target bacterial, fungal and encapsulated viral membranes (McCormick & Weinberg, 2010). Defensins contain clusters of amino acids with positively charged side chains and clusters of amino acids with hydrophobic side chains. Due to this property, they interact with microbial membranes. When defensins congregate on microbial cell membranes they form pores, and this leads to permeabilisation. Although defensins are secreted by epithelial cells, they are in much higher concentration in granulocytes (Ganz, 2003). Cathlecidin, produced by the stratum corneum and gastrointestinal tract, also functions by interaction with microbial cell membranes, leading to permeabilisation (Kosciuczuk et al., 2012). Like defensins, cathlecidins are also present in higher concentration in granulocytes.

Other antimicrobial peptides include RNase 7, produced by the stratum corneum, which has potent ribonuclease activity (Harder & Schroder, 2002), and calprotectin, a manganese and zinc chelator, which deprives pathogens of essential metals. Calprotectin comprises as much as 60% of the soluble protein content of the cytosol of neutrophils (Striz & Trebichavsky, 2004). Other metal binding antimicrobial peptides include psoriasin, produced by the stratum corneum, that likely functions by sequestration of zinc (Glaser et al., 2005) and neutrophil gelatinase-associated lipocalin, which sequesters iron (Miethke & Skerra, 2010).

The reactive innate immune system

Should pathogens penetrate the physical barriers of the body, the reactive innate immune system and adaptive immune system detect the presence of invading organisms. One key difference between the innate and adaptive immune systems is that the innate immune system has a rather limited repertoire of receptors to detect invading pathogens. Innate immune system receptors are inherited on a germline basis. In contrast, the adaptive immune system has an extremely diverse, semi-randomly generated repertoire of receptors (Turvey & Broide, 2010).

The innate immune system has three different recognition strategies for infection. Not only does it recognize foreign pathogens, also called 'microbial non-self' or pathogen-associated molecular patterns (PAMP), but it is able to detect metabolic consequences of cell injury, also called damage-associated molecular patterns (DAMP). The innate immune system is also able to recognise "missing self" molecules only expressed in normal healthy cells, and not in damaged or infected cells. These pattern recognition receptors (PRR) are expressed on antigen presenting cells such as dendritic cells and macrophages, other immune cells such as fibroblasts, mast cells, monocytes and neutrophils, and on non-immune cells such as epithelial cells.

Pattern recognition receptors include toll-like receptors (TLR), nucleotide oligomerization domain (NOD)-like receptors, collectin (collagen-containing C-type lectin) family receptors, retinoic-acid-inducible gene (RIG-1)-like receptors and absent in melanoma-2 (AIM2)-like receptors (Takeuchi & Akira, 2010; Turvey & Broide, 2010). TLR can detect a limited but highly conserved number of foreign antigens in the

extracellular space or in the lumen of endocytic vesicles. The human TLR family consists of 10 receptors. Examples include TLR 4, which recognizes lipopolysaccharide, TLR 5, which recognizes flagellin and TLRs 1,2 and 6, which recognise lipoproteins.

The collectin family of receptors are soluble pattern recognition receptors that bind to oligosaccharides or lipids on microorganism surfaces. An example of this is mannose binding lectin, which binds to microbial terminal mannose residues.

NOD-like receptors on the other hand recognise foreign antigen in the intracellular environment. There are 23 members of the NOD family. Examples include NOD2, that recognises muramyl dipeptide and Ipaf, that recognises flagellin. RIG-1-like receptors also recognise foreign antigens in the cytoplasm, in the form of double stranded RNA.

During cell lysis from infection, DAMPs are released and recognised through NOD-like receptors. Examples of DAMPs include high mobility group box 1, heat shock proteins and uric acid. On the other hand, recognition of normal cells, through expression of MHC Class I molecules by MHC class I-specific inhibitory receptors, inhibits activation of the immune response against uninfected tissue.

Activation of PRR upregulates transcription of multiple inflammatory response genes (Takeuchi & Akira, 2010). These genes encode enzymes, chemokines and pro-inflammatory cytokines. These include tumour necrosis factor, interleukin-1 and interleukin-6, type I interferons, adhesion molecules, regulators of the extracellular

matrix and antimicrobial proteins. Signaling pathways involved for each PRR are different and are reviewed in reference (Takeuchi & Akira, 2010).

Production of these substances leads to recruitment and activation of innate immune system leucocytes including natural killer cells, mast cells, eosinophils, basophils, macrophages and neutrophils. This process is assisted by vascular alteration by mediators including histamine, prostaglandins and nitric oxide, which cause vasodilation and increase vascular permeability. Tumour necrosis factor and interleukin-1 promote leucocyte extravasation by increasing leukocyte adhesion molecules on endothelial cells (Newton & Dixit, 2012).

Activated immune cells at the site of infection include dendritic cells, macrophages and neutrophils. These immune cells remove foreign particles and host debris by phagocytosis and subsequent digestion by production of reactive oxygen species, release of antimicrobial peptides as described previously and in the case of neutrophils, expulsion of nuclear contents to form neutrophil extracellular traps (Mayadas et al., 2014). Further specialized roles of tissue macrophages outside of inflammation are discussed in reference (Epelman et al., 2014). Inflammation also activates the complement cascade. For further information see (Dunkelberger & Song, 2010).

In summary, the reactive innate immune system consists of sentinel cells that express PRR and detect infection. Activation of these sentinel PRR induces the production of

inflammatory molecules, leading to recruitment of effector immune cells, which is facilitated by vascular changes.

The reactive innate immune system functions intimately with the adaptive immune system to provide a comprehensive system that can react with the speed inherent to the innate system, following-up with the specificity of the adaptive immune system. Although dendritic cells are part of the innate response, they serve a critical role in linking the innate immune system with the adaptive immune system.

The adaptive immune system

Although the innate immune system rapidly senses and eliminates pathogens, it is only able to recognise a limited set of PAMPs and DAMPs. As a result, pathogens have evolved mechanisms to avoid detection by the innate immune system.

With constant evolutionary drive between vertebrates and pathogenic microbes, an adaptive immune system provides an extra layer of protection with a broader and more finely tuned repertoire of recognition for self and non-self-antigens (Bonilla & Oettgen, 2010). The adaptive immune system also provides immunological memory, whereby on subsequent encounters with the same pathogen, a more rapid, robust and protective response is elicited (Bonilla & Oettgen, 2010).

The innate immune response makes crucial contributions to adaptive immunity by presenting antigen to the cells of the adaptive immune system, and activating

lymphocytes (cells that are part of the adaptive immune system) with complement fragments on microbial surfaces (Janeway et al., 2001).

Antigen presentation by dendritic cells

Induction of the adaptive immune system occurs when an immature dendritic cell phagocytoses a pathogen after binding and activation of a PRR. Micropinocytosis, whereby immature dendritic cells also continually take up extracellular material, can also lead to uptake of virus and bacteria and activate PRR. After activation of PRR, dendritic cells travel to a nearby lymph node and mature into a highly effective antigen presenting cell. At this stage, they are able to activate pathogen-specific lymphocytes (Janeway et al., 2001).

Antigens that are presented to lymphocytes have been processed by dendritic cells. Cytosolic and nuclear antigens are degraded in the cytosol by proteasomes to peptides. These peptides are then transported to the endoplasmic reticulum and stabilize newly synthesized MHC Class I heavy chain and B2 microglobulin and move to the cell surface. Interaction of the T cell receptor with antigenic peptides complexed with MHC Class I molecules activates naïve CD8 T cells. Activated CD8 T cells are then able to kill infected cells that present the same antigenic peptide/MHC Class I complex.

Antigens from the extracellular space are enclosed into endosomes in dendritic cells and are degraded into peptides. MHC Class II molecules in conjunction with invariant chain interact with these peptides. CLIP, a peptide place holder of invariant chain, dissociates by digestion once in the endosome, allowing peptides in the endosome to

bind to MHC Class II. The complex is then transported to the cell surface and can activate CD4 T cells expressing cognate antigen receptors. Different PRR control CD4 T cell function by determining the origin of the antigens recognised by the antigen receptors expressed on T cells. Different classes of dendritic cells produce different cytokines and, in doing so, instruct lymphocytes to induce the appropriate effector immune response (Iwasaki & Medzhitov, 2015). CD4 T cell subsets have distinct functions, including activation of other lymphocytes to kill microbes (T_H1), production of antibody and expulsion of helminths (T_H2), induction of inflammatory responses (T_H17), and dampening of immune activation (regulatory T cells). See (Bonilla & Oettgen, 2010; Iwasaki & Medzhitov, 2015) for further information.

T cell receptors are immensely diverse. Generation of T cell receptor diversity is similar to generation of antibody diversity in B cells and is discussed below. See (Dudley et al., 2005) for further information.

B cells provide humoral immunity by production of antibodies

While T cell functions include killing of infected target cells, provision of signals to enhance B cell and T cell responses, activation of mononuclear phagocytes and also regulation of immune responses, B cells primarily produce defensive antibodies (Bonilla & Oettgen, 2010), although they play important antigen-presenting and regulatory functions as well (LeBien & Tedder, 2008).

Antibodies are initially expressed as membrane-bound B cell receptors, but are secreted into blood plasma following activation of antigen-specific clones. Antibodies

function to neutralize toxins, prevent organisms adhering to mucosal surfaces, activate complement, opsonize bacteria for phagocytosis, and sensitise cells for antibody dependent cytotoxic attack (Parkin & Cohen, 2001). Antibody structure is critical for these properties. Each antibody consists of two types of polypeptide chains; two identical heavy chains and two identical light chains are linked together with disulfide bonds. This forms a Y shaped structure (see Figure 1.1 (Loureiro et al., 2015)).

Each antibody chain consists of variable (V) and constant (C) regions. Antigen binds to paired heavy and light chain V regions of the Y shaped structure. As the heavy and light chains are duplicated, each antibody bears two identical antigen binding sites. The base of the Y shape structure consists of the paired heavy chain C regions, forming part of the F(ab) region and the Fc region. It is the Fc region that interacts with effector cells by binding to the Fc receptor on effector cells or binding to other molecules such as complement (Schroeder & Cavacini, 2010). There are 5 classes of Ig: IgM, IgD, IgG, IgA and IgE. See (Schroeder & Cavacini, 2010) for further information.

The structure of the antigen binding site is determined by the unique amino acid sequence of the three complementary determining regions located inside the V region (see Figure 1.1) (Jung et al., 2006). While there are a very limited number of PRR involved in innate immunity, humans are capable of producing antibodies with about $5x10^{13}$ different antigen binding sites (Pieper et al., 2013).

Each B cell produces one type of antibody only, as a consequence of allelic exclusion, which will be explained later. Each B cell's unique antigen binding site is brought about

by somatic "V(D)J recombination" of genomic DNA within the *Ig* loci during B cell development to create heavy chain and light V regions unique to each B cell clone. V(D)J recombination is a semi-random process that recombines separate V_H, D_H and J_H germline gene segments to produce the heavy chain VDJ_H variable region encoded by the *Igh* locus (IgH heavy chain), then V_L and J_L germline gene segments recombine to encode the light chain variable region using either the *Igk* locus (Igκ light chain) or the *Igl* locus (Igλ light chain). V(D)J recombination is a process of precise double stranded break (DSB) formation, followed by imprecise non-homologous end joining (NHEJ) that permanently deletes large segments of genomic DNA up to a megabase in length (Fugmann et al., 2000).

V(D)J recombination creates antibody diversity

The heavy chain locus (*Igh*) is located on Chromosome 12 or 14 in mouse or human, respectively. The light chain *Igκ* and *Igλ* loci are located on chromosomes 6 and 16, respectively, in mice and on chromosomes 2 and 22, respectively, in humans (Honjo, 1983). In mice, the *Igh* locus includes 141 of V_H (in strain C57/BL6 (de Bono et al., 2004)), 13 of D_H and 4 of J_H. In humans, there are about 65 of V_H, 27 of D_H and 6 of J_H; exact numbers of germline V gene segments vary between mouse strains and between human haplotypes. As the light chain sequence can be derived from either the *Igκ* or *Igλ* light chain loci, the possible VJ-rearrangements can include combinations in mice of any one of 100-300 of V_κ with any one of 5 of J_κ, or 2 of V_λ with 2 of J_λ (see Figure 1.2) and 40 of V_κ with 5 of J_κ or 30 of V_λ with 4 of J_λ in humans. Functional V, D and J gene segments are flanked by non-coding recombination signal sequences (RSS). Recombination signal sequences consist of a highly conserved

heptamer (consensus 5'-CACAGTG) followed by a poorly conserved spacer usually 12 or 23 bases long, followed by a conserved nonamer (consensus 5'-ACAAAAACC) (Fugmann et al., 2000). There is variation between heptamer and nonamer of individual RSS, with those most closely resembling the consensus sequences being the most efficiently re-arranged (Dudley et al., 2005).

In mice and humans, V_H segments are 3'-flanked by RSS with 23-bp spacers, while the D_H segments are flanked at both the 3' and 5' ends by RSS with 12-bp spacers. Like the 3' end of V_H, the 5' end of J_H is also flanked by RSS with 23-bp spacers. V_κ segments are 3'-flanked by RSS with 12-bp spacers. J_κ are flanked at their 5' ends by RSS with 23-bp spacers. V_λ segments are flanked by RSS with 23-bp spacers at their 3' ends and J_λ are 5'-flanked by RSS with 12-bp spacers (see Figure 1.2).

The initiator of V(D)J recombination is the RAG endonuclease. The genes that produce RAG endonuclease are *recombination-activating genes 1 and 2* (*RAG1 and RAG2*) (Jung et al., 2006), which are only expressed in developing lymphocytes. A null deficiency for either RAG 1 or 2 leads to a complete block in B and T lymphocyte development at progenitor stages and null mutations in *RAG 1 or 2* are amongst the causes of severe combined immune deficiency in humans and mice (Dudley et al., 2005; Tasher & Dalal, 2012). RAG endonuclease introduces DNA double stranded breaks specifically between RSS and their associated *V, D or J* segments (Fugmann et al., 2000). Efficient recombination occurs between gene segments with a 12-base spacer RSS to gene segments with a 23-base spacer RSS, which is known as the 12/23

rule (Fugmann et al., 2000). Enforcement of the 12/23 rule appears to occur at the level of binding and assembly of RAGs 1 and 2 to paired RSS and at the subsequent cleavage step. The 12/23 rule prevents non-productive *V* to *V* or *J* to *J* joining and ensures the incorporation of D_H segments into the heavy chain variable region (Dudley et al., 2005).

The process of cleavage begins with bringing RAGs 1 and 2 to a single 12-bp RSS, followed by integration of the companion 23-bp RSS. The DNA bending proteins HMG1 and HMG2 facilitate the binding of both the 12-bp and 23-bp RSS and promote RAG-mediated cleavage (Dudley et al., 2005). RAGs introduce a single strand nick in the DNA between the border of the RSS heptamer and the gene-coding segment creating a 3'-OH on one DNA strand of the gene-coding segment and a 5' phosphate on the RSS strand. The 3'-OH then acts as a nucleophile and attacks the opposite DNA strand in a transesterification reaction, and forms a covalently sealed hairpin on the coding strand and a blunt 5'-phosphorylated RSS end (Dudley et al., 2005).

RAG-mediated DSBs are joined by non-homologous end joining (NHEJ). NHEJ proteins include Ku70, Ku86, XRCC4, XLF, ligase IV, DNA-PK_{cs} and Artemis. Initially Ku70 and Ku86 form a heterodimer that binds to RAG generated breaks. The Ku heterodimer then recruits DNA-PK_{cs} that bridges gene segment junctions (Malu et al., 2012). Artemis, in conjunction with DNA-PK_{cs}, mediates hairpin opening within four or five nucleotides of the apex of the hairpin, and this results in short 5' or 3' overhangs (Dudley et al., 2005). This ultimately increases junctional diversity, and, in doing so, increases the repertoire of antibodies generated. Polymerase μ is able to promote

NHEJ of ends with non-complementary 3' overhangs, by priming synthesis from a 3' overhang from one DNA end, and using a noncomplementary overhang from a second end as a template (Davis et al., 2008). Polymerase μ performs this by adding nucleotides in a template-independent manner and generates microhomology between junctions for pairing and ligation (Chang et al., 2017). Nucleases are able to chew back 3' overhangs, and this causes deletion of sequences at the junction. Terminal deoxynucleotidyl transferase, a lymphoid specific protein, also adds non-templated nucleotides to the heavy chain junctions, further increasing junctional diversity. Finally, ligation is catalyzed by ligase IV, XRCC4 and XLF. XLF and XRCC4 interact to form filaments that bridge DNA ends and stimulate ligase IV recruitment to perform ligation (Andres et al., 2012; Roy et al., 2015).

Like all blood cells, B lymphocytes are derived from haematopoietic stem cells (HSC) resident in the liver during embryogenesis and in the bone marrow after birth. HSC originate from the aorta-gonad mesonephros, formed from the mesoderm (Pieper et al., 2013). D_H to J_H V(D)J-recombination initially occurs in the *Igh* locus in liver pro-B cells in the foetus or in bone marrow pro-B cells post-partum. This is followed by V_H to DJ_H V(D)J-recombination. Once an IgH polypeptide is produced, the *Igh* loci in both chromosomes become inaccessible to RAG-endonuclease; a brief burst of proliferation occurs to "amplify" the successful VDJ_H-recombination, then *V* to *J* recombination commences in one allele of either the *Igκ* or *Igλ* loci in the resulting pre-B cells. If V(D)J-recombination fails to produce an IgL chain, it will continue in the same allele or in another *Igκ* or *Igλ* allele until all alleles are unable to undergo further recombination. Pre-B cells that fail to produce IgH and IgL chains will die by apoptosis,

but cells that have produced IgH and IgL chains differentiate into immature B lymphocytes expressing surface IgM and IgD. Immature B cells undergo negative selection to minimize export of self-reactive cells from the fetal liver or bone marrow. Most immature B cells with autoreactive receptors undergo apoptosis or overcome auto-reactivity *via* receptor editing driven by further V(D)J-recombination or by somatic mutation, but "central tolerance" is not perfect, and potentially auto-reactive B cells are exported. For further information see (Nemazee, 2006). B cells that survive negative selection leave the bone marrow as transitional B cells, and mature into follicular B cells (Cambier et al., 2007).

Somatic mutation creates high affinity antibodies in response to antigen challenge

Prior to activation by antigen, B cells circulate through the body through the vasculature. B cells then enter secondary lymphoid organs and will filter through lymphoid follicles for approximately 24 hours. If activation does not occur, then the B cell re-enters the circulation.

B cell activation occurs in lymphoid B-cell follicles, which occur adjacent to T cell zones (see (Hamel et al., 2012) for review). Here, naïve B cells encounter T-dependent antigens that enter the lymph node through lymphatics. Binding of the B cells' antigen receptor (BCR, formed from CD79a + CD79b + membrane Ig produced by alternative splicing of the ultimate CH exon) by antigen activates B cells via signaling from the BCR to the nucleus (Healy et al., 1997). Activated B cells migrate to the T cell/B cell border, where they can be rescued from programmed cell death by T_H cell-derived secondary

signals as long as they successfully present peptides derived from internalization of BCR-bound antigen as peptide/MHC class II complexes to cognate T_H cells (Shulman et al., 2014). This dependence on a second signal from antigen-specific T_H cells inhibits inappropriate activation of auto-immune B cells that escaped central tolerance.

Following secondary activation by cognate T cells, some B cells migrate to the central region of the B cell follicle and proliferate rapidly to form germinal centres (GC) (see Figure 1.3). In the GC, proliferating B cells are called centroblasts (Z. Li et al., 2004). Proliferating centroblasts undergo somatic hypermutation, a process which introduces nucleotide substitutions (and less commonly, insertions and deletions) in the *V(D)J* and 3' flanking regions of the rearranged antibody genes, leaving the C regions unmutated. Somatic hypermutation introduces mutations at a rate 10^6 times higher than the rest of the genome (Bardwell et al., 2004). This leads to production of daughter clones with antibody variants that bind antigen with differing affinities. B cells producing antibody with maintained or increased affinity are positively selected and undergo a proliferative burst (Neuberger & Milstein, 1995); the extent of the burst being proportional to the amount of antigen that the daughter cell has been able to capture from follicular dendritic cells and re-present to follicular T_H cells (Gitlin et al., 2015; Gitlin et al., 2014). Selection occurs when centroblasts drop out of cell cycle to become centrocytes, and is provided by both follicular dendritic cells and T-follicular helper cells (Hamel et al., 2012) (see Figure 1.3 (Heesters et al., 2014)). Proliferation and somatic hypermutation occur in the dark zone of the germinal centre, while selection occurs in the light zone (Gitlin et al., 2015; Gitlin et al., 2014).

Two processes are therefore involved in creating high affinity antibodies. Firstly, high levels of nucleotide substitutions in the variable region, termed somatic hypermutation, is combined with secondly, clonal selection. Daughter lymphocytes with higher affinity antibodies continue to proliferate, while lymphocytes with lower affinity undergo apoptosis (Z. Li et al., 2004). B cells with the highest affinity are rapidly selected into the antibody secreting pool (Krautler et al., 2017).

Class Switch Recombination alters production of antibody isotype

While somatic hypermutation is involved in creating high affinity antibodies, class switch recombination (CSR) is a process that alters the isotypes of antibodies produced from the IgM and IgD (predominantly IgM) isotypes produced by all naïve B cells to IgG, IgE or IgA isotypes following B cell activation. Switching from pentameric IgM to the much smaller IgG, E and A isotypes allows antibodies to diffuse out of the blood into other tissues, and also changes antibody effector functions; both potentially improve the ability of antibodies to eliminate target pathogens (Chaudhuri et al., 2003; Stavnezer & Schrader, 2014). In the germline, exons encoding IgM ($C\mu$) and IgD ($C\delta$) occur 3' to the J_H gene segments, separated from the J_H by the J-C_H intron. Either $C\mu$ or $C\delta$ exons are spliced to the VDJ_H exon during transcription in naïve B cells, such that mature naïve B cells express both IgM and IgD. Downstream (i.e. 3') of $C\delta$ are additional constant region exons that are not used by naïve B cells which encode Fc distinct from IgM and IgD. In humans, the order (5' to 3') of these additional constant region gene segments are $C\gamma_3$, $C\gamma_1$, $C\alpha_1$, $C\gamma_2$, $C\gamma_4$, $C\varepsilon$ and $C\alpha_2$, which encode

the Fc of IgG_3, IgG_1, IgA_1, IgG_2, IgG_4, IgE and IgA_2, respectively. Located upstream (i.e. 5') of every constant region gene lies a non-coding switch (*S*) region sequence.

Class switch recombination is the process of deletional recombination between the $S\mu$ switch region and one of the 3' $S\gamma$, $S\varepsilon$ or $S\alpha$ regions (see Figure 1.4). Combination between $S\mu$ and a downstream *S* region changes the constant region exons placed immediately 3' to the splice acceptor in the $J\text{-}_H$ intron, leading to the production of a different antibody isotype without changing antigen specificity. Switch (S) regions are typically 1-10kb in length, G-rich in the top strand and are highly repetitive, consisting of $G_{4\text{-}5}$ tracts, which can form G-quadruplexes when single-stranded, interspersed with WGCW (W = A or T) motifs. The non-canonical switch region lying upstream of $C\delta$ (known as σ_δ) has residual function and switching to IgD is a rare event (Choi et al., 2017). (see Figure 1.4).

Somatic hypermutation and Class Switch Recombination are initiated by Activation Induced Cytidine Deaminase (AID)

Activation induced cytidine deaminase (AID, gene *Aicda*) is essential for both somatic hypermutation and class switch recombination. AID deficiency causes a complete defect in class switching and somatic hypermutation (Muramatsu et al., 2000), and is the cause of the human disorder Hyper-IgM syndrome 2 (HIGM2) (Revy et al., 2000).

AID initiates somatic hypermutation and class switch recombination by de-aminating genomic cytosine residues, converting them into uracil residues (see Figure 1.5) (Frieder et al., 2009). In the absence of further processing, replication uracils produces

transition C:G to T:A mutations, because replicative DNA polymerases incorporate A opposite U in DNA (Frieder et al., 2009).

The sequence homology between AID and Apobec-1, which is the catalytic component of the complex that deaminates cytidine to uracil in Apolipoprotein B mRNA, initially suggested that the mechanism of AID mutation might be deamination of cytidine to uracil in a specific unknown RNA target (Muramatsu et al., 1999). However, the first evidence that AID functions by directly mutating cytosine bases in genomic DNA was produced when AID was overexpressed in *E.coli*. Subsequent sequencing of *E.coli* DNA sequences revealed that 80% of AID-induced mutations in *E. coli* were C:G to T:A transitions and that these mutations were enhanced when the *E. coli ung* gene was inactivated; the UNG protein removes uracils from DNA, but not from RNA (Petersen-Mahrt et al., 2002). Direct evidence that AID converts genomic cytosines into uracils *in vivo* was produced by quantifying uracil residues in immunoglobulin variable regions in genomic DNA from murine germinal centre B cells (Maul et al., 2011). *In-vitro,* DNA deamination by AID in the absence of cofactor proteins was observed in single stranded DNA and not in double stranded DNA, DNA/RNA hybrids or RNA (Bransteitter et al., 2003; Dickerson et al., 2003). Transcription exposes single stranded DNA (ssDNA) and enhances AID mediated cytidine deamination; this initially supported the concept that AID requires ssDNA as a substrate (Ramiro et al., 2003). In the presence of co-factors, however, AID deaminates both ssDNA and dsDNA templates *in vitro* (Basu et al., 2011).

Nucleotide substitutions are not randomly distributed along the variable region of *Ig* genes. Mutation hotspots occur where nucleotide substitutions are likely to be effective at generating useful amino acid substitutions to optimize affinity maturation of the immune response (Jolly et al., 1996; Neuberger, 2008). AID deaminates C residues more often if they occur as part of a WRCY motif (W=A/T; R=A/G; Y=C/T) in either DNA strand (Rogozin & Diaz, 2004; Wang et al., 2010). AGCT is a palindromic variant of WRCY that encodes apposed hotspots on both strands; that is, the C on both strands is favoured by AID at AGCT sites. Of all WRCY motifs, AGCT, and outside *Ig* genes AACT, are the most mutated hotspots (Alvarez-Prado et al., 2018). WRCY motifs are enriched in the CDRs of *Ig* V genes and depleted from the framework regions (Jolly et al., 1996; Neuberger, 2008), and as mentioned earlier, AGCT motifs are very dense in switch regions (Zarrin et al., 2004), which facilitates CSR (Han et al., 2011; Yeap et al., 2015). Substitution of AID's DNA-binding site with the DNA binding sites from AID homologues (APOBEC proteins) that target motifs distinct from WRCY skewed mutation during *Igλ* mutation in chicken DT40 cells away from WRCY motifs and towards the motifs preferred by the donor APOBEC proteins (Wang et al., 2010). This elegantly proved that the mutator function of AID was determined by AID's acting directly on genomic DNA *in vivo*.

The action of AID is tightly regulated *in vivo*

As AID is a potent mutator (58), its expression is tightly controlled. In B cells, AID is expressed after activation (Muramatsu et al., 1999). AID is also expressed in the ovary, where its function may lie in epigenetic programming (Orthwein & Di Noia, 2012). The mechanism for controlled expression of AID includes a combination of binding sites

for factors such as NF-κB, Stat 6, Hox C4 and Pax5 in *Aicda's* promoter (Maul & Gearhart, 2010a) and also enhancers and silencers located within four upstream regulatory regions (Tran et al., 2010). *Aicda* transcription in B cells is induced by engagement of B cell CD40 by CD40 ligand (CD154) and by cytokines such as interleukin-4 (IL-4) (Dedeoglu et al., 2004), both of which are provided by T cells during immune synapse formation with activated B cells (Orthwein & Di Noia, 2012).

Given the preference of AID for single stranded DNA *in vitro*, and, given that single stranded DNA is exposed during transcription, it is thought the immunoglobulin gene is targeted when it is actively transcribed. This is indirectly supported by the observation that immunoglobulin transgenes undergo somatic hypermutation at levels that correlate with their transcription rates (Bachl et al., 2001). An analysis of the 118 most expressed genes in germinal centre B lymphocytes revealed 43% of these genes were also mutated by AID. The rates of mutation in these genes, however, were at least 100-fold lower than the variable region (Liu et al., 2008). Active transcription, therefore, is not the sole reason that AID targets *Ig* genes.

Recruitment of AID is more closely correlated to sites of RNA polymerase II (RNA Pol II) stalling than to sites of transcription. In chromatin immunoprecipitation analyses, AID was associated with the 5' ends of many transcription domains, along with RNA Pol II, but the AID co-factor RPA, which is required for AID activity on dsDNA *in vitro* was predominantly associated with *Ig* loci (Yamane et al., 2011). A spike in RNA Pol II density was seen in immunoglobulin switch regions in activated B lymphocytes and is thought to be due to elongation stalling (Rajagopal et al., 2009). Elongation stalling is

an intrinsic feature of immunoglobulin locus transcription (Pavri & Nussenzweig, 2011). Spt5 is a factor that is recruited to sites of stalling and is required for CSR. Spt5 was found to facilitate the association between RNA Pol II and AID and their recruitment to immunoglobulin targets (Pavri et al., 2010).

Interacting with RNA Pol II is the RNA exosome, which disassociates stalled RNA Pol II from DNA (Andrulis et al., 2002). The RNA exosome is involved in 3' processing of various stable RNA species. Its function is crucial for RNA quality control in the nucleus (Kilchert et al., 2016). It degrades many types of cryptic transcripts formed by pervasive transcription and removes aberrant RNA molecules that fail to mature correctly. *In-vitro* and *in-vivo* experiments found that the RNA exosome bound to AID and stimulated AID activity on both template and non-template strands *in-vitro* (Basu et al., 2011). This suggested that stalled RNA Pol II recruited the RNA exosome to the *Igh* locus and in doing so recruited AID. Supporting this theory is the presence of DNA sequences within the immunoglobulin loci that express non-coding RNAs that are substrates for the RNA exosome (Laffleur et al., 2017).

During somatic mutation, processing downstream of UNG2 introduces transversions at G and C bases

On its own, conversion of cytosine into uracil can only produce C:G to T:A transition mutations, by subsequent replication across U. Although these are the most common mutations in *Ig* genes that have not undergone antigen selection, mutations at A:T base pairs and transversion mutations at C:G are also abundant in hypermutated B cells (Methot & Di Noia, 2017). Uracil is directly recognized as DNA damage by DNA

base excision repair (BER) factors UNG2, SMUG1, TDG and MBD4; furthermore U:G base pairs are recognized as DNA mismatches by the DNA mismatch repair factor MutSα, which is a heterodimer of MSH2 and MSH6 (Jacobs & Schar, 2012). B cells deficient in both the BER factor uracil N-glycosylase (UNG2) and MutSα accumulate only C:G to T:A transition mutations in their *Ig* loci, at elevated frequencies compared to wild-type B cells (Rada et al., 2004; Shen et al., 2005), while somatic hypermutation of *Ig* genes in wild type mice includes all possible nucleotide substitutions. Processing of U:G base pairs by these two proteins are therefore required for transversions at G and C bases and for mutations at A and T bases (see Figure 1.6 (Petersen-Mahrt et al., 2002)).

When UNG2 is inactivated; transversions at C and G bases are reduced by greater than 80% (Rada et al., 2002; Storb & Stavnezer, 2002). This is distinct to the high-fidelity function of UNG2 in the canonical base excision repair (BER) pathway. In canonical BER, UNG2 cleaves the N-glycosylic bond of genomic uracils, creating an apyrimidinic (AP) site (Krijger et al., 2009). Subsequently, two different pathways can process the apyrimidinic site. Short-patch BER recognizes the site by APE1, which creates a nick in the DNA immediately 5' to the AP site. The AP-lyase activity of APE1 or of DNA polymerase β then removes the AP base. The original base is then restored by DNA polymerase β and XRCC1/DNA ligase III (Krokan & Bjoras, 2013). DNA polymerase λ can substitute for Pol β. Long-patch BER excises at least two nucleotides and up to 20 nucleotides, with DNA synthesis via polymerase β or polymerase δ/ε (Sattler et al., 2003; Woodrick et al., 2017). Unlike single-nucleotide BER, long-patch BER requires

Proliferating cell nuclear antigen (PCNA), the trimeric sliding clamp that tethers DNA polymerases to the template during replication (Krokan & Bjoras, 2013).

More specifically, when the activity of UNG2 is restricted to G1-phase by cell cycle restricted overexpression of a uracil glycosylase inhibitor polypeptide (ugi), it was found that UNG2 excision in G1-phase was responsible for mutagenic processing of AID induced uracils, and surprisingly also for most of the faithful repair (Sharbeen et al., 2012).

Current unresolved questions involve the mechanisms downstream of uracil excision to achieve this result. It is highly likely that replication across AP sites with a low fidelity polymerase is the major mechanism. Supporting this theory is that mice deficient for Rev1 DNA polymerase, a translesion synthase that adds dC opposite AP sites, have an almost complete loss of C:G to G:C transversions, while the frequencies of C:G to A:T transversions are unaffected (Jansen et al., 2006; Maul & Gearhart, 2010b). B cells double-deficient for Rev1 and the translesion polymerase Pol η, which can accommodate several DNA lesions in its active site (Parijs et al., 1999), or double-deficient for REV1 and MSH2 have almost no *Ig* C:G to G:C transversions at all, but again C:G to A:T transversions are unaffected (Kano et al., 2012; Krijger et al., 2013). Thus, REV1 and Pol η activity downstream of UNG2 and MutSα, respectively, are responsible for C/G to G/C transversions, but the source of C:G to A:T transversions remains unknown. It is also unknown when in the cell cycle bypass of AP sites by REV1, Pol η or other translesion polymerases might occur. It was proposed that lesion bypass

occurs in S-phase, when unrepaired AP sites persist from G1-phase (Sharbeen et al., 2012), but this is far from certain.

The mismatch repair pathway introduces mutations at A and T bases through recruitment of DNA polymerase η

In conventional post-replicative mismatch repair (MMR), which occurs in S-phase (Hombauer et al., 2011), the MutSα protein (a heterodimer of MSH2 and MSH6) binds to single base mismatches, most of which have been produced by occasional mis-incorporation during replication fork progression. Activation of MutSα by mismatches leads to MutSα translocation and recruitment by MutSα of MutLα (a heterodimer of PMS2 and MLH1). The DNA is then nicked distal to the mismatch site by PMS2. Exonuclease 1 (Exo1) binds MSH2 and MLH1 and creates an excision patch greater than 100bp long starting from the nicked site; DNA Polymerase δ (one of the two DNA polymerases used for conventional high-fidelity replication) then fills in the excision patch. This process repeats until the mismatch driving MutSα activation is repaired (Jiricny, 2013). AID-induced MMR, on the other hand is distinct from this pathway, because it is highly mutagenic and because it involves factors that are not involved in canonical post-replicative MMR (see below).

Low fidelity MMR is the main contributor of mutations at A and T bases. More specifically, deficiency in MSH2, MSH6 or Exo1 is associated with an 80-95% loss of *Ig* mutations at A and T bases. A complete loss in mutations at A and T bases, however, is observed in UNG/Msh2 and UNG/Msh6 double-deficient GC B cells and, therefore, UNG2 is able to generate some mutations at A:T independent of MSH2, perhaps *via*

error-prone long patch BER (Krijger et al., 2009; Rada et al., 2004; Shen et al., 2008). A deficiency in UNG2 alone is associated with a modest and insignificant reduction of mutations at A:T bases (Rada et al., 2002).

While B cells deficient in PMS2 and MLH1 display unchanged mutation frequencies (Maul & Gearhart, 2010b), a deficiency in both UNG and PMS2 leads to a 50% reduction in mutations at A:T bases. Reduction in mutation at A:T bases is also seen in a $Ung^{ko/ko}$ background with additional deficiencies in other enzymes with uracil-DNA glycosylase activity; these are TDG and SMUG1 (Girelli Zubani et al., 2017). Double-deficiency for the BER factors APE1 and APE2 also impacts A:T mutation significantly (Stavnezer & Schrader, 2014). Overall, the data suggest that uracil BER and MutLα create nicks from deamination sites redundantly, to then allow DNA entry by Exo1 during AID-induced MMR (Girelli Zubani et al., 2017).

The mutagenic phase of AID-induced MMR has been associated with polymerase η, the same translesional DNA polymerase (TLS) responsible for C:G to G:C transversions downstream of MMR. Pol η can bypass DNA lesions at replication fork stall sites; TLS are also error-prone (Weill & Reynaud, 2008). *In-vitro,* MSH2 was shown to bind to polymerase η and the MSH2-MSH6 heteroduplex stimulated the catalytic activity of polymerase η (Wilson et al., 2005). In humans with xeroderma pigmentosum variant (XPV), a condition with defective polymerase η, analysis of DNA from lymphocytes showed reduced mutation at A and T bases (Masuda et al., 2008; Mayorov et al., 2005; Yavuz et al., 2002; Zeng et al., 2004). There was also an increase in mutations at C bases, especially transversions (Zeng et al., 2004). In mice, inactivation of *POLH* (the

gene for polymerase η) also leads to decreased mutation at A and T bases (Delbos et al., 2005; Weill & Reynaud, 2008), and an increase in transition mutations at C and G bases (Weill & Reynaud, 2008). In the absence of polymerase η, it appears that polymerase κ contributes to residual mutations at A and T bases. Mice deficient for both polymerase η and polymerase κ show a further reduced mutation frequency at A and T bases compared to polymerase η deficiency alone (Faili et al., 2009).

Interestingly, studies using MSH2, polymerase η and combined MSH2 and polymerase η deficient mice showed that mutations at A and T bases as a percentage of total mutations in mouse *Ig* somatic hypermutation to be 10, 16 and 1% respectively (Delbos et al., 2007). This reinforced the theory that polymerase κ contributes to residual mutations in the absence of polymerase η, but also that mutations at A and T bases due to polymerase η are not entirely downstream of MSH2; some are presumably downstream of UNG2 and are introduced during long-patch BER (Delbos et al., 2007).

Of the other TLSs, it is uncertain if polymerase θ contributes to somatic hypermutation downstream of MMR (Masuda et al., 2007; Masuda et al., 2006). It does not appear that polymerase ι contributes to mutations downstream of MMR (Weill & Reynaud, 2008).

Mechanisms for polymerase eta recruitment

Proliferating cell nuclear antigen (PCNA) is a homo-trimeric sliding clamp that tethers DNA polymerases to the DNA template during replication (Weill & Reynaud, 2008).

Ubiquitination of PCNA occurs during replication fork stall, which recruits TLS polymerases (Weill & Reynaud, 2008). Specifically, monoubiquitination at lysine 164 is responsible for recruitment of polymerase η (Hoege et al., 2002). It has been shown that after UV damage, PCNA becomes ubiquitinated, and it is the monoubiquitinated form of PCNA that interacts with polymerase η (Kannouche et al., 2004). Studies in mice revealed that B cells with PCNA resistant to ubiquitination at lysine 164 showed a large reduction of mutations at A and T bases in *Ig* genes (Roa et al., 2008). In G1-phase human B cells, PCNA ubiquitination is dependent on MMR; furthermore PCNA ubiquitination recruits polymerase η (Pena-Diaz et al., 2012).

Class Switch Recombination requires double strand break formation

Class switch recombination is an event that recombines double strand DNA breaks between the donor μ switch region (Sμ) and a downstream acceptor switch region. This process begins with cytosine deamination by AID. Further processing by BER alone, or together with MMR produces the vast majority of double strand breaks. *Ung*-knockout (producing uracil-BER deficiency) results in a 95-99% reduction in CSR (Stavnezer & Schrader, 2014). Another uracil DNA glycosylase: SMUG1, is able to partially substitute for UNG in CSR when over-expressed (Di Noia et al., 2006), and *Smug1*-knockout five-fold further reduced CSR in UNG deficient mice from its already very low base. No effect on CSR was seen in *Smug1*-knockout, UNG competent mice, however (Dingler et al., 2014). $Msh2^{ko/ko}$ cells switch at a rate of 15-50% of wild type cells and, in general, deficiencies in MMR factors including MSH2, MSH6, MLH1, PMS2 and EXO1 have reduced CSR by two to seven-fold. A combination of UNG and MSH2 deficiency ablates CSR (Stavnezer & Schrader, 2006).

A simple model for DSB generation in CSR has emerged. Most deamination sites are detected by UNG and are converted into AP sites in G1-phase (Sharbeen et al., 2012). APE1 and/or APE2 then cuts at these AP sites, creating ssDNA nicks (Stavnezer & Schrader, 2014). When two ssDNA nicks occur in close proximity on opposite strands, this creates a double strand break. When ssDNA nicks are too far apart to generate a double strand break, MMR can convert the single strand nicks to double strand breaks provided some U:G mismatches able to activate MutSα are left in place by UNG2 (Schrader et al., 2007; Stavnezer & Schrader, 2014) (see Figure 1.7).

Usually, BER is a highly active and error free pathway (Stavnezer & Schrader, 2014), however given the large numbers of deaminations at switch regions, it appears to be overwhelmed before repair can be completed (Stavnezer & Schrader, 2014). Indeed, germinal centre B cells express very low levels of APE1 compared to other cells (Stavnezer & Schrader, 2014). APE2 is a much less efficient endonuclease compared to APE1 with around 1000-fold lower activity that may impair efficient BER. Thus, APE2 actvity may promote DSB formation by reducing faithful BER of uracils.

Non-homologous end joining recombines S region double strand breaks

Two major DSB repair pathways have been elucidated: homologous recombination (HR) and non-homologous end joining (NHEJ) (Davis & Chen, 2013). NHEJ is the DSB repair pathway that leads to CSR.

In NHEJ, after DSB formation, the Ku heterodimer (composed of Ku70 and Ku86) binds to both DNA ends flanking the DSB. Ku can also interact with various DNA fragments that contain nicks (Yang et al., 2016). Ku serves as a scaffold to bind other NHEJ factors; after binding to DNA sequences, the conformation of Ku is modified. Ku is then able to bind other NHEJ factors. DNA-dependent protein kinase catalytic subunit (DNA-PK_{cs}) binds to the Ku heterodimer and forms the DNA-PK complex (Chang et al., 2017). The binding of DNA-PKcs is assisted by scaffolding of Ku and DNA-PKcs by the long noncoding RNA LINP1 (Y. Zhang et al., 2016). DNA-PK_{cs} is thought to function to co-ordinate processing and ligation of DSBs (Yang et al., 2016).

Processing of ends are required if they are incompatible; this is to achieve short regions of microhomology (less than four nucleotides) for end recombination (Chang et al., 2017). DNA-PK_{cs} recruits Artemis, which has intrinsic 5' exonuclease activity, and when in complex with DNA-PK_{cs}, has endonuclease activity on both 5' and 3' overhangs (Chang et al., 2017). Artemis also functions to bridge DNA ligases to the DSBs. The Ku-DNA complex also recruits polymerases μ and λ. Pol μ adds nucleotides in a template independent manner at 3' incompatible overhangs to achieve microhomology, while Pol λ promotes ligation of terminally compatible overhangs that require fill-in activity (Chang et al., 2017).

Proteins belonging to the XRCC4-like family include XRCC4 and XLF. These proteins function to co-stabilise DSB ends to enable bridging and ligation. Ligation is then performed by the XRCC4-XLF-Lig4 complex (Yang et al., 2016).

The effect of cell cycle on Class Switch Recombination

AID dependent DSBs are predominantly detected in the G1-phase of the cell cycle. This was found *in-vitro* by activating mouse splenic B cells with lipopolysaccharide and IL-4 and then sorting for G1-phase vs S/G2/M-phase fractions (Schrader et al., 2007). This showed that AID-dependent DSBs were more abundant in the G1-phase fraction. Independent experiments showed that AID activity is restricted to early G1-phase (Wang et al., 2017); AID (as an AID-EGFP fusion protein) was transiently localised in the nucleus from prometaphase until early G1-phase. AID deamination of the S region was also seen to be restricted to early G1-phase. Supporting this, experiments that enforced nuclear localization of AID to G1-phase accelerated CSR and somatic hypermutation, while enforced nuclear localization of AID to S/G2/M compromised cell viability (Le & Maizels, 2015).

It is also G1-phase UNG that leads to DSB formation during CSR (Sharbeen et al., 2012). Although the cell cycle phase of MMR involved in CSR is not known, mouse B cells with PCNA-K164 mutation (which makes PCNA resistant to ubiquitination) displayed a decrease in CSR as well as a reduction in mutation at A:T, with a reduction in switching to IgG3 by 50% and IgG1 by 25% (Roa et al., 2008).

The ligation step in CSR occurs in S-phase. It was shown that activation of DNA replication origins in the *Igh* locus is required for CSR (Wiedemann et al., 2016).

Current unresolved questions

It is clear that AID hotspots exist because AID more often targets C in WRCY motifs than outside these motifs. Interestingly, MMR deficiency increases focusing of mutation in AID hotspots in mice with normal levels of UNG (Bardwell et al., 2004; Frey et al., 1998; Jacobs et al., 1998; Li et al., 2006; Phung et al., 1998; Rada et al., 1998; Wiesendanger et al., 2000). We do not know why this occurs but possible explanations include MMR-induced AID targeting (Li et al., 2006), or that C deaminations at AID hotspots are resistant to UNG processing compared to non-hotspot C deaminations (Delbos et al., 2007).

As previously stated, AID-induced MMR processing leads to mutation at A:T and involves mono-ubiquitination of PCNA and recruitment of polymerase η. We do not know why PCNA becomes mono-ubiquitinated, however, given that G1-phase MMR leads to PCNA mono-ubiquitination, then differences in cellular environments during G1 versus post G1-phase, such as dNTP pools may be responsible. We previously hypothesised that dNTP paucity in G1 is the cause of polymerase δ stalling and recruitment of polymerase η during AID-induced MMR (Sharbeen et al., 2010).

dNTP concentrations vary across the cell cycle

An optimal concentration of dNTPs is crucial for the fidelity of DNA synthesis during replication and repair (Rampazzo et al., 2010), but the supply of dNTPs is not constant through the cell cycle. Uncontrolled and high dNTP concentrations are known to be mutagenic for genome replication (Stillman, 2013). A limited pool of dNTPs also limits viral reverse transcription (Lahouassa et al., 2012). dNTP production and turnover are

regulated to increase dNTP levels in cycling cells and restrict levels in non-cycling cells (Ayinde et al., 2012). In S-phase, when replication of DNA occurs, dNTP pools are in the range of 10-100pmol for each dNTP per 10^6 cells, while in G0/G1-phase, dNTPs are about 10-fold lower (Rampazzo et al., 2010) (see Figure 1.8).

dNTP pools are determined by ribonucleotide reductase and SAMHD1

dNTP synthesis occurs via *de-novo* synthesis and salvage pathways. The *de-novo* pathway is responsible for the bulk of dNTP synthesis and is regulated by the activity of ribonucleotide reductase (RNR) (Nordlund & Reichard, 2006). RNR is a heterotetramer containing two copies of a large subunit, R1 and two small subunits, R2 or p53R2. While R1 is abundant through the cell cycle, R2 is restricted to S-phase. This restriction to S-phase translates to the much higher levels of dNTPs during S-phase. p53R2 is much less abundant than R2 but is expressed throughout the cell cycle to allow for mitochondrial DNA synthesis (Rampazzo et al., 2010). p53R2 is also transcriptionally activated by p53 after DNA damage, and is required for optimal dNTP pools for DNA repair in quiescent cells (Pontarin et al., 2012). RNR is recruited to sites of DNA damage *via* interaction of the R1 subunit with the histone acetyl transferase TiP60. This interaction is essential for G1-phase DSB repair, but is redundant in S-phase cells (Niida et al., 2010), and presumably overcomes local dNTP deficiency to enable G1-phase DNA synthesis during DSB end-processing.

The salvage pathway is dependent of the activity of thymidine kinase 1 and deoxycytidine kinase (Kunos et al., 2011). These enzymes phosphorylate

deoxyribonucleosides to produce deoxyribonucleoside monophosphates. Like ribonucleotide reductase, thymidine kinase 1 is restricted to S-phase (Kunos et al., 2011). Deoxycytidine kinase is constitutively expressed through the cell cycle (Kunos et al., 2011).

SAMHD1 is a cellular enzyme with deoxynucleoside triphosphate triphosphohydrolase (dNTPase) activity

On top of the regulation of dNTP pools by RNR and thymidine kinase 1, sterile alpha motif domain and histidine-aspartic domain containing protein 1 (SAMHD1) also contributes to the lower dNTP pools in G0/G1-phase. *SAMHD1* was initially identified in 2000 as a novel cDNA expressed in human peripheral blood monocyte derived macrophages and was designated as dendritic cell-derived IFN- γ (DCIP) (Li et al., 2000) . However, its function was not identified until 2009, when mutations in the gene were found in patients with Aicardi-Goutières syndrome (Rice et al., 2009).

SAMHD1 is a metal dependent deoxynucleotide-triphosphohydrolase (dNTPase) which degrades dNTPs to deoxynucleosides and triphosphate products (Powell et al., 2011). *In-vitro* studies show the recombinant human and mouse enzymes to cleave all dNTPs: dGTP, dATP, dTTP and dCTP (Goldstone et al., 2011; Powell et al., 2011). Studies in THP-1 human monocytic cell lines showed that shRNA-mediated knockdown of *SAMHD1* led to increased dNTP levels (Amie et al., 2013b; Lahouassa et al., 2012). Knockdown of *SAMHD1* by siRNA also showed increased dNTP levels compared to controls in lung and skin fibroblasts. *Samhd1* knockout mice (produced

by gene-trapping in embryonic stem cells) have 40-fold elevated dNTP levels (Behrendt et al., 2013).

SAMHD1 is a protein of 626 amino acids in human and 658 amino acids in mice. *In-vitro* studies on SAMHD1 mutants reveal residues 110-599 are responsible for dNTPase activity and these correspond to the HD1-like domain (Goldstone et al., 2011; Powell et al., 2011; Sze et al., 2013; White et al., 2013a). The HD1-like domain is sufficient to achieve dNTPase activity without the rest of the protein (Goldstone et al., 2011). This structure of SAMHD1 is a predicted phosphohydrolase HD (His---His-Asp---Asp) domain (Powell et al., 2011; Rice et al., 2009) with conserved doublet of histidine and aspartate (Franzolin et al., 2013). Similar enzymes that share the conserved doublets exist in *Escherichia coli*, which hydrolyses dGTP to deoxyguanosine and triphosphate, *Enterococcus faecalis* which hydrolyses all canonical dNTPs and also in *Thermus thermophilus,* which is also able to hydrolyse all canonical dNTPs (Franzolin et al., 2013).

Initial studies showed that SAMHD1 dNTPase activity is activated in the presence of dGTP. SAMHD1 is able to hydrolyse dGTP alone. However, to hydrolyse dATP, dUTP, dTTP and dCTP, dGTP has to be present as well (Goldstone et al., 2011). dNTPase activity was found to be twenty-fold higher in the presence of dGTP (Powell et al., 2011). Further studies have shown that GTP is also able to activate dNTPase activity, and rather than dGTP, is the likely activator of SAMHD1, as GTP levels are 1000-fold higher than dGTP levels (Amie et al., 2013a).

It is likely that SAMHD1 is regulated by GTP/dGTP binding to an allosteric site. Comparison of SAMHD1 protein to the related HD proteins from *Enterococcus faecalis* and *Aquifex aeolicus* suggested allosteric activation of the enzyme by dGTP. Examination of the HD domain and residues that interact with guanosine show that Asp 137, Gln 142 and Arg 145 recognize the activating guanosine edge, with identical hydrogen bonding with Watson-Crick and Hoogsteen (N7) sites of the guanine base (Koharudin et al., 2014). Base substitutions at Arg 145 and Asp 137 abolish dNTPase activity (Goldstone et al., 2011). Base substitution at Gln 142 reduces the dNTPase activity five-fold.

Examination of the catalytic site shows binding of dNTPs via the triphosphates with a metal ion, coordinated through two histidine residues and two aspartate residues (Ahn, 2016). H167, H206, D207 and D311 coordinate the metal ion binding in the active site and R164, H233, K321 and Y315 form hydrogen bonds with the triphosphate (Arnold et al., 2015).

Subsequent research has elicited the crystal structure of SAMHD1 as a tetramer in its active state, while usually existing in a monomer and dimer form in a non-active state. In the presence of dGTP/GTP binding to the activator site (A1), monomers and dimers assemble into the tetramer form. Subsequent formation of the tetramer form induces a conformational change in the dNTP binding pocket (A2) to yield the catalytically active enzyme that can bind all dNTPs (Ji et al., 2013; Koharudin et al., 2014). Out of all dNTPs binding to the A2 site, dATP is the most potent cofactor for inducing SAMHD1 tetramerisation (Ballana & Este, 2015) (see Figure 1.9 (Seamon et al., 2014)).

Further *in-vitro* studies have found that when the activator site, A1, was bound by GTP and the catalytic site, A2, bound by dNTP, this led to a long lived activated state that remained even after dNTP levels were depleted (Hansen et al., 2014). To summarise, SAMHD1 is a tetramer in its active state, and this is controlled by levels of GTP and dNTPs. However once in tetramer form, SAMHD1 remains in a long lived active state despite dNTP depletion and is still active, even when there is a very low level of dNTPs later on.

The rest of the SAMHD1 protein consists of the following features:

- An N-terminus extended loop with nuclear localization signal (amino acids 1-39); in humans this corresponds to residues ^{11}KRPR^{14}(Brandariz-Nunez et al., 2012). Mutagenesis of these residues change SAMHD1's nuclear distribution to the cytoplasm (Wu, 2013a). Patients with *SAMHD1* nuclear localization signal mutations have a cytosolic distribution of SAMHD1 (Goncalves et al., 2012).
- Putative protein-protein interaction sterile motif (SAM) domain (amino acids 40-109) consisting of four alpha helices (Powell et al., 2011). SAMs are 65–70 residues in length and can serve as protein-interaction modules mediating interactions with other SAM domain and non-SAM domain–containing proteins. In SAMHD1, however, a function of the SAM domain has not been yet elucidated (Rice et al., 2009). The SAM domain may possibly contribute to binding of ssDNA and ssRNA (Beloglazova et al., 2013).
- C terminus tetramerisation site (amino acids 600-626). Tetramerisation relies on interactions at these residues (Ahn, 2016). Deletion of the C-terminal region

of SAMHD1 prevents tetramerisation and ablates HIV restriction activity (Arnold et al., 2015).

Regulation of SAMHD1 deoxynucleotide-triphosphohydrolase activity

SAMHD1 levels have been shown to remain unchanged throughout the cell cycle in primary T lymphocytes and the monocyte cell line THP-1 (Yan et al., 2015). Initial reports showed that the T592 residue was phosphorylated in cycling THP-1 cells and in U937 cells transduced to express SAMHD1-FLAG (U937 cells do not express SAMHD1), but not in resting cells (White et al., 2013b). Initial *in-vivo* and *in-vitro* experiments however showed that T592 phosphomimetic mutants still had dNTPase activity, and that dNTP levels were comparable (White et al., 2013b). These early *in-vitro* assays, however, provided a level of dNTPs far in excess of intracellular levels.

Subsequent experiments showed, via size exclusion chromatography and multi-angle light scattering (SEC-MALLS), that T592 phosphorylated SAMHD1 (pSAMHD1) inhibited stable tetramer formation. This leads to disordered protein formation and loss of tetramer stabilizing interactions (Arnold et al., 2015). Experiments showed that SAMHD1 was able to hydrolyse dNTPs at much lower concentrations compared to T592 phosphorylated SAMHD1. This suggested that phosphorylation impaired SAMHD1 tetramerisation in a long lived active state.

It has been shown that T592 phosphorylated SAMHD1 was detectable by western blot in extracts from S/G2/M-phase but not G1-phase cells (Yan et al., 2015). *In-vitro* experiments with recombinant proteins demonstrated T592 phosphorylation by

Cyclin A2/CDK1 (Pauls et al., 2014b). Cyclin A2/CDK1 associates with the c-terminus of SAMHD1 and phosphorylates it at T592 (Cribier et al., 2013; St Gelais et al., 2014; Yan et al., 2015). When the cyclin A2/CDK binding site in the c-terminus was abolished, SAMHD1 was not phosphorylated at T592 (Yan et al., 2015). As Cyclin A2 is synthesized at the onset of S-phase, this suggests that the *in-vivo* phosphorylation of SAMHD1 occurs at entry into S-phase. Studies reveal the equivalent phosphorylation site to be T603 in mice (Wittmann et al., 2015). Similar to human SAMHD1, activity in mice is regulated by T603 phosphorylation. Given that SAMHD1 activity is regulated to G1-phase cells, while RNR is regulated to post-G1-phase, SAMHD1 activity has a greater effect on dNTP pools in G1.

CDK2 and p21 have also been shown to be regulators of SAMHD1 phosphorylation, with CDK2 expression triggering S-phase (Pauls et al., 2014c), while CDK6 is an upstream regulator of CDK2 (Pauls et al., 2014a; Pauls et al., 2014b). siRNA mediated knockdown of CDK2 and CDK6 were shown to decrease SAMHD1 phosphorylation (Pauls et al., 2014b). Pan-CDK inhibitors have been shown to decrease dNTP concentrations in monocyte-derived macrophages (Pauls et al., 2014b).

SAMHD1 has nuclease activity in addition to dNTPase activity

SAMHD1 can bind to single stranded nucleic acids. Using fluorescence cross correlation spectroscopy, SAMHD1 was shown to interact with ssDNA and ssRNA (Tungler et al., 2013). Also, incubating immobilized nucleic acids with cell extracts also showed that SAMHD1 binds these substances (Goncalves et al., 2012). It was noted also that while tetrameric SAMHD1 is a dNTPase, monomeric SAMHD1 binds ssRNA

(Seamon et al., 2015), while other data suggests the phosphomimetic mutant of SAMHD1 T592E has decreased RNAse activity (Koharudin et al., 2014).

Studies showed *in-vitro* 3' -> 5' exonuclease activity against ssDNA, ssRNA and RNA in DNA/RNA duplexes using full length recombinant SAMHD1 and ^{32}P-ssRNA and ssRNA as substrates with the addition of magnesium (Beloglazova et al., 2013). Other studies confirm the degradation of ssRNAs, however without activity on ssDNA or RNA in RNA:DNA hybrids (Ryoo et al., 2014).

SAMHD1 is an HIV restriction factor

SAMHD1 was found to be an HIV restriction factor while looking for a binding partner to the HIV-2 accessory protein Vpx (Laguette et al., 2011). Vpx was initially discovered in 1989 as an open reading frame in HIV-2 and SIV, which was not present in HIV-1. This accessory protein was dispensable when infecting lymphocyte cell lines, as *Vpx* mutants were still infectious. These viruses, however, had a severe defect of infectivity in peripheral blood lymphocytes, macrophages and dendritic cells (Goujon et al., 2007; Guyader et al., 1989; Yu et al., 1991). Vpx, which is not usually expressed in HIV-1, was also found to enhance HIV-1 infection in dendritic and myeloid cells (Goujon et al., 2007). Vpx was later found to load human SAMHD1 (but not mouse SAMHD1) onto CRL4^{DCAF1} E3 ubiquitin ligase (Ahn et al., 2012; Hrecka et al., 2011), leading to highly efficient proteasome-dependent degradation of SAMHD1.

It was also found that SAMHD1 was expressed in cells resistant to HIV-1 infection, which included THP-1 cells, primary monocytes, monocyte derived macrophages,

dendritic cells and resting CD4+ T cells. Active CD4+ T cells and U937 cells, which are susceptible to HIV-1, did not express SAMHD1 (St Gelais & Wu, 2011). Subsequent experiments showed that *SAMHD1* knockdown by shRNAs in THP-1 cell lines and resting CD4+ T cells significantly increased susceptibility HIV-1 infection. SAMHD1 was subsequently labelled as an HIV-1 restriction factor (Baldauf et al., 2012; Laguette et al., 2011).

Multiple HIV restriction factors exist in addition to SAMHD1, and include apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3G (APOBEC3G), which is related to AID, Tripatite Motif 5 alpha (TRIM5a) and BST-2/Tetherin (Cribier et al., 2013; St Gelais & Wu, 2011). These cellular factors "mediate a cell intrinsic resistance that arrest the viral life cycle at a specific step". The following table demonstrates the characteristics of various HIV-1 restriction factors (Sze et al., 2013):

Restriction factor	Function	Retroviral countermeasure
APOBEC-3G	Deaminates cytidine to uracil, leading to G-to-A hypermutations	HIV-1: Vif HTVL-1: nuclear capsid
TRIM5α	Targets retroviral capsids for proteasomal degradation	Capsid proteins cannot be bound by human TRIM5α
Tetherin	Cross-links new HIV-1 particles to the cell surface, preventing release	HIV-1: Vpu HTLV-1: cell-to-cell transmission strategy
SAMHD1	Hydrolyzes host dNTPs that are required for reverse transcription	HIV-2: Vpx

Table 1.1: Known HIV restriction factors, function in HIV restriction and retroviral countermeasures.

The defining characteristics of a restriction factor are the following (Ballana & Este, 2015):

1. A restriction factor must directly and dominantly cause a significant decrease in virus infectivity.
2. If a restriction factor is a true threat to viral replication, then the predecessors of HIV invariably evolved an equally potent counter-restriction mechanism that still exists in the present-day virus.
3. Because the interactions between restriction and counter-restriction factors occur through direct protein-protein interactions, the restriction factor often shows signatures of rapid evolution.
4. The expression of each restriction factor is often hard wired to the innate immune response.

SAMHD1 fulfills all these requirements.

As stated previously, the resistance mechanism of Vpx is to load SAMHD1 onto $CRL4^{DCAF1}$ E3 ubiquitin ligase (Ahn et al., 2012; Hrecka et al., 2011), and this leads to highly efficient proteasome-dependent degradation. This mechanism of ubiquitin ligase recruitment is similar to other non-structural accessory proteins. Vif promotes degradation of APOBEC3G by interacting with E3 ubiquitin ligase complex consisting of cullin 5, elongin B and elongin C. Vpu and Nef degrade Tetherin by recruiting another ubiquitin ligase, the SKP-cullin 1-βTRCP complex (Ayinde et al., 2012).

SAMHD1 has been shown to restrict multiple retroviruses, not just HIV-1 in non-cycling cells. Alpha, beta and gamma retroviruses are also restricted by SAMHD1 in myeloid cells (Gramberg et al., 2013), as is Hepatitis B virus in hepatocytes (Chen et al., 2014), Herpes Simplex virus 1 in macrophages (Kim et al., 2013) and vaccinia (Hollenbaugh et al., 2013).

Mutations in SAMHD1 are responsible for AGS5, a subset of Aicardi Goutières syndrome.

Aicardi Goutières syndrome (AGS) is a rare disease with the clinical manifestations of congenital encephalopathy, which leads to severe intellectual disability and motor deficits including tetraplegia, pyramidal and extrapyramidal signs, abnormal eye movements, nystagmus and poor visual performance (Crow, 1993; Orcesi et al., 2009). AGS also manifests with clinical signs consistent with systemic lupus erythematosus (SLE), including recurrent fevers, inflammatory arthropathy and chilblains (Kretschmer et al., 2015; Ramesh et al., 2010). Mutations in multiple genes have been found in AGS, including *SAMHD1*. Patients with *SAMHD1* mutation also manifest with cerebral arteriopathy on top AGS manifestations, and present with both cerebral vascular occlusion and haemorrhage and also peripheral vascular disease. As expected, resting CD4+ T cells from AGS patients are permissive to HIV-1 infection (Berger et al., 2011).

AGS patients have increased IFN-α levels regardless of the mutation conferring AGS (Dussaix et al., 1985; Goldstone et al., 2011). The manifestations of AGS mimic congenital viral infection and also SLE (Rice et al., 2009) as these diseases all display increased production of interferon alpha (IFN-α) (Dussaix et al., 1985). In viral

infection, IFN-α is produced to elicit antiviral activity from target cells and induce apoptosis in virus-infected cells (Taniguchi & Takaoka, 2002). This involves nucleic acid-binding PRR including toll-like receptors, RNA and DNA receptors (Crow & Manel, 2015). In SLE, IFN-α production is also characteristic of the disease (Niewold et al., 2010). Double stranded DNA, RNA and DNA/RNA hybrids are the primary autoimmune targets in SLE and activate PRR. This induces production of IFN-α, which activates the adaptive immune system by promoting dendritic cell activation of T cells, and also enhances proliferation of primary B cells.

AGS is associated with autosomal recessive mutations in nuclease genes including *TREX1* and *RNASEH2*. Mutations in these genes lead to failure to clear nucleic acid debris and subsequent activation of the innate immune system (Rice et al., 2009).

This table summarizes the known mutations associated with AGS (Crow & Rehwinkel, 2009; Leshinsky-Silver et al., 2011):

AGS variant	Gene	Locus	Proportion of AGS	Protein Function
AGS1	*TREX1/DNaseIII*	3p21	35	3' -> 5' exonuclease
AGS2	*RNASEH2B*	13q14	45	Cleaves RNA of RNA/DNA hybrids
AGS3	*RNASEH2C*	11q13	15	Cleaves RNA of RNA/DNA hybrids
AGS4	*RNASEH2A*	19p13	<5	Cleaves RNA of RNA/DNA hybrids
AGS5	*SAMHD1*	20q11	10	dNTPase/single stranded ribonuclease

Table 1.2: Aicardi Goutières Syndrome variants, genetic aetiology and protein function.

The role of SAMHD1 in malignancy

The most substantial research in the role of SAMHD1 in malignancy has been in chronic lymphocytic leukaemia (CLL). *SAMHD1* mutation frequency was found to be 5% of CLL patients (Clifford et al., 2014). A higher proportion of patients with relapsed/refractory CLL have *SAMHD1* mutations (11% vs 3%). Other cancers implicated with *SAMHD1* mutation include pancreatic, lung, glioblastoma multiforme, breast and colorectal cancers (Kohnken et al., 2015).

One study showed that SAMHD1 deficiency causes genomic instability in fibroblasts of patients with AGS (Kretschmer et al., 2015), with increased comet tail length in the absence of exogenous genotoxic stress. It is now clear that SAMHD1 is involved in the DNA damage response, especially DNA double strand break repair (Clifford et al., 2014). SAMHD1 has been found to reside in a high molecular weight complex in untreated cells and a low molecular weight complex after treatment with etoposide. SAMHD1 also co-localises with 53BP1 foci in camptothecin treated and untreated cells. CtIP is an endonuclease which processes the 5' ends of DSBs during the initial stages of HR. Independent experiments showed that SAMHD1 complexes with CtIP to facilitate HR (Daddacha et al., 2017).

SAMHD1 in mouse

Phenotypically, *SAMHD1* knockout mice (produced by gene-trapping in embryonic stem cells) are fertile, have no macroscopic abnormalities and were also healthy beyond the age of 96 weeks. Their cells have 40-fold elevated dNTP levels (Behrendt et al., 2013).

Compared to human, murine SAMHD1 has some similarities and differences.

Similarities in human and mouse SAMHD1 include (Behrendt et al., 2013; Rehwinkel et al., 2013):

1. Murine SAMHD1 also restricts HIV replication *in-vitro*, however SAMHD1 competent mouse cells have a higher concentration of dNTPs than human cells. dNTP concentrations in mice are higher than the K_M of HIV reverse transcriptase, and so *in-vivo* HIV restriction is not present in wild type mice, except for HIV mutants with reverse transcriptase that has a lower affinity for dNTPs.
2. *SAMHD1* knockout mice also have activation of IFN-α.
3. *SAMHD1* knockout mice also have elevated dNTP concentrations in bone marrow derived dendritic cells and macrophages, splenic B and T cells and mouse embryonic fibroblasts (Behrendt et al., 2013; Wu, 2013b).

Differences in human and mouse SAMHD1 include:

1. Vpx does not lead to murine SAMHD1 degradation (Lahouassa et al., 2012).

Aims:

The first aim of this research project is to understand why a deficiency in MMR leads to an increase in focusing of mutation in AID hotspots.

Specifically, this involves:

1. Generation of large mutation datasets using the SW_{HEL} adoptive transfer model (Sharbeen et al., 2010; Sharbeen et al., 2012) in the following groups:
    i. SW_{HEL} B cells.
    ii. SW_{HEL} B cells deficient for UNG2 *via* retroviral expression of the uracil glycosylase inhibitor (ugi).
    iii. SW_{HEL} B cells deficient for MutSα by crossing $Msh2^{ko}$ alleles into SW_{HEL} mice.
    iv. SW_{HEL} B cells deficient for both UNG2 and MutSα.
2. Analysis of the mutation spectrum by sequence context by comparing the above groups.

<u>Hypothesis:</u>

Downstream processing of AID-induced deamination by UNG2 BER along the *Ig* gene is dependent on sequence context of surrounding bases.

The second aim of this research project is to determine the effect of increasing dNTP pools on somatic hypermutation by knockdown and knockout of *Samhd1,* in combination with retroviral manipulation of R2 levels.

Specifically, this involves:

Samhd1 knockdown by retroviral expression of shRNAs against *Samhd1* and measurement of effect on somatic hypermutation:

1. Development of shRNA to inhibit *Samhd1* expression.
2. Creation of retroviral reagents containing these sequences.
3. Determination of efficacy of knock-down of each shRNA by western blot analysis of transduced mouse fibroblast cell line 3T3 cells.
4. Measurement of the effects of *Samhd1* knockdown on somatic hypermutation by the SW_{HEL} adoptive transfer model.
5. Analysis of the mutation spectrum compared to controls.

Creation of homozygous *Samhd1* knockout mice and measurement of the effect on somatic hypermutation (CRISPR/Cas9 frameshift deletions in *Samhd1* were created by Robert Brink, Garvan Institute, Sydney):

1. Breeding of founder mice to achieve homozygous *Samhd1* frameshift deleted mice.
2. Measurement of the expression of SAMHD1 in homozygous *Samhd1* mutants by western blot.
3. Measurement of effect on somatic hypermutation by adoptive transfer model.
4. Analysis of mutation spectrum compared to controls.

Hypothesis:

Polymerase η recruitment during somatic hypermutation is caused by nucleotide paucity during G1 cell cycle phase, due to stalling of DNA replication complexes during AID-induced MMR.

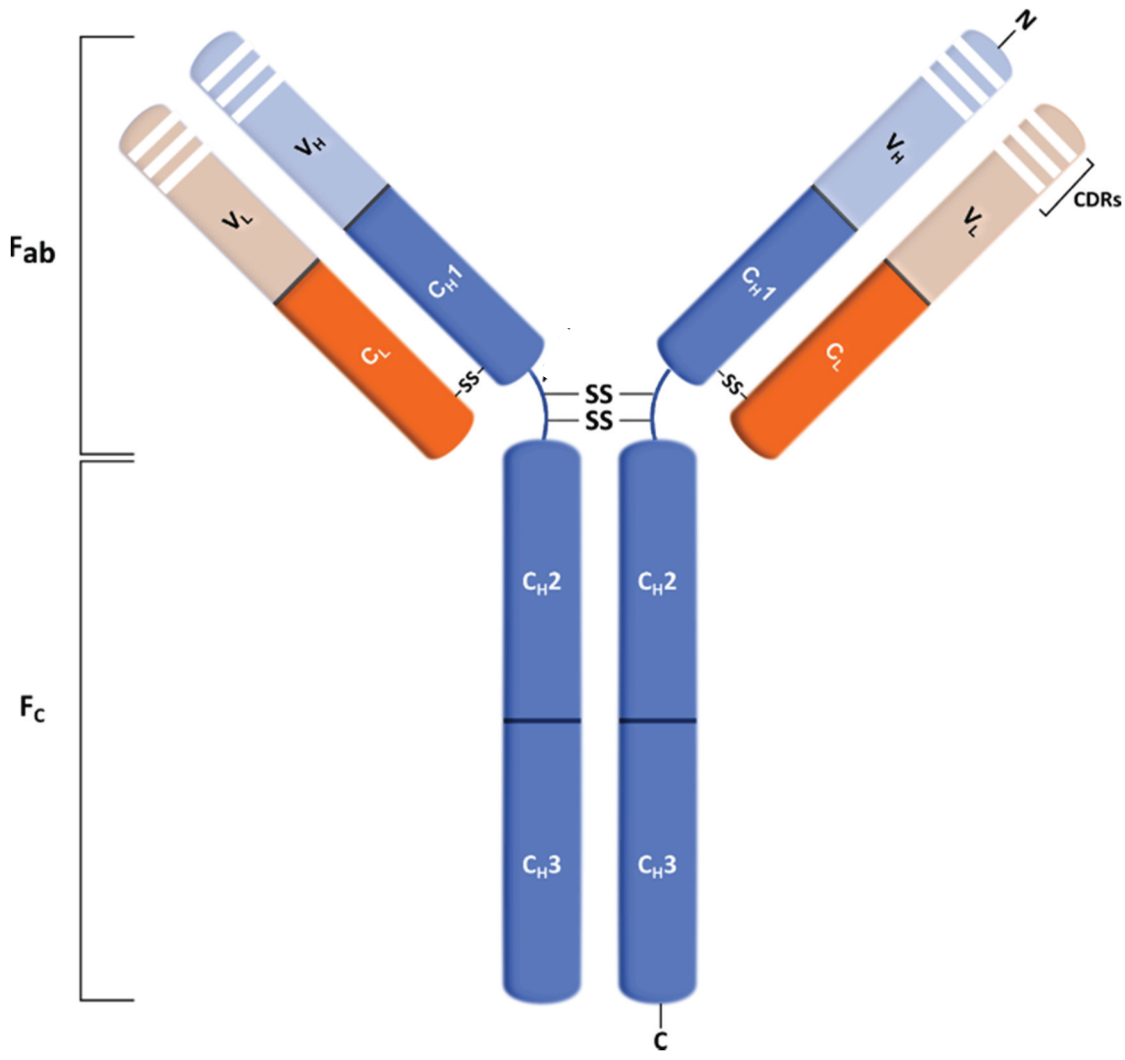


Figure 1.1: Schematic representation of an antibody. Each antibody consists of two types of polypeptide chains: two identical heavy (**H**) chains (**blue**) and two identical light (**L**) chains (**orange**) linked together by disulfide bonds (**SS**). The heavy chain is composed of three to four constant domains (**C_H1-C_H3**) and one variable domain (**V_H**). The light chain is composed of one constant domain (**C_L**) and one variable domain (**V_L**). The **Fab** consists of **V_L**, **C_L**, **V_H** and **C_H1**, while **Fc** consists of **C_H2 to C_H4**. Complementary determing regions **(CDRs)** are located within the V regions (from Loureiro *et al.* 2015).

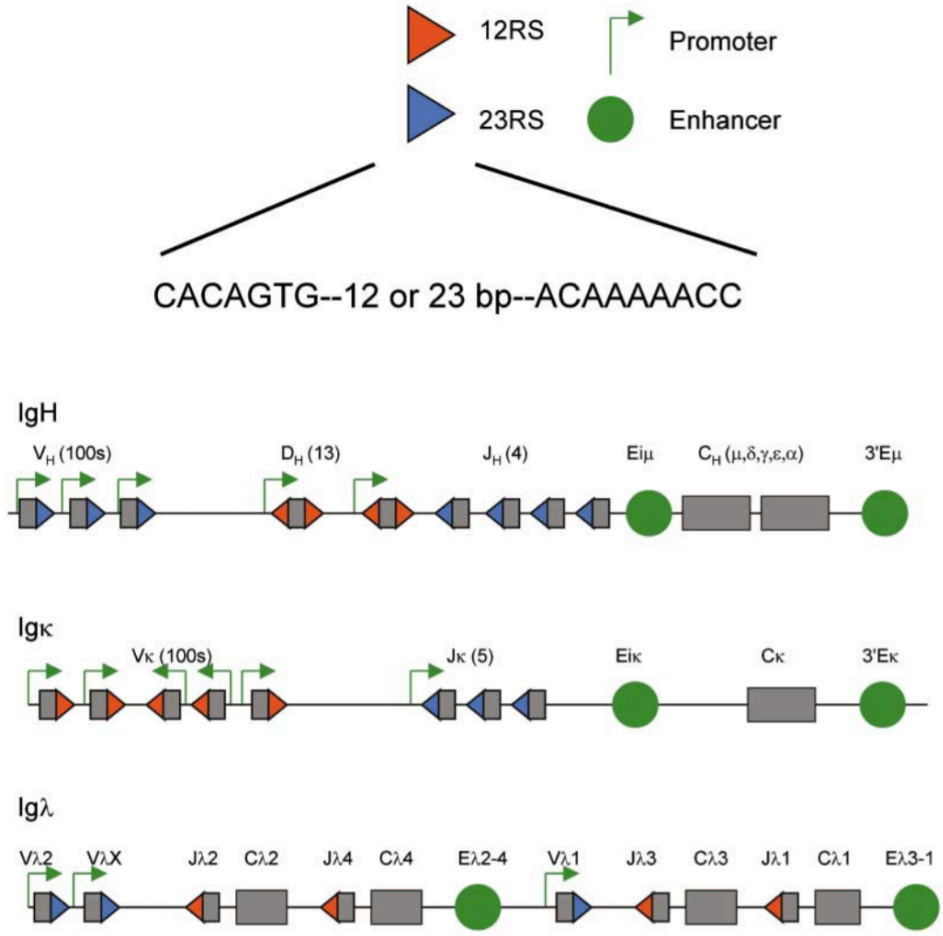


Figure 1.2: Genomic structures of the antigen receptors of IgH and IgL loci in mice. Functional V, D, and J gene segments are flanked by non-coding recombination signal sequences (**RSS**). V_H segments are 3'-flanked by RSS with 23-bp spacers, while the D_H segments are flanked at both the 3' and 5' ends by RSS with 12-bp spacers. Like the 3' end of V_H, the 5' end of J_H is also flanked by RSS with 23-bp spacers. V_K segments are 3'-flanked by RSS with 12-bp spacers. J_K are flanked at their 5' ends by RSS with 23-bp spacers. V_λ segments are flanked by RSS with 23-bp spacers at their 3' ends and J_λ are 5'-flanked by RSS with 12-bp spacers (from Jung *et al.* 2004).

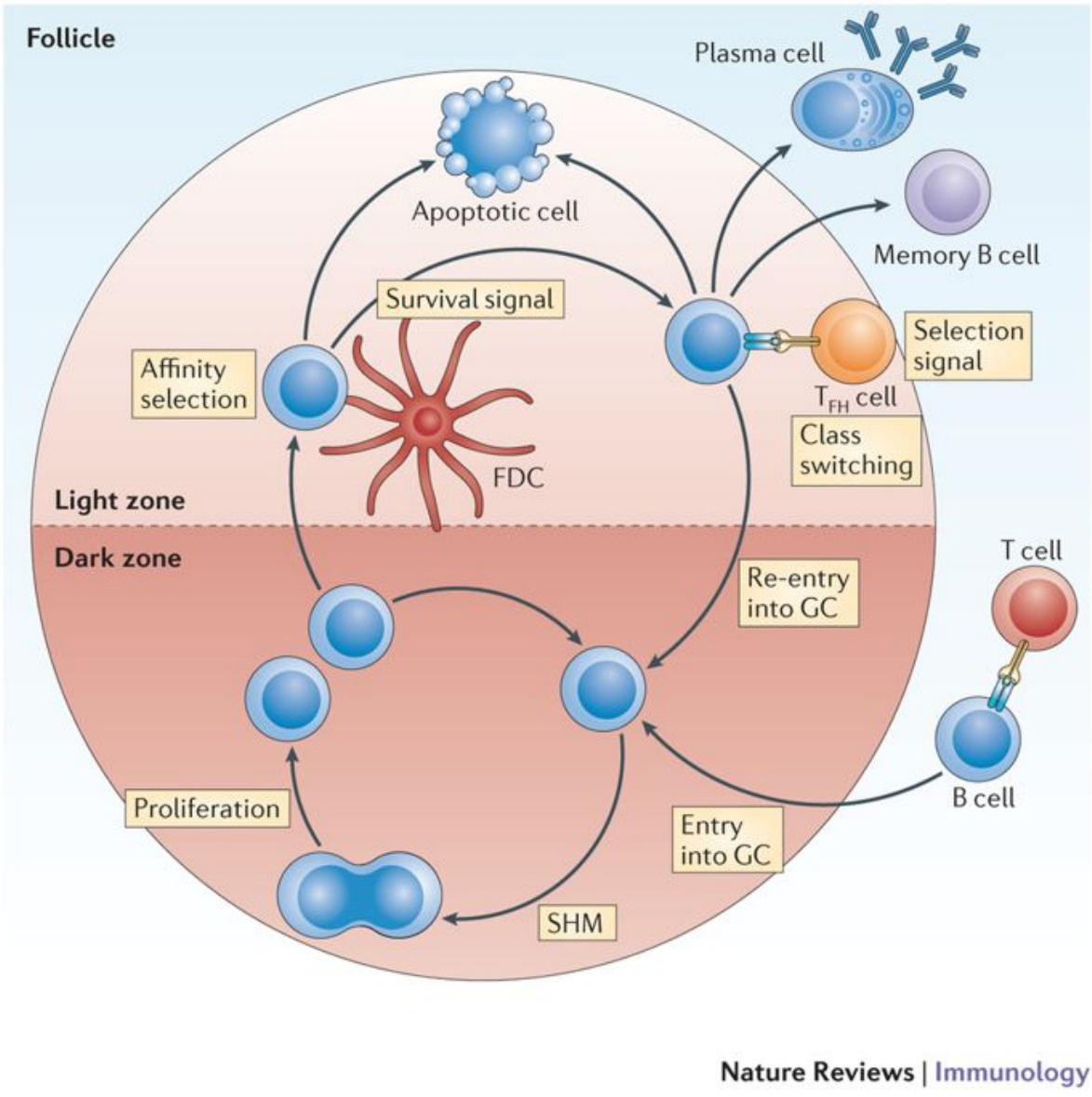


Figure 1.3: Visual representation of the germinal centre in the lymph node. Following secondary activation by cognate **T cells**, **B cells** proliferate rapidly to form germinal centres (large circle). Germinal centres have two different zones. In the **Dark zone,** B cells undergo proliferation and somatic hypermutation, producing daughter clones with antibody variants that bind with different affinities. B cells migrate into the **Light zone** to undergo **affinity selection**, provided by survival signals from follicular dendritic cells (**FDC**) and T-folliculr helper cells (**TFH**) (from Heesters *et al.* 2014).

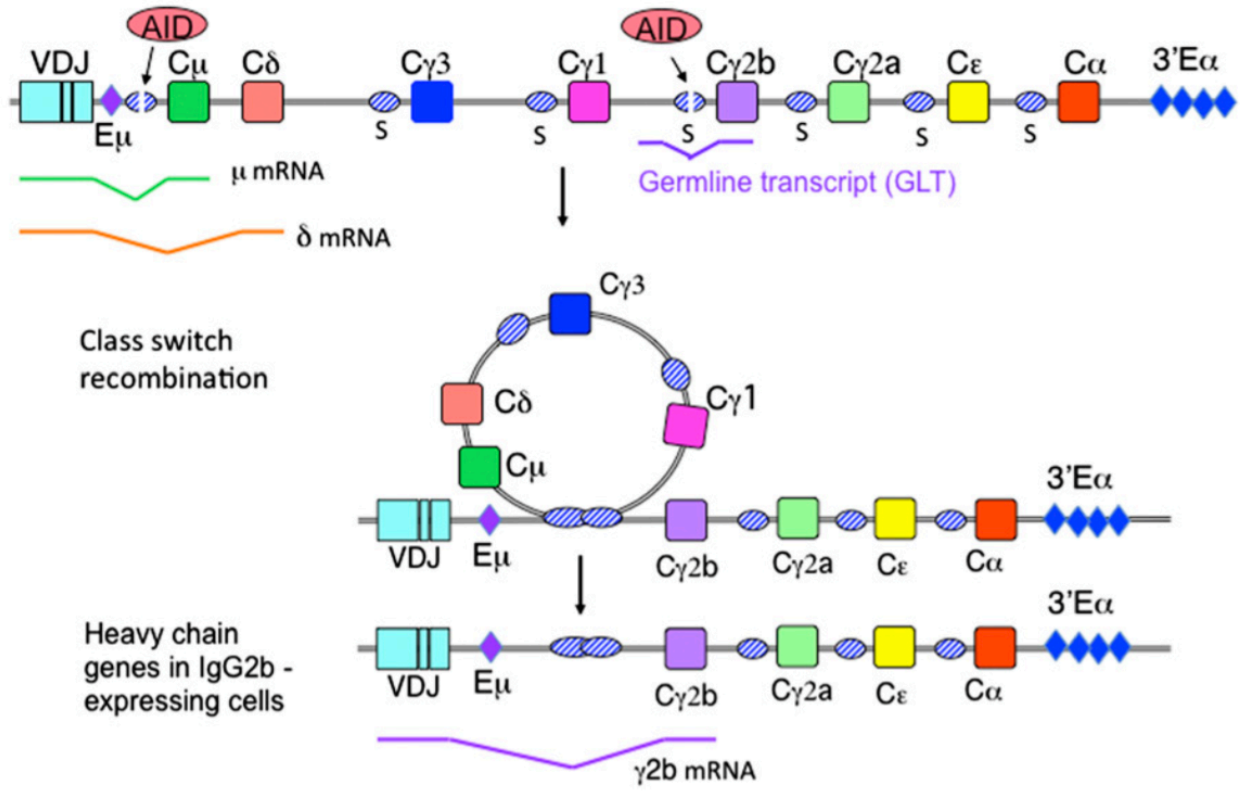


Figure 1.4: Structure of the mouse immunoglobulin heavy chain gene and Class Switch Recombination. In the germline, exons encoding IgM (***Cμ***) and IgD (***Cə***) occur 3' to the J_H gene segments. Either ***Cμ*** or ***Cə*** exons are spliced to the VDJ_H exon during transcription in naïve B cells, such that mature naïve B cells express both IgM and IgD. Downstream of ***Cə*** are additional constant region exons which encode Fc distinct from IgM and IgD that are not used by naïve B cells. Located upstream (i.e. 5') of every constant region gene lies a non-coding switch (***S***) region sequence. Class switch recombination is a process of deletional recombination between the *Sμ* switch region and a downstream ***S*** region, the example here is to **IgG2b** (from Stavnezer *et al.* 2014).

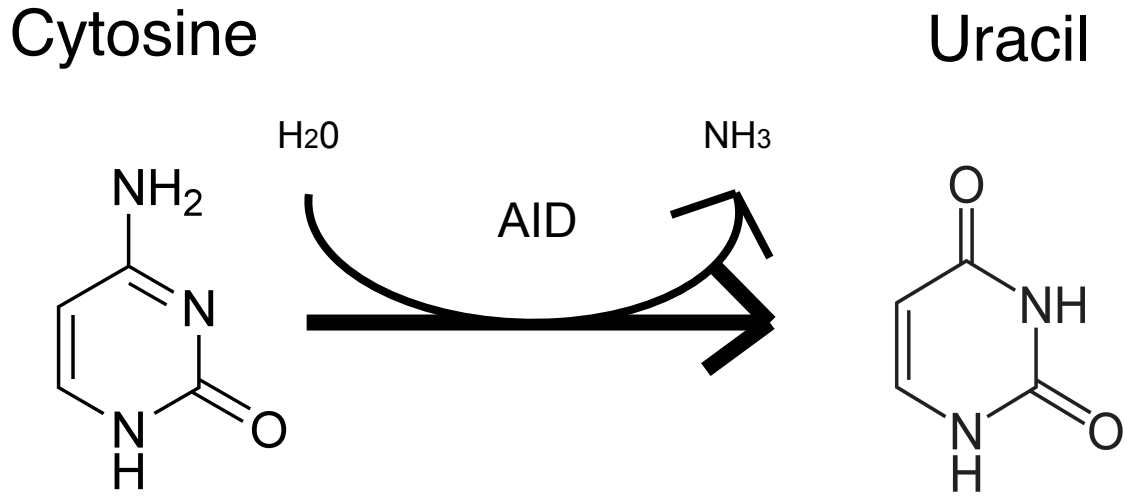


Figure 1.5: Diagram of the deamination of cytosine to uracil by Activation induced cytidine deaminase (AID).

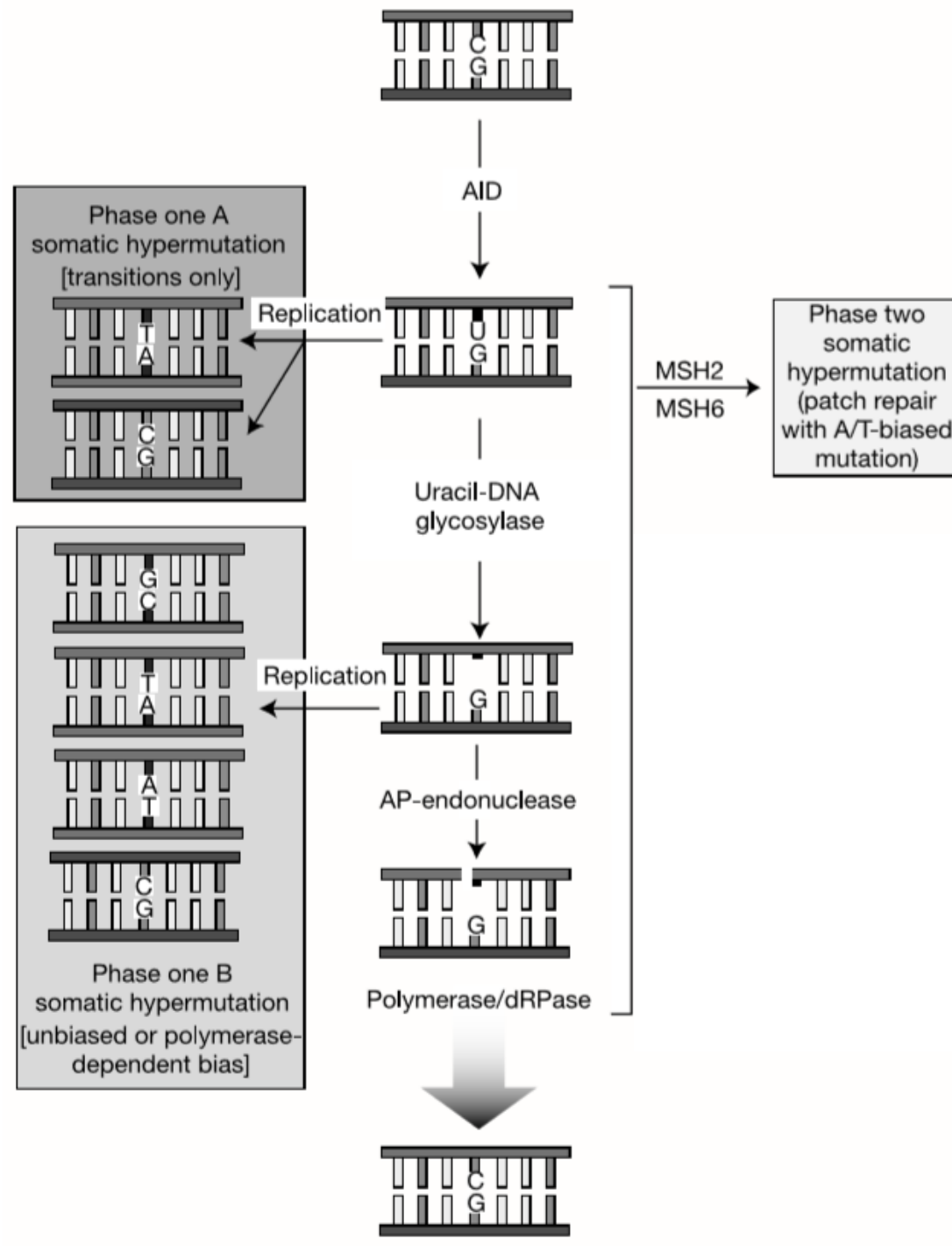


Figure 1.6: The Neuberger model of Somatic Hypermutation. Activation induced cytidine deaminase (**AID**) initiates somatic hypermutation by de-aminating genomic C residues to U. Without further processing, replication leads to transition mutations (**Phase one A**). Processing of U:G mismatches by the mismatch repair pathway (**Msh2/Msh6)** introduces mutations at surrounding A and T bases (**Phase two**). Processing of U:G mismatches by the base excision repair protein Uracil-DNA glycosylase (UNG2) can repair the initial C to U mutation but also introduces transversion mutations (**Phase one B**) (from Petersen-Mahrt *et al.* 2002).

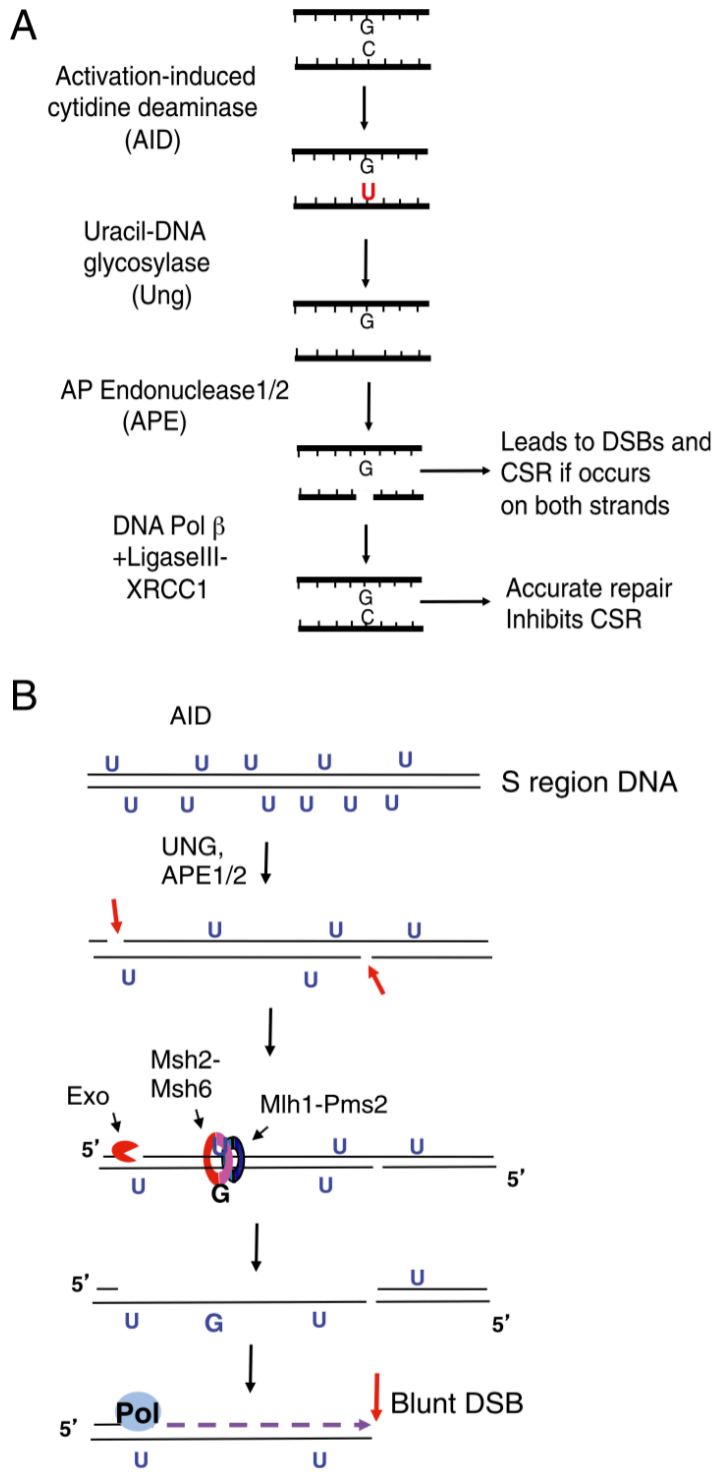


Figure 1.7: The model of double strand break formation (DSB) in Class Switch Recombination. A) Mismatch repair independent (MMR) DSB formation occurs when **Uracil-DNA glycosylase (UNG)** converts Activation induced cytidine deaminase (**AID**) induced deaminations into apyrimidinic (**AP**) sites. **AP endonuclease 1 and/or 2 (APE)** cut these AP sites 5', creating a single stranded DNA (ssDNA) nick. When two ssDNA nicks occur in close proximity on opposite strands, this creates a **DSB**. B) MMR dependent **DSB** formation occurs when ssDNA nicks are too far apart to generate a **DSB**. MMR can convert the ssDNA nick to a **DSB**, provided some U:G mismatches able to activate **Msh2/Msh6** and **Mlh1/Pms2** are left in place by UNG (from Stavnezer 2014).

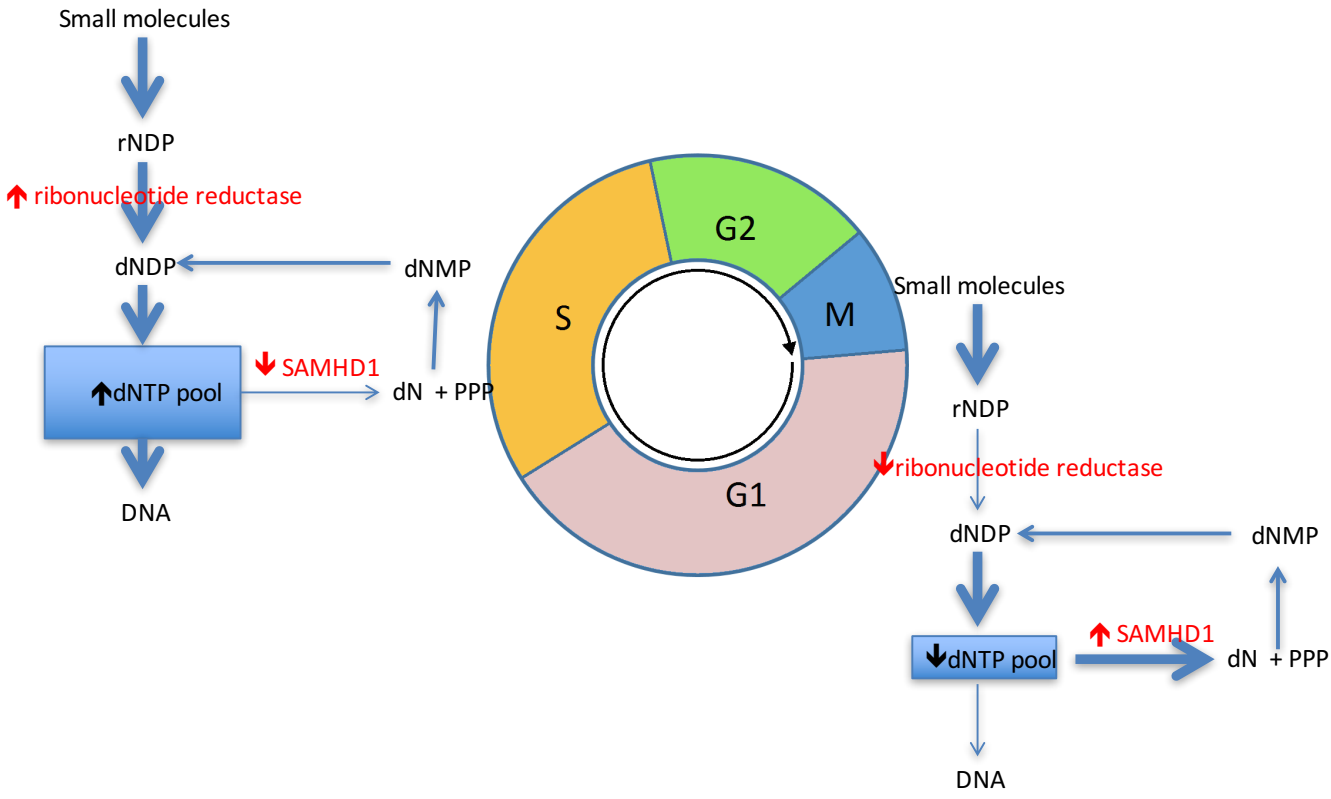


Figure 1.8: Diagram of the regulation of dNTP pools in different phases of the cell cycle. The synthesis of **dNTPs** is dependent on **ribonucleotide reductase** (RNR) while degradation is dependent on the dNTPas activity of **SAMHD1**. During **S** phase, **RNR l**evels are high, while **SAMHD1's** dNTPase is inactive. In **G1** phase, RNR is low, while **SAMHD1's** dNTPase is active. This leads to higher **dNTP** levels in **S** phase versus **G1** phase.

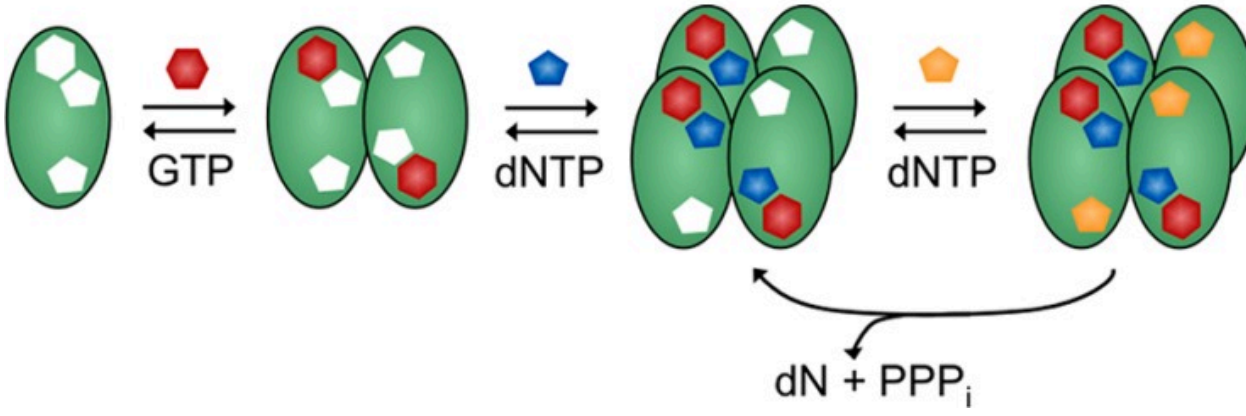


Figure 1.9: Diagram of SAMHD1 tetramer formation. SAMHD1's dNTPase is active in a tetramer state and inactive in a monomer or dimer state. dGTP/GTP (**red**) and dNTPS (**blue**) bind to activator sites, leading to formation of the SAMHD1 tetramer. This induces a conformational change int the dNTP binding pocket to yield the catalytically active enzyme that cleaves dNTPs into nucleosides (**dN**) plus triphosphate (**PPPi**) (from Seamon *et al.* 2014).

2. Materials and methods

Mice

SW_{HEL} rag1$^{ko/ko}$ mice carry a gene-targeted *VDJ_H*-rearrangement (SW3) and a transgenic Ig_K gene, both from the HyHEL-10 hybridoma, together conferring Ig specificity for hen egg lysozyme (HEL). Inactivation of *rag1* or *rag2* in *SW_{HEL}* mice prevents V(D)J receptor editing, ensuring that all B cells express identical HEL-specific receptors (Cook et al., 2003; Phan et al., 2003).

Samhd1$^{\Lambda/\Lambda}$ SW_{HEL} rag1$^{ko/ko}$ mice were constructed by Australian BioResources by targeting frameshift deletion to exon 1 of *Samhd1* by CRISPR/Cas9 in mouse embryos on a *SW_{HEL}* background. 25, 41, 55 and 74 base pair deletions in Exon 1 were obtained and confirmed by PCR. Mutant alleles were bred to homozygosity after back-crossing founders to congenic *SW_{HEL}* mice.

Male C57BL/6 adoptive transfer experiment host mice were purchased from Animal Resources Centre (Canning Vale, Western Australia) and Australian BioResources (Moss Vale, New South Wales) and were used in experiments when 8–12 weeks old.

SW_{HEL} rag1$^{ko/ko}$ msh2$^{ko/ko}$ mice were created by inter-crossing *msh2$^{ko/ko}$* and *SW_{HEL} rag1$^{ko/ko}$* mice (de Wind et al., 1995; Sharbeen et al., 2010).

Fucci transgenic mice have fused red and green-emitting fluorescent protein to E3 ligase substrates. *mKO* is fused to the post G1 degron portion of *cdt* (amino acids 30-120) and therefore labels cells in G1. *mAG* is fused to the G1 degron portion of *geminin* (last 100 amino acids) and therefore labels cells in post G1 (Sakaue-Sawano et al., 2008)

Mice were bred and maintained under SPF conditions in the Centenary Animal Facility. Mouse experiments were approved and monitored by the University of Sydney Animal Ethics committee in accordance with the New South Wales Animal Research Act (1985).

Materials

Cell lines

Cell line	Source	Culture medium
NIH/3T3	ATCC	HEK
Platinum E	Cell Biolabs, Inc.	HEK
A20	ATCC	10FR
HEK293	ATCC	HEK

Table 2.1: Cell lines used in experiments.

Plasmids

Name	Gene	Vector	Source
GS015	*ugi-GFP*	pMiG	George Sharbeen
DB04	*R2-IRES-GFP*	pMiG	Daniel Bosnjak
DB05	*R2^{AAN}-IRES-GFP*	pMiG	Daniel Bosnjak
mCherryLuc	*mCherry-Luciferase*	pHIV	Michelle Van Geldermalsen

V057	*GFP*	pMiG	commercial
GS026	*mKO2*	pMiG	George Sharbeen
Samhd1 shRNA 1-15	*mIR30-shRNA*	pLMP	created

Table 2.2: Plasmids used in experiments.

Recipes

B cell medium:

Constituent	concentration
L-glutamine	2mM
Streptomycin	100ug/mL
Penicillin	100U/mL
Non-essential amino acids	0.1mM
2-mercaptoethanol	50uM
Sodium pyruvate	1mM
Heat inactivated FCS	10%
in IMDM	

HEK-293 medium:

Constituent	concentration
L-glutamine	2mM
Streptomycin	100ug/mL
Penicillin	100U/mL
Heat inactivated FCS	10%
Sodium pyruvate	1mM
in DMEM	

10FR medium:

Constituent	concentration

L-glutamine	2mM
Streptomycin	100ug/mL
Penicillin	100U/mL
Heat inactivated FCS	10%
in RPMI	

Lysogeny Broth/Luria Bertani medium (LB):

Constituent	concentration
Tryptone	1% (w/v)
Yeast extract	0.5% (w/v)
Sodium chloride	0.1M
in distilled water	

TBST:

Constituent	concentration
Tris pH 7.4	10mM
Sodium chloride	0.9% (w/v)
Tween 20	0.05% (v/v)
in distilled water	

Single cell sort buffer:

Constituent	concentration
Proteinase K	0.5mg/mL
EDTA	0.1mM
Taq buffer	1x
Tween 20	0.1%
in MilliQ water	

Tris-acetate-EDTA electrophoresis buffer (TAE):

Constituent	concentration
Tris pH 7.4	0.04M
Acetic acid	20mM
in distilled water	

Laemmli sample buffer:

Constituent	concentration
SDS	2% (w/v)
Glycerol	10% (v/v)
Tris pH 7.4	60mM
Bromophenol blue	0.01% (w/v)
in distilled water	

Bjerrum Schafer-Nielsen buffer:

Constituent	concentration
Tris pH 7.4	48mM
Glycine	39mM
Ethanol	15% v/v
in distilled water	

Conjugation buffer:

Constituent	concentration
Mannitol	0.35M
Sodium chloride	0.001M
in distilled water	

passed through 0.22um filter.

Cell cycle staining solution

Constituent	concentration
DAPI	1μg/mL
NP-40	0.1% (v/v)
BSA	20μg/mL
EDTA	2mM
in PBS	

Methods:

Production of retroviral supernatants

Ecotropic viral supernatants were produced by calcium phosphate transfection of the Platinum-E retroviral packaging cell line (Cell Biolabs, Inc.). Culture medium (DMEM, 10% fetal calf serum, 1mM sodium pyruvate, 2mM L-alanyl-L-glutamine dipeptide, penicillin/streptomycin) was replaced at 12 hours, then retroviral supernatants were harvested at 60 hours, 0.45 μm filtered, snap frozen in liquid nitrogen and stored at -70 °C.

Transduction of NIH-3T3 cells

Log phase mouse immortalized NIH-3T3 fibroblasts were transduced by adding thawed retroviral supernatants supplemented with 8μg/mL polybrene to HEK medium.

Adoptive Transfer of splenocytes:

Mouse spleens were harvested from $C0_2$ euthanased mice into 5mL B cell medium, passed through 70μm sieve and washed with a further 5mL B cell medium. These samples were centrifuged at 500g 4 degrees Celcius 5 minutes. The supernatant was removed, and the cells were resuspended in 4mL B cell medium. Splenocytes were isolated from red blood cells by using Histopaque 1083 (Sigma Aldrich). Cells were then suspended in B cell medium with recombinant CD40L-FLAG fusion protein and anti-FLAG antibody (4μg/mL clone M2) (Invitrogen) for 24 hours and then transduced by centrifugation with viral supernatant with 16μg/mL polybrene (1100g, 60min, 32 °C). Splenocytes were then washed and incubated for 48 hours. 10^4 GFP+ cells were sorted using BD FACS Aria II or BD influx, combined with 10^8 sheep red blood cells (SRBC) (Applied Biological Products) conjugated to HEL in B cell medium and tail vein injected into 8-12-week hosts previously intraperitonally injected with 10^8 SRBC in PBS 7 days prior. 6 days later host spleens were harvested, splenocytes isolated by Histopaque 1083 and sorted on BD FACS Aria II or BD influx for GFP^+ and HEL^+ one cell per well in 96 well plate containing 15 μl/well of ice-cold single cell sort buffer. Collection of mutation data followed the methods from (Sharbeen et al., 2010) and statistical analysis of mutation data followed methods from (Sharbeen et al., 2012).

Conjugation of hen egg lysozyme to sheep red blood cells

5mL of SRBC in Alsever's solution was transferred to 50mL centrifuge tube, and 4°C sterile PBS was added to 50ml. This was spun at 2300 rpm (1111 g) for 5 minutes at 4°C without brakes. and repeated twice. This was resuspended with 20 ml of 4°C conjugation buffer and centrifuged at 2200 rpm for 5 min at 4°C without brakes. This

was resuspended in 5.0 ml 4°C conjugation buffer and split SRBC into two 50ml centifuge tubes: 3mL for normal conjugation, 2mL for mock. 100ug of HEL was added to normal conjugation tube and incubated on the rocker on ice for 10 minutes. 0.05g of EDCI was added to both tubes and incubated on rocker for 30 minutes.

After this 4°C sterile PBS was added to a total of 50mL in both tubes and centrifuged at 2200 rpm for 5 min at 4°C without brakes. This was repeated thrice.

To confirm successful conjugation, 10ul of mock & conjugated SRBC were combined with 200uL HyHEL9 hybridoma supernatant for 30 minutes on ice, washed with 2mL PBS with 2% albumin (PBA), and centrifuged at 1111 g for 5 minutes at 4°C without brakes, and further stained with anti-IgK biotin (1:100) on ice 20 min, then washed with 2mL PBA and centrifuged at 1111 g for 5 minutes at 4°C without brakes, and stained with streptavidin-FITC (1:1000) on ice 20 min, and washed with 2mL PBA Spin at 2300 rpm (1111 g) for 5 minutes at 4°C without brakes. The samples were then resuspended in 500uL PBA and analysed on BD Fortessa for change in GFP fluorescence to confirm successful conjugation.

Preparation of cell extracts, western blot analysis:

Total protein extracts from unsorted cells, *Fucci* cells sorted on BD Influx for GFP^+ or $mKO2^+$, or 3T3 G1 restricted $mKO2^+$ transduced cells were prepared by suspending cell pellets $2x10^7$ cells /mL in Laemmli sample buffer and incubated for 30min at 20°C with 50U Benzonase (Novagen) per sample. Extracts were electrophoresed on 4-12% Bolt Bis-Tris (+MES buffer) polyacrylamide gels (Invitrogen) and transferred to

nitrocellulose membranes in Bjerrum Schafer-Nielsen buffer on Novex Semi dry blotter (Invitrogen). Membranes were stained with Memcode total protein stain (Invitrogen) as per manufacturer's protocol and imaged using a ChemiDoc MP (Bio-Rad). Membranes were blocked with 5% (w/v) skim milk powder dissolved in TBST were incubated overnight at 4°C with mouse anti-SAMHD1 [1A1] antibody (Abcam), goat anti R2 antibody (e-16) (Santa Cruz Biotechnology), or rabbit anti-RRM1 [epr8483] (ab137114) (Abcam) in 5% skim milk TBST. Washed membranes were incubated with appropriate secondary antibody conjugated to horseradish peroxidase (Santa Cruz Biotechnology) for 1h at 20°C, then washed again and incubated with Supersignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific) and luminescence recorded using a ChemiDoc MP (Bio-Rad).

Retroviral and lentiviral transduction test

Samhd1 ko and control mouse spleens were harvested and splenocytes isolated as per adoptive transfer, but cultured in B cell medium with anti-CD40 HM40-3 (BD Biosciences) at 5μg/mL and IL-4 (BD Biosciences) at 20ng/mL for 24 hours then transduced with pMiG retroviral supernatant method as per adoptive transfer at 5-fold dilutions and analysed for GFP^+ by flow cytometry.

C57/BL6 mice were harvested, splenocytes isolated and cultured as per above method and transduced with R2-IRES-GFP and GFP retroviruses for 48 hours then further transduced with mKO2 retrovirus at 5-fold dilutions and analysed 24 hours later for GFP^+ and $mKO2^+$ by flow cytometry.

Creation and validation of shRNA against SAMHD1

The target for designing shRNAs against mouse Samhd1 was obtained from ensemble.org (mouse SAMHD1 cDNA ENSMUST000000057725). Potential 21-mer shRNA constructs against SAMHD1 mRNA were created as per protocol described in (Dow et al., 2012). Identification of potential candidates used a three-step process. An online siRNA prediction tool 'Designer of Small Interfering RNAs-DSIR' (Vert et al., 2006) identified potential sequences. These sequences were then cross checked against the BLAST database to exclude sequences matching against off-target genes. Sequences were then filtered again through a series of seven Sensor exclusion criteria (Dow et al., 2012). In total 11 shRNA sequences made by this method and 4 commercial shRNA sequences available from Sigma were selected (TRCN0000099800, TRCN0000099801, TRCN0000099803, TRCN0000099804).

97-mer templates were then created from these sequences and synthesized by Integrated DNA Technologies (Coralville, Iowa, USA).

shRNA sequences were then amplified using Phusion HF (Finnzymes) using miR30 5' EcoR1 and 3' Xho1 oligonucleotide adaptors (ACTTAGAAGAATTCCGAGGCAGTAGG and TACAATACTCGAGAAGGTATATTGCTG respectively) in the following PCR reaction in a 200μL well:

Reagent	μL
TDW	32.5
5xPhusion HF buffer	10
10mM dNTPs	1
10pmol/μL miR30 5' EcoR1 oligonucleotide	2.5
10pmol/μL miR30 3' Xho1 oligonucleotide	2.5
0.02ng/μL 97-mer template	1
Phusion HF	0.5
Total	50

The PCR reaction was:

1x:

98 degrees Celcius - 0:30

35x:

98 - 0:10

54 - 0:30

72 - 0:15

1x:

72 - 5:00

PCR products were confirmed on 3% agarose gel.

These products were cloned into a mIR30 site of LMP 1066 (Zhou et al., 2008) (Addgene) vector co-expressing using EcoR1 and Xho1 restriction sites. LMP 1066 is a MSCV based retroviral vector which expresses shRNAmir, with a puromycin resistance cassette and GFP as a marker for retroviral integration.

1. These products were then digested with EcoR1 and Xho1 at 37 degrees Celcius for 1 hour:

Reagent	μL
TDW	2
10X Cutsmart buffer	2
EcoRI HF	0.5
Xho I	0.5
PCR product	15
Total	20

Products were purified using Promega PCR purification kit as per instructions and eluted into 25μL of $T_{10}E_{0.1}$

2. Digestion of pLMP vector containing mIR30 site at 37 degrees Celsius for 1 hour:

Reagent	μL
TDW	80
10X Cutsmart buffer	10
EcoRI HF	2.5
Xho I	2.5
pLMP vector containing mIR30 site (300ng/μL)	5
Total	100

Products were run on 1% Agarose gel,

7880bp band cut and purified using Promega PCR purification kit as per instructions and eluted into 50μL of $T_{10}E_{0.1}$.

3. Ligation of digested PCR product and backbone at RT for 1 hour:

Reagent	μL
10x T4 ligase buffer	1
T4 DNA ligase	1
PCR product	7
Backbone	1
Total	10

Products were then transformed into Top10 competent *E.coli* by putting on ice for 5 minutes, followed by 5μL ligation product added to 50μL vial of Top10 competent E.coli and incubating on ice for 30 minutes. *E.coli* were transformed by putting into 42 degree Celcius water bath for 40 seconds, then transferring on to ice for 5 minutes. 1mL LB medium was then added to vial and put into 37-degree Celcius shaker at 250rpm for one hour. After this, the vial was centrifuged for 3 minutes at 1000g at RT, aspirating 950uL of supernatant and resuspending the pellet in the rest. This suspension was plated onto LB agar with ampicillin, incubated at 37 degrees Celsius overnight. Positive colonies were purified using Promega Miniprep kit (Promega, Madison WI).

Sequences were verified by sequencing of mIR30 containing the shRNA cassette by Macrogen Inc. (Seoul, Korea).

Out of 15 constructs, 14 were successfully cloned into the mIR30 vector. One luciferase control (CJ171) and control (ET06) were also created by the above cloning method.

Ecotropic retroviral supernatants were produced as per Methods: Production of retroviral supernatants.

3T3 cells at 50% confluence in 12 well plates were transduced with 8μL of viral supernatant with 4μg of polybrene in a total volume of 500μL and incubated for 48 hours. Transduction efficiency was 1-2% ensuring single copy transduction.

Cell cultures were then selected with puromycin at a concentration of 1 μg/mL for 48 hours. Dead cells were then washed off and remaining cells re-plated into T25 flasks without puromycin and grown to confluence.

Class switching experiment

Mouse spleens were harvested from $C0_2$ euthanased mice into 5mL B cell medium, passed through 70μm sieve and washed with a further 5mL B cell medium. Samples were centrifuged, the supernatant was aspirated, and the cells were resuspended in 4mL B cell medium. Splenocytes were isolated from red blood cells by using Histopaque 1083. Cells were then suspended in B cell medium with anti-CD40 HM40-3 (BD Biosciences) at 5μg/mL and IL-4 (BD Biosciences) at 20ng/mL for 24 hours and then transduced by centrifugation with viral supernatant with 16μg/mL polybrene (1100g, 60min, 32 °C). Splenocytes were then washed and incubated for 48 hours.

Cells were then fixed, permeabilised and stained for IgG by the following technique applicable to $5x10^6$ cell samples, incubations performed in the dark: Cultures were transferred to 15mL centrifuge tube and then centrifuged at 500g for 5 minutes at RT. The supernatant was aspirated. Cells were then resuspended in 500μL of 10% formalin for 10 minutes at RT. 1500μL PBA was then added and cells spun at conditions above. Cells were then resuspended in 100μL BD Cytofix/Cytoperm (BD Biosciences) for 30 minutes on ice. 1mL of BD Perm/Wash buffer was then added and the sample was centifuged. The supernatant was aspirated and the cells were then resuspended in 100μL Cytoperm Plus for 10 minutes on ice and 1mL of BD Perm/Wash buffer was then added and the samples were centrifuged. The supernatant was aspirated and the cells were then resuspended in100μL BD Cytofix/Cytoperm for 5 minutes on ice. 1mL of BD Perm/Wash buffer was then added. 10% of the sample was set aside for unstained controls, the main sample centrifuged. To stain, cells were resuspended in 100μL rat anti-mouse IgG1-Biotin conjugate (BD Pharmingen) at 2.5 μg/ml in Perm/Wash buffer for 30 minutes, 1mL of BD Perm/Wash buffer was then added then centrifuged. The supernatant was aspirated and the cells were then resuspended in 100μL Streptavidin-APC conjugate at 0.2 μg/ml (BD Pharmingen) in Perm/Wash buffer for 30 minutes on ice, 1mL of BD Perm/Wash buffer was then added then centrifuged. The supernatant was aspirated and the cells were then resuspended in 500μL PBA and analysed on BD-Fortessa.

Preparation of samples for dNTP assay

Mouse splenocytes were used for dNTP analysis by pelleting at 500g for 5 min at 4 degrees Celsius. Samples were washed with cold PBS thrice with pelleting at 500g for 5 min at 4 degrees Celsius. Samples were then resuspended in ice cold 60% methanol at 10^7 cells per mL. cells were then heated at 95 degrees Celsius for 3 minutes, and then sonicated for 30 seconds at 70% maximum amplitude.

Samples were then dried in vacuum centrifuge at 60 degrees Celsius, 500g, 0.001 bar until dry.

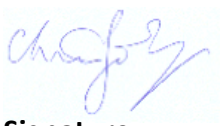
3. Proximity to AGCT sequences dictates MMR-independent versus MMR-dependent mechanisms for AID-induced mutation *via* UNG2

Thientosapol ES, Sharbeen G, Lau KKE, Bosnjak D, Durack T, Stevanovski I, Weninger W, Jolly CJ. Proximity to AGCT sequences dictates MMR-independent versus MMR-dependent mechanisms for AID-induced mutation via UNG2. *Nucleic Acids Research*: 2017 Apr 7;45(6):3146-3157. doi: 10.1093/nar/gkw1300. PMID: 28039326.

Author Contributions:

C.J.J. designed research; Experiments from Figures 2, 3, 4 were performed by E.S.T., G.S. and D.B. with assistance by T.D., I.S. and C.J.J. Experiments from Figure 5 were performed by K.K.E.L.; W.W. contributed new reagents/analytic tools; E.S.T., G.S., K.K.E.L. and C.J.J. analysed data. E.S.T. and C.J.J. wrote the paper.

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Chris Jolly		27/6/2018
Supervisor Name	**Signature**	**Date**

3146–3157 *Nucleic Acids Research, 2017, Vol. 45, No. 6* *Published online 30 December 2016*
doi: 10.1093/nar/gkw1300

Proximity to AGCT sequences dictates MMR-independent versus MMR-dependent mechanisms for AID-induced mutation *via* UNG2

Eddy Sanchai Thientosapol†, George Sharbeen†, K.K. Edwin Lau†, Daniel Bosnjak, Timothy Durack, Igor Stevanovski, Wolfgang Weninger and Christopher J. Jolly*

Centenary Institute, Royal Prince Alfred Hospital, Camperdown NSW 2050, and Sydney Medical School, The University of Sydney, Sydney NSW 2006, Australia

Received October 28, 2016; Revised December 12, 2016; Editorial Decision December 12, 2016; Accepted December 16, 2016

ABSTRACT

AID deaminates C to U in either strand of *Ig* genes, exclusively producing C:G/G:C to T:A/A:T transition mutations if U is left unrepaired. Error-prone processing by UNG2 or mismatch repair diversifies mutation, predominantly at C:G or A:T base pairs, respectively. Here, we show that transversions at C:G base pairs occur by two distinct processing pathways that are dictated by sequence context. Within and near AGCT mutation hotspots, transversion mutation at C:G was driven by UNG2 without requirement for mismatch repair. Deaminations in AGCT were refractive both to processing by UNG2 and to high-fidelity base excision repair (BER) downstream of UNG2, regardless of mismatch repair activity. We propose that AGCT sequences resist faithful BER because they bind BER-inhibitory protein(s) and/or because hemi-deaminated AGCT motifs innately form a BER-resistant DNA structure. Distal to AGCT sequences, transversions at G were largely co-dependent on UNG2 and mismatch repair. We propose that AGCT-distal transversions are produced when apyrimidinic sites are exposed in mismatch excision patches, because completion of mismatch repair would require bypass of these sites.

INTRODUCTION

During adaptive immune responses, the affinity of antigen-specific antibodies increases over a time frame of a few days to weeks. Mutations are introduced into the *Ig* genes of activated B cells proliferating in germinal centres (1) by the DNA editing enzyme AID (activation-induced deaminase, gene *Aicda*) (2,3). AID deaminates C to U in targeted genes, producing a U:G mismatch founder mutation (4). AID has a preference to deaminate C within 5′-WRCH-3′, on either DNA strand (4,5). *Ig* genes concentrate this motif in the hypervariable regions of *Ig* V gene segments, and have undergone selection against it in most of the framework regions. As a consequence, mutation hotspots are concentrated in the hypervariable regions (6). A palindromic iteration of WRCH: AGCT, is the most favoured hotspot and AGCT motifs are highly abundant in Ig *S* regions (7–11).

If deaminations are not repaired, one daughter cell will inherit a U:A base pair in place of the original C:G base pair (a phase 1A mutation) and the other daughter will be un-mutated. DNA repair enzymes can correct U:G mismatches, but error-prone processing also diversifies mutation (reviewed in (12)). Specifically, the U can be excised by the uracil-specific enzyme UNG2 (uracil N-glycosylase), creating an apyrimidinic (AP) site (12). Conventionally, AP-endonucleases (APE1 or 2) can nick 5' to AP sites to facilitate repair synthesis by polymerase β, restoring the original base pair (13). Loss of UNG2 activity reduces the frequency of transversion mutations at C:G base pairs by >80% (14,15). It is frequently proposed that some AP-sites produced by the sequential action of AID and UNG2 are replicated before they are processed further by BER enzymes, requiring bypass of the AP sites by translesion DNA polymerases (phase 1B mutation) (12,16–19). Alternatively, the AP-lyase activity of the MRN complex may induce error-prone BER (20). Loss of UNG2 activity can also increase the frequency of C:G to T:A transition mutations, presumably as a result of increased uracil replication. This suggests that high fidelity uracil BER does occur in hypermutating B cells (9,11,19,21).

U:G mismatches are recognized by the mismatch binding protein MutSα: a heterodimer of MSH2 and MSH6.

*To whom correspondence should be addressed. Email: c.jolly@centenary.org.au
† These authors contributed equally to this work as first authors.
Present addresses:
George Sharbeen, Lowy Cancer Research Centre, Prince of Wales Clinical School, UNSW Sydney, NSW 2052, Australia.
K.K. Edwin Lau, Haematology Coagulation, Westmead Hospital, Westmead NSW 2145, Australia.

MutSα-deficiency substantially reduces the frequency of *Ig* mutations at A:T base pairs and also reduces transversion mutations at C:G base pairs. This has produced a consensus view that mismatch repair (MMR) plays a dominant role in AID-induced mutation at A:T base pairs, and a minor role in generating transversion mutations at C:G base pairs (12,16). Canonical MMR occurs post-replication: the binding of MutSα to a mismatch induces MutSα translocation, the formation of a MutSα/MutLα complex, and the recruitment of Exonuclease I (*ExoI*). Exo I excises a large single strand patch, starting from the nearest available 3′-end (22,23). The excision patch is then in-filled by DNA polymerases δ or ϵ (24). In contrast, *Ig* A:T mutation exhibits little or no dependence on MutLα (see (12)), but still requires Exo I (25). It additionally requires PCNA capable of ubiquitination at K164 (26,27) and the translesion DNA polymerase η (pol η, (28)). $K164^{Ub}$-PCNA/pol η is presumed to introduce mutations within Exo I excision patches and preferentially at A:T base pairs (phase II mutation, (16)), perhaps as a consequence of pol η's inherent infidelity (29), and/or as a consequence of low dNTP levels in G1-phase cells (30,31). It's not clear how nicks are created to enable Exo I entry during AID-induced MMR. UNG2 and SMUG1 induce at least some of the nicks *via* APE, but are semi-redundant with other nick generators; perhaps TDG, MBD4, MutLα or OGG1 (32–36).

Genetic ablation of MutSα did not increase the frequency of C:G to T:A transition mutations in hypermutating B cells in several studies (37–39), leading to the suggestion that faithful mismatch repair is poorly recruited to AID-induced founder mutations (21). However, it's possible that faithful MMR has been under-estimated because the most proliferated germinal center B cells are preferentially lost in MMR-deficient mice (37,38). Ablation of both UNG2 and MutSα results in a mutation spectrum consisting entirely of C:G to T:A transitions, on both strands. (9,19,40–41). This suggests that although SMUG1 and MBD4 do excise AID-induced uracils in MMR-competent cells (32,36,42), processing by these *N*-glycosylases does not occur in the absence of MutSα, or is non-mutagenic in V-regions when it does occur.

Thousands of AP sites are produced daily in normal cells independently of AID (43), but do not cause wholesale point mutation. The factors rendering AID-induced uracil excision mutagenic remain unclear. Here, we exploit mutation datasets collated using the SW_{HEL} transduction/adoptive transfer model (see (30)) to analyse hypermutation at day 6 in a model immune response, prior to the narrowing of the response to a limited number of B cell dynasties. We find that both UNG2 and MMR repair many AID-induced uracils, especially in the top strand. Excluding C:G base pairs within and near AGCT hotspots—wherein UNG2 is mutagenic regardless of MMR activity—we find that interaction between UNG2 and MMR is required for up to 90% of transversions at C:G base pairs. Our data demonstrate that UNG2-mediated mutation occurs by distinct MMR-independent or MMR-dependent pathways, which are dictated by local sequence context.

MATERIALS AND METHODS

Mice

Male C57Bl/6 host mice were purchased from Animal Resources Centre (Canning Vale, Western Australia) and were used in experiments when 8–16 weeks old. SW_{HEL}, $Msh2^{ko/ko}$ and $Rag1^{ko/ko}$ mice—all C57Bl/6 background (30,44) were interbred and maintained under SPF conditions in the Centenary Institute Animal Facility. Mouse experiments were approved and monitored by the University of Sydney Animal Ethics committee in accordance with the New South Wales Animal Research Act (1985 No. 123).

Retroviral transduction and adoptive transfer

pMiG-based retroviruses expressing EGFP and ugi-EGFP fusion protein have been described (19). cDNAs encoding mouse UNG2 (CCDS19560) or SMUG1 (CCDS27898) were cloned immediately 5′ to the internal ribosome entry site (IRES) of the pMiG retroviral vector (45). cDNAs encoding UNG2-EGFP or SMUG1-EGFP fusion proteins were cloned in place of the IRES and EGFP sequences in the pMiG vector. Retroviral supernatants were produced using calcium phosphate-mediated transient transfection of Plat-E packaging cells (46). Primary SW_{HEL} splenocytes were purified by density gradient separation on Histopaque 1083 (Sigma-Aldrich) and activated overnight by culture with recombinant CD40L as described (30), transduced by 'spinfection' (1100 *g*, 45 min, 20°C) with retrovirus supernatants in the presence of 4 μg/ml Polybrene, washed, then incubated for a further 2 days in activating medium. GFP^{+ve} cells were purified by flow cytometric sorting (BD FACSAria IIU or Influx sorter), mixed with freshly HEL-conjugated sheep red blood cells (prepared as described (30)) and injected via a tail vein into 8–12 week old male C57Bl/6 hosts that had been primed i.p. with 10^8 non-conjugated SRBC in PBS seven days earlier. Each host received $\leq 10^4$ GFP^{+ve} cells, along with 10^8 HEL-SRBC in a bolus of 0.25 ml culture medium. Retroviral experiments were performed in accordance with a permit from the Office of the Gene Technology Regulator (Australia), overseen by the Royal Prince Alfred Hospital Institute Biosafety Committee (Sydney).

Measurement of SW_{HEL} VDJ-mutation in single cells

Six days after adoptive transfer, individual transduced SW_{HEL} B cells were recovered from the spleens of host mice into 96-well PCR plates (Bio-Rad) as GFP^{+ve} cells that bound HEL conjugated to Alexa Fluor 647 (30). Following cell lysis and protein digestion, the VDJ_H-region of the single SW_{HEL} allele present in each well was amplified by nested PCR, as described (30). PCR products from ≤60 wells per host in which amplification was successful were Sanger sequenced by Macrogen (South Korea). Mutations where secondary peaks formed <30% of the signal were confirmed using Sequencher software (version 5.1, Gene Codes Corporation). Processed sequences were sorted into phylogenetic trees (using neighbor joining and uncorrected 'p') with MacVector software (version 12.7.5 MacVector Inc) to check for clonal dynasties, then collated

for a window spanning nucleotides 17–539 (counting from the translation start ATG codon as bases 1–3) using custom Microsoft Excel spreadsheets (version 15.23 for Mac, Microsoft Corporation). Statistical analyses of mutation data exported from Excel were performed using Prism software (version 7.0a for Mac, GraphPad Software).

RESULTS

Production of large datasets using the SW_{HEL} adoptive transfer model

SW_{HEL} mice carry the HyHEL10 VDJ_H3-rearrangement targeted to the *IgH* locus plus the HyHEL10 *Igk* as a low copy number randomly-integrated transgene, in the C57BL/6 strain. This enforces high affinity for hen egg lysozyme (HEL) upon B cells—as long as receptor editing has not occurred at the *IgH* locus (47). Editing is blocked in $Rag1^{ko/ko}$ SW_{HEL} mice, so all B cells are specific for HEL in this line (44). We blocked UNG2 activity in $Rag1^{ko/ko}$ SW_{HEL} (hereon simply called SW_{HEL}) B cells *via* retroviral expression of the uracil glycosylase inhibitor (ugi) protein from phage (tagged with GFP), which blocks 98% of UNG activity (48). Retroviral expression of GFP alone served as a control. MutSα-deficient B cells were created by crossing $Msh2^{ko}$ alleles (49) into SW_{HEL} mice, and $Msh2^{ko/ko}$ SW_{HEL} cells were transduced to express either GFP or ugi-GFP (19). We used our established procedure (30) to analyse hypermutation in transduced adoptive SW_{HEL} cells (Figure 1). Briefly, C57BL/6 host mice were primed i.p. with 10^8 sheep red blood cells (SRBC) on day −7, to activate SRBC-specific T-cell help. On day −3, naïve donor SW_{HEL} splenocytes were placed into culture with recombinant soluble CD40L. On day −2, these cells were transduced with retroviruses derived from pMiG (45). On day 0, GFP^+ cells were purified and $\leq 10^4$ GFP^+ cells plus 10^8 HEL-conjugated SRBC were injected i.v. into each SRBC-primed host. By day +6, almost all adoptive SW_{HEL} cells are in germinal centres (50) and comprise 0.1–0.5% of total splenocytes (data not shown). On day +6, individual HEL-binding GFP^+ cells were recovered from host spleens into 96-well PCR plates, and the SW_{HEL} VDJ_H allele present in each well was amplified by nested single cell PCR from genomic DNA, usually with ∼70% efficiency (30).

Most of the mutations we collate are in exons (Supplementary Figure S1A), and subject therefore to Ag-selection, which has potential to bias the mutation spectrum towards rare mutations that favour antigen-binding. However, the HyHEL10 mAb from which SW_{HEL} mice were derived has very high affinity for HEL (51), so Ag-selection is likely only to be against deleterious mutations, rather than for affinity enhancing mutations, especially given that adoptive SW_{HEL} cells enter germinal centers only three days prior to our harvesting of SHM data (50). In support of this assertion, the distribution of mutations across all twelve possible point mutations in the adoptive SW_{HEL} model is comparable to non-selected datasets derived from J_H–C_H intron sequences and from passenger *Ig* transgenes (based on comparison of Tables in (19) and (12)). Furthermore, as long as no more than fifty sequences are collated per host mouse, clonal dynasties are rare (30).

From now on, we will refer to cells expressing ugi-GFP or GFP as 'UNG$^-$' or 'UNG$^+$' cells, respectively, and refer to $Msh2^{ko/ko}$ or $Msh2^{wt/wt}$ cells as 'M$^-$' or 'M$^+$' cells, respectively. We collated mutations in UNG^+M^+, UNG^-M^+, UNG^+M^- or UNG^-M^- SW_{HEL} cells, 6 days after antigen encounter, using 10 or 11 host mice per dataset (Supplementary Figure S1A, Table 1, Supplementary Table S1). As indicated in Table 1 (column *s*), sub-sets of the data collated here have been published for other purposes (19). Mutation frequencies were acceptably reproducible between independent hosts (Figure 2). The same hotspots were preferentially targeted in multiple mice within each dataset (Supplementary Figure S1B), demonstrating that hotspots arose as a result of repeat targeting, and were not an artifact of clonal dynasties. This indicated that our model could be used to quantitatively analyze antibody hypermutation.

Deficiency for UNG2 or MMR increases transition mutation at C:G base pairs

Deficiency for UNG2 activity (in UNG^-M^+ cells) or for Mutsα (in UNG^+M^- cells) increased transitions at C:G 2.3-fold or 1.6-fold, respectively, relative to UNG^+M^+ cells (Table 1). Reduced UNG activity increased transitions at C ($p = 0.002$) and at G ($p < 0.0001$) almost equally, while *Msh2*-knockout significantly increased transitions at C ($p = 0.019$), but not at G (Figure 2A). In double-deficient UNG^-M^- cells, transitions at C:G were increased 5-fold overall, relative to UNG^+M^+ cells (Table 1), with mutations at C increasing 2.3-fold more than mutations at G (Figure 2A). These transition increases in UNG^-M^- cells were highly significant, even in comparison to the UNG^-M^+ and UNG^+M^- datasets (Figure 2A). The data are consistent with MMR and UNG2 (especially UNG2) driving more faithful than mutagenic repair at deamination sites, but it is also possible that MMR and/or UNG2 curtail ongoing deamination by AID, such that deficiency for UNG2 or MMR gives AID greater leeway to deaminate cytosines in the first place. In other studies, mutation at C was also increased more than at G by double-deficiency for UNG2 and MutSα, but in two of those studies (9,40) the bias towards increased mutation at C was not as high as in this or another study (41). This may simply reflect experimental variability, which is better controlled in our study (because $n \geq 10$ in each of our datasets) than in previous studies, or it may be due to differences in compensatory mechanisms induced by *Ung*-knockout versus ugi over-expression.

UNG2 and MMR semi-redundantly excise uracils

Transversion mutation at C:G (mostly phase 1B mutations (12)) and mutations at A:T (phase II mutations (12)) were almost eliminated in UNG^-M^- cells (Figure 2B and C, Table 1, Supplementary Table S1, Supplementary Figure S1A), as expected (40). The residual non-phase 1A mutations (16/999 mutations from 368 cells, Table 1) could be be due to a combination of PCR error (3 mutations per 375 cells—see Table 1), residual UNG2 activity (excess ugi leaves 2% of UNG activity intact (48)), and possibly the activity of other *N*-glycosylases such as SMUG1, TDG or MBD4. Thus, transition mutation at C:G in UNG^-M^- cells

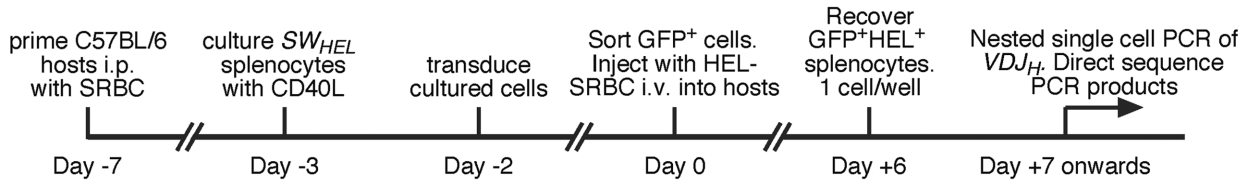


Figure 1. Timeline for SW_{HEL} transduction/adoptive transfer experiments.

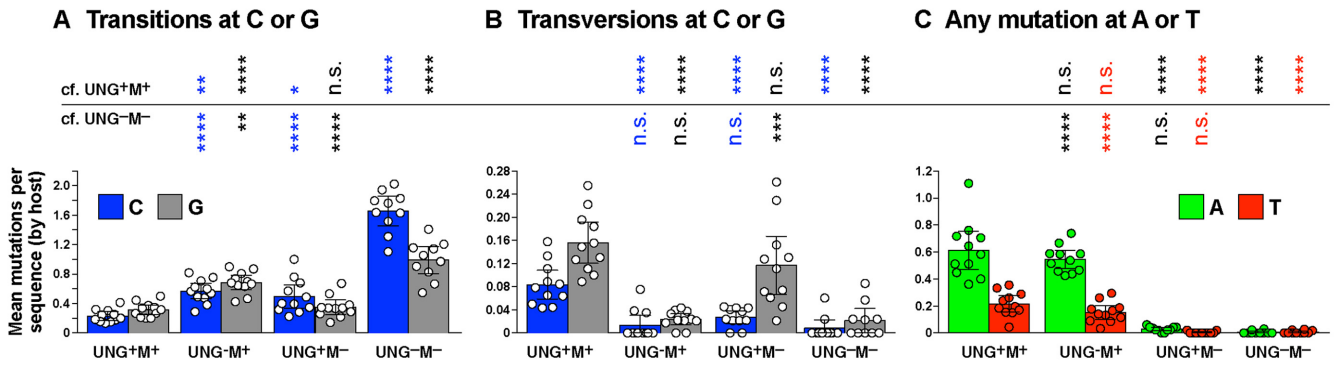


Figure 2. Mutations in the top strand of the SW_{HEL} VDJ_H sequence window were averaged per host mouse (dots), with the overall mean of hosts (± 95% confidence interval) indicated by histograms. Significant differences (according to Tukey's ordinary one-way ANOVA) compared to ('cf.') the UNG^+M^+ or UNG^-M^- datasets are indicated: ****$p < 0.0001$; ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; n.s., not significant. Mutations are divided into three classes: (**A**) transition mutations at C or G; (**B**) transversion mutations at C or G; (**C**) any mutation at A or T.

Table 1. Summary of datasets and estimates of uracil processing as indicated by drops in transition mutations at C or G relative to the UNG^-M^- 'baseline', plus estimates of the relative contributions of UNG2 and MMR to other point mutations

a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*	*k*	*l*	*m*	*n*	*o*	*p*	*q*	*r*	*s*
						Mutations per 523bp sequence window:						Processing of baseline U						
				No. of		G		C				drop in baseline Ts		U converted to Tv		contribution to:		
Transduced cells	Transduced protein	Shorthand	Hosts	Seq's	Mut'ns	Tv	Ts	Tv	Ts	A	T	G	C	G	C	A	T	published hosts (19)
Msh2$^{ko/ko}$SW$_{HEL}$	ugi-GFP	UNG–M–	**10**	**368**	**999**	**0.02**	**0.97**	**0.008**	**1.70**	**0.005**	**0.008**		double-deficient baseline					3
Msh2$^{ko/ko}$SW$_{HEL}$	GFP	UNG+M–	11	514	527	0.12	0.35	0.027	0.50	0.03	0.00	64%	71%	10%	1.1%	4%	-2%	3
SW$_{HEL}$	ugi-GFP	UNG–M+	11	478	948	0.03	0.68	0.015	0.56	0.54	0.16	30%	67%	0.3%	0.4%	86%	75%	4
SW$_{HEL}$	GFP	UNG+M+	11	539	867	0.15	0.31	0.082	0.23	0.62	0.21	68%	86%	14%	4.3%	100%	100%	6
	PCR background[a]	no AID	2	375	3	0	0.005	0	0.003	0	0							1

[a]PCR background from sequencing $CD4^+B220^-SW_{HEL}Rag1^{+/+}$ cells.

can be presumed to closely approximate the baseline deamination signature of AID in the absence of significant mutagenic uracil processing (52), and the residual phase 1B and phase II mutations we detected in UNG^-M^- cells can be considered little more than background noise.

We estimated the fraction of deaminations processed by UNG2 or MutSα by comparing C:G transition mutation in single-deficient UNG^+M^- cells or UNG^-M^+ cells to double-deficient UNG^-M^- cells. In UNG^+M^- cells, transitions were reduced by 71% at C and by 64% at G, relative to the UNG^-M^- baseline (Table 1). This suggests that UNG2 excises about two-thirds of deaminations in the absence of MMR, with 1.1-fold bias toward the non-template (upper) strand. Of course, dATP might be incorporated opposite deamination sites—indeed, dATP preferentially incorporates opposite AP sites *in vitro* (53,54). We presume, therefore, we have estimated a *minimum* amount of MMR-independent processing by UNG2. In UNG^-M^+ cells, transitions were reduced by 67% at C and by 30% at G, relative to the UNG^-M^- baseline (Table 1). Presumably this is also a minimum estimate of UNG2-independent processing by MMR in the absence of UNG2 activity, because Exo I excision of the G-bearing strand instead of the U-bearing strand of deamination sites would be indistinguishable from ignorant replication. MMR thus appeared to be much more biased towards excising upper strand deaminations than UNG2.

In UNG^+M^+ cells, transitions were reduced by 86% at C and by 68% at G, relative to the UNG^-M^- baseline; both of which are only a little greater than the uracil excision rate apparent in UNG^+M^-cells (Table 1). This result is consistent with incomplete redundancy between UNG2 and MMR for access to AID-induced uracils.

Collaboration between UNG2 and MMR to diversify mutation

As expected from the Neuberger model (12), transversion mutation at C:G was largely UNG2-dependent (because it was reduced by 85% in UNG^-M^+ cells, relative to

UNG^+M^+ cells) and mutation at A:T was largely MSH2-dependent (because it was reduced by 96% in UNG^+M^- cells, relative to UNG^+M^+ cells; Table 1, Figure 2B-C, Supplementary Figure S1A). Some groups reported a similarly high dependence of A/T mutation on MMR (38,55), while others reported dependencies nearer 50% (37,39). The reason for this variability remains unknown, but it is unlikely to be due to variations in antigen selection, and might be due to variation between mouse strains.

Transversion mutation at C was substantially co-dependent on UNG and MMR. It was reduced by 84% in UNG^-M^+ cells ($p < 0.0001$) and by 68% in UNG^+M^- cells, relative to UNG^+M^+ cells ($p < 0.0001$, Figure 2B). Transversion mutation at G was reduced by 85% in UNG^-M^+ cells ($p < 0.0001$), but only by 25% in UNG^+M^- cells ($p = 0.25$, Figure 2B), so it was much less dependent on MMR. Mutation at A:T base pairs also appeared to be somewhat UNG2/MMR co-dependent, because the frequencies of A and T mutations in UNG^+M^- cells and UNG^-M^+ cells summed to values lower than their frequencies in UNG^+M^+ cells (see columns *q-r* in Table 1). Co-dependence of some phase 1B and phase II mutations on MMR and BER is already well recognised (12,32–33,35,56–57).

Processing of U:G base pairs by MMR or UNG2 appears to be dictated by local sequence context

We were particularly interested to know whether local sequence context influenced the processing of deaminations. MMR-deficiency increases focusing of mutation on AID hotspots (25,37–39,55,58–59), but whether this occurs by altered AID-targeting (59) and/or by uneven uracil repair downstream of AID (28,33) is unclear. We wanted to quantify focusing in an unbiased manner. First, we sorted C or G mutation sites in each dataset from most to least mutated sites (Supplementary Figure S1B). We then calculated cumulative mutation at C or at G accordingly. The UNG^+M^- dataset produced outlying cumulative mutation plots for both C and G (Figure 3A). In the UNG^+M^- dataset, five C mutation sites or four G mutation sites (numbered in Figure 3B) accounted for ≥50% of mutation at C or at G, respectively (see 50% intercepts marked in Figure 3A). This was about half the number of mutation sites required in the other datasets. Seven of these nine most prominent UNG^+M^- hotspots occurred within AGCT or AGCA motifs (AGCW motifs, Figure 3B), in which the C on both strands conforms to the WRCH hotspot consensus. In other words, mutation focusing at C and at G in UNG^+M^- cells was about 2-fold greater than in the other treatment groups, and focussed towards palindromic hotspots (AGCT) or hotspots that are 'quasi-palindromic' from AID's viewpoint (AGCA). A direct role for MSH6 in AID-targeting (59) may contribute to mutation focusing in UNG^+M^- cells, but it is difficult for this role to explain the relaxed focusing evident in UNG^-M^- cells (see Figure 3A). Interference from MMR influences the fidelity of UNG2-mediated repair *in vitro* (33). This supports the proposal of Delbos *et al.* that mutation focusses to motifs that become hyper-resistant to faithful UNG2-mediated repair in MMR-deficient cells (28).

To test the proposal of Delbos *et al.*, we measured changes in C or G transition frequency *on a per-site basis* in UNG^-M^+ cells and UNG^+M^- cells, compared to the UNG^-M^- baseline. This approach was based on the presumption that most C or G transitions are phase 1A mutations. To exclude rarely deaminated sites, only sites where >7 mutations were present in the UNG^-M^- dataset were considered. This arbitrary cut-off still encompassed 75% or 74% of mutation at C or G, respectively, in the UNG^-M^- dataset. Both MMR or UNG2 reduced baseline C and G transition at all sites considered, except for G531, G239 and G273 in the UNG^-M^+ dataset (Figure 4A). G239 and G273 were also noticeable outliers in the UNG^+M^+ dataset (Figure 4A). These outliers might be sites where MSH6 enhances AID-targeting, as proposed by Scharff's group (59). In MMR-competent UNG^-M^+ and UNG^+M^+ cells, baseline transitions were reduced outside and within AGCW sites almost equally (Figure 4A). However, in UNG^+M^- cells baseline transitions were reduced at sites *outside* AGCW almost twice as much as *within* AGCW motifs ($p = 0.0016$, Figure 4A). To confirm that these findings were consistent across replicate hosts, we plotted C:G transition frequencies outside or within AGCW motifs, for each host of UNG^+M^+ or UNG^+M^- cells (Figure 4B). This confirmed that C:G transition frequencies outside AGCW sites were comparable between UNG^+M^+ and UNG^+M^- cells, but within AGCW motifs were significantly higher in UNG^+M^- cells than in UNG^+M^+ cells ($p = 0.014$).

Thus, the AGCW sites upon which hypermutation became most focused in UNG^+M^- cells were sites that *appeared to be* especially resistant to MMR-independent processing by UNG2, as proposed by the groups of Weill and Ramiro (11,28). Nonetheless, the site-dependent changes in C:G transition mutations revealed in Figure 4A could be due to changes in targeting by AID, rather than changes in uracil-processing. If this were so, then the ratio between mutagenic and repair outcomes at AGCW versus non-AGCW deamination sites would not be expected to vary. We therefore determined if the frequency of mutagenic outcomes from deamination within AGCW versus outside AGCW varied.

Outside AGCW motifs, transversion mutation at C:G base pairs requires interaction between UNG2 and MMR

In the mutation skyline of UNG^+M^- cells shown in Figure 3B, transversion mutations at C/G were rare outside hotspots. We therefore calculated the percentage of mutations that were transversions at each C or G mutation site for the UNG^+M^- and UNG^+M^+ datasets (Figure 3D). Analysis was restricted to sites that acquired >7 mutations in the UNG^-M^- dataset. The percentages of mutations at C:G that were transversions were comparable within versus outside AGCW sites in UNG^+M^+ cells ($p = 0.82$, Figure 3D bottom). In contrast, they formed a reduced proportion of mutations at C:G outside AGCW sites, compared to inside AGCW sites in UNG^+M^- cells ($p = 0.023$, Figure 3D top). To confirm consistency across biological replicates, we plotted the frequencies of transversion mutations at C:G that lay outside or inside AGCW motifs, for each host of UNG^+M^+ or UNG^+M^- cells (Figure 4C). Transver-

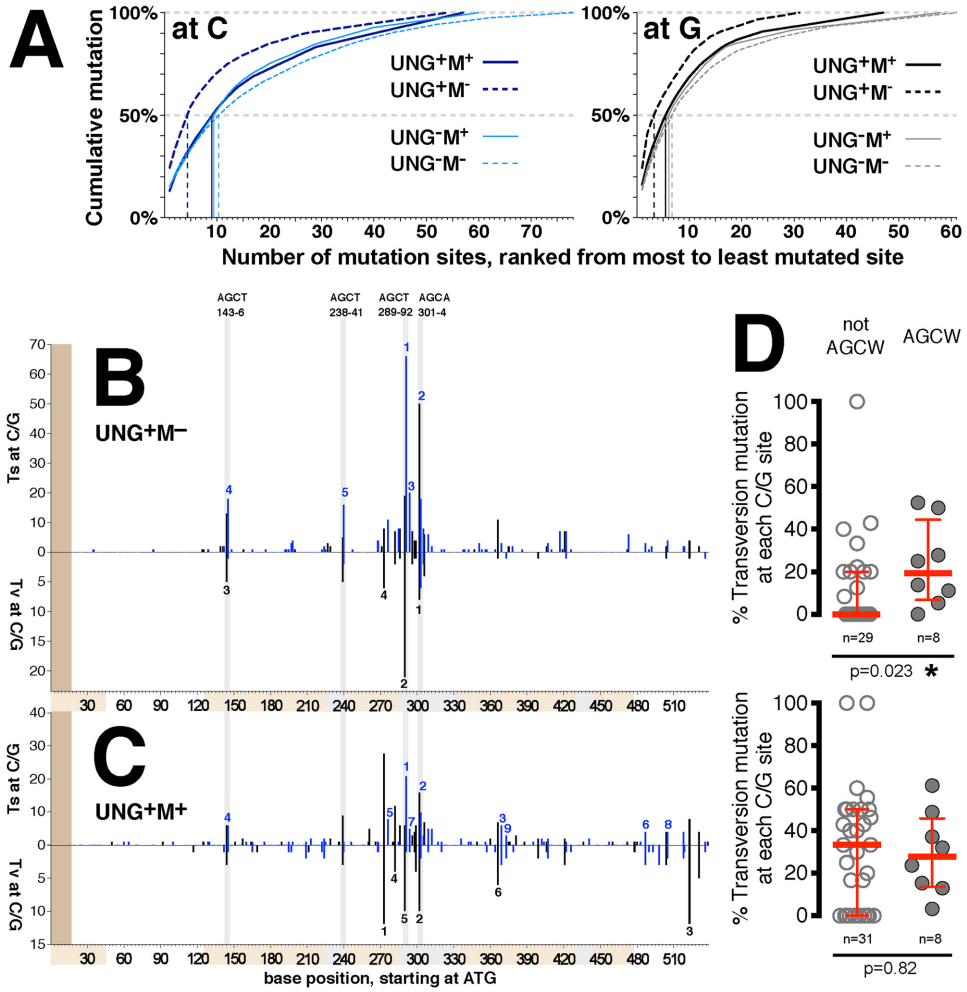


Figure 3. Mutation focusing. (**A**) Mean mutations per mutation site ranked from most to least mutated (see Supplementary Figure S1B) were transformed into plots for cumulative mutation at C (blue, left) or at G (black, right) in the top strand. The x-axis indicates rank position (leftmost rank = most mutated C or G site), while the y-axis indicates cumulative mutation at C or at G. 50% intercepts indicate the least number of mutation sites required to account for ≥50% of all mutation at C or at G for each dataset. (B and C) Skylines of (above the x-axis) total transition or (below the x-axis, with a different y-scale) transversion mutations at (blue) C or (black) G in (**B**) UNG^+M^- cells or (**C**) UNG^+M^+ cells. Mutations were not collated in the most 5′ region shaded brown. Grey stripes identify AGCW motifs. Beige and gray boxes at bottom indicate exons and CDRs (according to http://www.imgt.org), respectively. Top-ranked mutation sites that together accounted for ≥50% mutation at C or at G are indicated by integers above or below the x-axes, respectively. (D) Percentage of mutations that were transversions at each C or G site in the (top) UNG^+M^- or (bottom) UNG^+M^+ datasets, restricted to sites that acquired >7 mutations in the UNG^-M^- dataset, and partitioned into those lying (open symbols) outside or (closed symbols) inside an AGCW motif. Medians and interquartile ranges are indicated in red. *p*-values from Mann–Whitney tests are indicated.

sion mutation at C:G outside AGCW sites was markedly reduced in UNG^+M^- compared to UNG^+M^+ cells (p = 0.007), but this was not the case within AGCW motifs (Figure 4C). This suggested that MMR was dispensable for UNG2-dependent mutation within AGCW motifs. Everywhere else, UNG2-dependent mutation was largely co-dependent on MMR. We noticed that most of the transversion mutations that remained outside AGCT motifs in our UNG^+M^- dataset were close to an AGCT site, including the transversions at AGCA (examine Figure 3B). This raised the possibility that proximity specifically to AGCT motifs, rather than AGCW motifs, might be the major factor governing transversion rate in UNG^+M^- cells. To normalize confounding influences, we plotted the ratio of transversion mutation in UNG^+M^- cells versus UNG^+M^+ cells as a function of absolute distance (i.e. 5′ or 3′) from the nearest AGCT motif. We only performed this analysis for transversions at G, because transversions at C were rare in all data sets (see Table 1). The plot produced was a good fit to exponential decay of MMR-independent G transversion mutation with distance from the nearest AGCT motif (Figure 4D; R square = 0.88; absolute sum of squares = 0.98), with a half-life of 5.3 bases (95% confidence interval = 3.1–8.3 bases), towards an asymptote of ∼0.1. In other words, repair by UNG2 was ∼90% error-free in M^- cells, except within a roughly five base window either side of AGCT motifs. Within this window, MMR was redundant for UNG2-dependent mutagenesis.

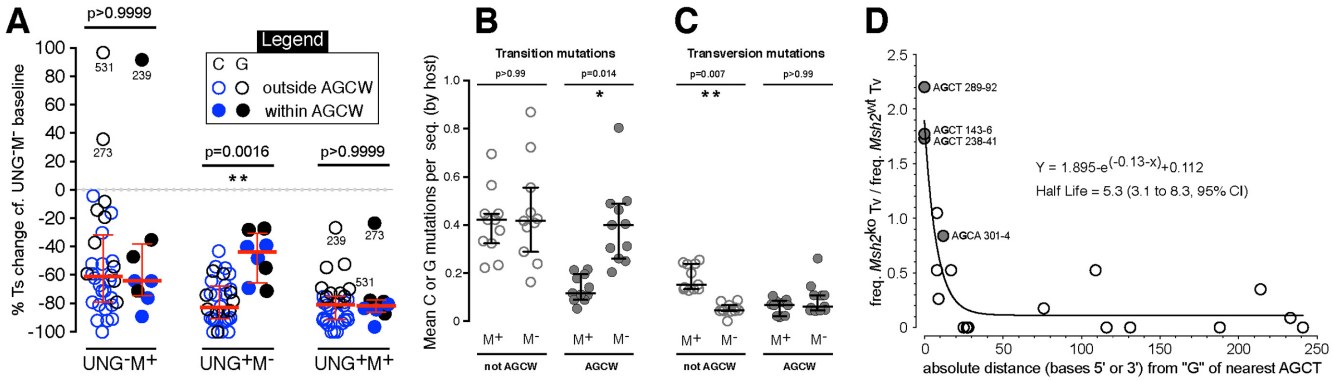


Figure 4. Processing of deaminations outside versus within AGCW motifs. (**A**) The change in transition mutation in the top strand at (blue) C or (black) G in each dataset, relative to the baseline $UNG^{-}M^{-}$ dataset, for every site that was mutated more than seven times in the $UNG^{-}M^{-}$ dataset. Medians and interquartile ranges are indicated in red. Significant differences between (solid symbols) AGCW and (open symbols) non-AGCW sites are indicated by Kruskal-Wallis *p*-values. (B and C) Mean frequencies per sequence of (**B**) transition mutations or (**C**) transversion mutations at C:G base pairs lying (open symbols) outside or (closed symbols) inside AGCW sites, for individual hosts of (M^{+}) $UNG^{+}M^{+}$ or (M^{-}) $UNG^{+}M^{-}$ cells. Bars indicate medians and inter-quartile ranges of hosts means. Significant differences between $UNG^{+}M^{+}$ cells and $UNG^{+}M^{-}$ cells are indicated by Kruskall-Wallis *p*-values. (**D**) The MMR-dependence of normalised transversion mutation at G (y-axis) upon distance from 'G' in the nearest AGCT motif (x-axis). The best-fitting exponential decay curve (Prism software) is overlaid.

UNG2 over-expression suppresses phase 1A and phase II mutations, but has little effect on phase 1B mutation

As a direct test for resistance of deaminated AGCW motifs to faithful processing by uracil *N*-glycosylases, we used the adoptive SW_{HEL} model to over-express SMUG1 or UNG2 in MMR-proficient B cells and measured the impact on mutation within and outside AGCW hotspots. SMUG1 over-expression substantially suppressed mutation at A:T base pairs ($p = 0.006$, Figure 5A). Outside AGCW motifs, it also suppressed C/G transition mutation ($p = 0.0017$, Figure 5B) and almost ablated C/G transversion mutation ($p = 0.0024$, Figure 5C). However, SMUG1 over-expression had no significant impact on C/G mutation within AGCW motifs (Figure 5B and C). This suggested that SMUG1 was largely unable to process deaminations within AGCW motifs, while outside AGCW motifs, SMUG1 induced non-mutagenic BER almost exclusively or prevented deamination outside AGCW in the first place. The overall reduction in hypermutation induced by SMUG1 over-expression (Figure 5 and Supplementary Figure S2) is consistent with a previous report using human SMUG1-transgenic mice (42). Like SMUG1, ectopic UNG2 substantially suppressed mutation at A:T and C/G transition mutation outside AGCW motifs ($p < 0.0001$, Figure 5A and B), and had no impact upon C/G transversion mutation within AGCW motifs (Figure 5C). In addition, UNG2 over-expression slightly reduced transition mutation within AGCW motifs ($p = 0.04$, Figure 5B) and slightly reduced transversion mutation outside AGCW motifs ($p = 0.04$, Figure 5C).

Regardless of any impact SMUG1 or UNG2 over-expression may have on deamination rates, Figure 5 indicates firstly that most *Ig* C and G transitions arise from replication of unrepaired uracils, and secondly that deaminations within AGCW sites are resistant to faithful repair by UNG2 or SMUG1, even when these *N*-glycosylases are over-expressed. Furthermore, the virtual absence of transversion mutations at C:G outside AGCW motifs in SMUG1 over-expressing cells (Figure 5C) suggests that, unlike UNG2, SMUG1 cannot collaborate with MMR to produce AGCT-distal phase 1B mutations.

DISCUSSION

Incomplete redundancy between UNG2 and MMR to process AID-induced uracils

In contrast to early studies using memory or Peyer's patch B cells, loss of either UNG2 or MutSα activity increased V-region transition mutation at C:G bases in our day 6 germinal center B cells, especially at C (Figure 2A). The increases could be due to increased ignorance of deaminations, and/or to these factors curtailing deamination by AID. In theory, excision of template strand uracils by UNG2 could curtail iterative deamination, because AP-sites block transcription in yeast (60) and might therefore be expected to alter recruitment of AID by transcription complexes. This could explain why C and G mutation in $UNG^{-}M^{-}$ cells were less focused than in wild-type $UNG^{+}M^{+}$ cells (Figure 3A). Nonetheless, we suspect the presence of AP sites does not have a major effect on AID-recruitment, firstly because mutation focusing in $UNG^{-}M^{-}$ cells was only slightly more relaxed than in $UNG^{+}M^{+}$ cells (Figure 3A), and secondly because transversion mutation at C:G was only marginally reduced in cells over-expressing UNG2 (Figure 5C).

If we accept that UNG2 or MutSα induce only minor curtailment of deamination, Table 1 presents evidence that Mutsα is largely, but not completely redundant to UNG2 for excision of AID-induced uracils. Furthermore, the ability of ectopic UNG2 and SMUG1 to suppress MMR-dependent mutation (Figure 5A) suggests that excess uracil *N*-glycosylase activity out-competes MMR to process U:G base pairs. Recent data collated from day 8 anti-NP cells implied that MMR is completely redundant to UNG2 for processing U:G base pairs, because UNG2, MMR or combined UNG2/MMR activities reduced transitions at C:G by 81%, 64% or 78%, respectively, relative to

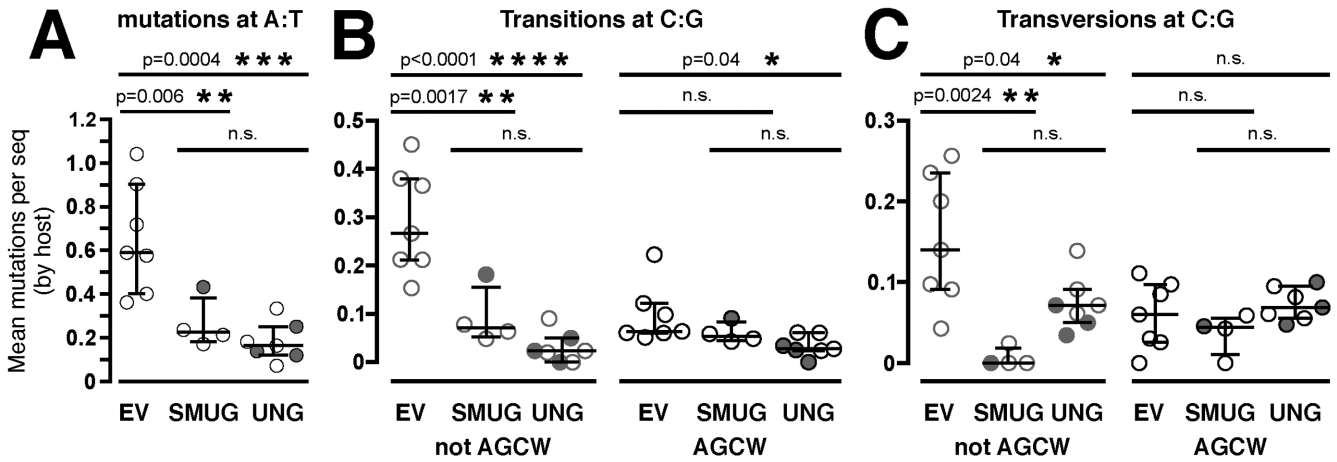


Figure 5. Uracil *N*-glycosylase over-expression suppresses mutation outside AGCW motifs. Mean frequencies of (**A**) mutations at A:T, (**B**) transitions at C:G or (**C**) transversions at C:G per sequence for each host mouse that received adoptive SW_{HEL} B cells transduced with (SMUG) SMUG1-expressing vector, (UNG) UNG2-expressing vector or (EV) empty pMiG vector. Open symbols: co-expression of native protein and GFP. Closed symbols: expression of SMUG1-GFP or UNG2-GFP fusion proteins. Bars indicate median ± inter-quartile range of the host means. Holm–Sidak's *p*-values are shown.

$Ung^{-/-}Msh2^{-/-}$ cells (our calculations from Supplemental Data in (9)). The deviations from our MMR data might be explained by target sequence composition, which has a profound impact on both AID-recruitment and processing downstream of AID ((8–9,11), this paper), or by the higher risk of sampling error in the NP study (due to fewer replicates per genotype) compared to this study, which has used the highest number of biological replicates per dataset to date. It is also possible that DNA repair-deficiency in T_{FH} cells impinges upon B cell hypermutation when $Ung^{-/-}$ or $Msh2^{-/-}$ mice are immunised. This complication would not occur in our datasets, because only the adoptive B cells carried DNA repair defects.

A role for uracil BER in phase II mutations?

Despite the ability of elevated uracil *N*-glysolysase activity to inhibit MMR-induced mutation (Figure 5A), loss of UNG2 activity did not increase MMR's ability to produce mutations at A:T base pairs in our study (see Table 1), nor in any other studies. Some have concluded from this that UNG2 and MMR do not compete to process deaminations (61). However, it is also consistent with a fraction of phase II mutation being UNG2 and MMR co-dependent. Any potential co-dependence cannot be explained by recruitment of MMR to AP sites, because MutSα does not recognize BER intermediates (62). However, entry by Exo I at BER-induced nicks (32–35) does provide a plausible collaboration mechanism. Suppression of A:T mutation by ectopic uracil *N*-glycosylases (Figure 5A) might seem to rule this idea out. However, excess uracil *N*-glycosylase probably deprives MutSα of U:G substrate, suppressing activation of MMR. UNG/SMUG1-double knockout and UNG/APE2-double knockout significantly reduced mutation at A:T base pairs (32,35). This is strong evidence of a major role for uracil BER in A:T mutation. UNG2/SMUG1-double knockout, but not UNG/APE2-double knockout, also reduced the strand bias of A:T mutation (32,35). AID and UNG2 seem to have 1.75-fold and 1.1-fold bias, respectively, towards activity in the top strand in our datasets (see Table 1). The reduction in transitions we induced with ectopic SMUG1 was 1.2-fold biased towards C over G (data not shown). Assuming that AP sites can be accessed by AP-endonucleases in each strand equally, we calculate that combined UNG2 and SMUG1 activity should produce 1.9 to 2.2-fold more nicks in the top strand than in the template strand of mutating SW_{HEL} alleles. This could bias Exo I entry into the top strand, should a nearby deamination site recruit MMR. Alternatively, AP-endonucleases may simply have better access to AP sites in the non-transcribed DNA strand. Either way, much of the strand bias of MMR-mediated uracil processing and of A:T mutation (see Table 1 and Figure 6) can be explained if BER is a major mediator of Exo I entry during AID-induced MMR. Semi-redundant nick generation by uracil BER and MutLα, and perhaps oxidation-induced BER (34) potentially explain why MutLα-knockout has only marginal impact on antibody hypermutation (57). Nick generation *via* MutLα might account for the markedly reduced strand bias of A:T mutation in $Ung^{-/-}Smug1^{-/-}$ mice (32), because there is no reason to assume that MutLα can distinguish between the non-deaminated and deaminated strands of hypermutating *Ig* genes.

Two distinct pathways for generating phase 1B mutations, dictated by proximity to AGCT motifs

Deaminations in AGCW motifs appeared more resistant to excision by UNG2 than deaminations at other sites (Figure 4A). In fact, the increase in C:G transition mutation induced by *Msh2*-knockout (relative to wild-type) was confined to these sites (Figure 4B). A similar phenomenon was noted in sequences 3′ to J_H4 (28). It's possible that targeting of AGCW motifs by AID increases when MMR is lost from UNG-competent cells because MSH6 can directly influence targeting by AID (see (59)). However, increased mutation-focusing in $UNG^{+}M^{-}$ cells cannot be fully explained by abilities of MSH6 or MMR to

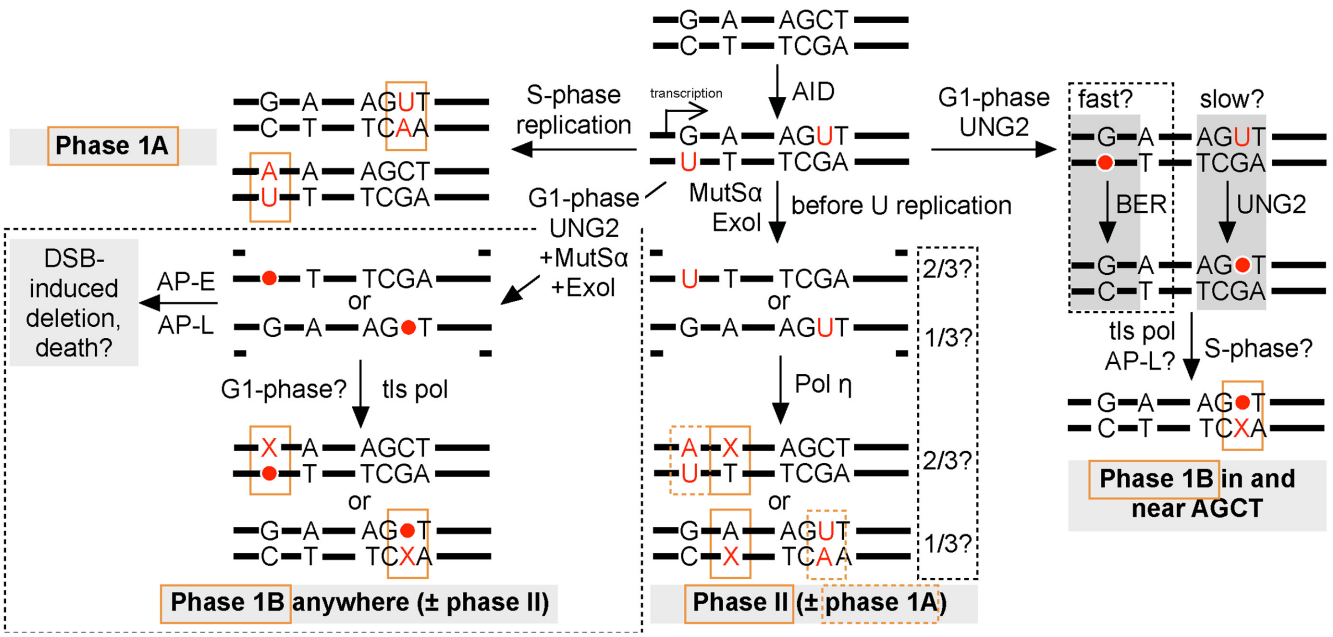


Figure 6. A revised hypermutation model based on (12,16,19–20). Dashed boxes enclose refinements added by this paper. The class of mutation (phase 1A, phase 1B or phase II) produced by each pathway is indicated by orange boxes. (•) AP site; (AP-E) AP-endonuclease; (AP-L) AP-lyase; (Pol η) DNA polymerase η; (tls pol) translesion polymerases; (Phase 1A) transition at C or G; (Phase 1B) transition or transversion at C or G; (Phase II) mutation at A or T.

influence AID targeting, because mutation was not similarly focussed in our UNG$^-$M$^-$cells (Figure 3A). Relaxed targeting was also apparent in *Ung*$^{-/-}$*Msh2*$^{-/-}$ cells (see (28,40)). Mutagenic processing was particularly focused to AGCW sites in UNG$^+$M$^-$ cells (Figure 3D). Thus, AGCW sites in UNG$^+$M$^-$ cells were resistant both to processing by UNG2 and to conventional BER steps downstream of UNG2, diverting to mutagenic outcomes at a higher rate than other sites in UNG$^+$M$^-$ cells. Even when UNG2 was over-expressed, C:G mutation rates within AGCW motifs were barely affected (Figure 5B and C). A similar phenomenon was observed in synthetic target genes in 3T3 cells expressing ectopic AID (11). Taken together, our results suggest that mutation-focusing previously described in UNG2-competent *Msh2*$^{-/-}$, *Msh6*$^{-/-}$, and *Exo1*$^{-/-}$ B cells (25,37–39,55,58) is more a consequence of altered uracil processing by UNG2 than of altered uracil production by AID.

We uncovered a mathematical relationship between distance from AGCT motifs and the MMR-dependence of transversion mutation at C:G (Figure 4D). We can predict from Figure 4D that the dependence of transversion mutation on MMR should decrease as the density of AGCT sequences in a target sequence increases. We can use the data of Chen *et al.*—who compared hypermutation in a conventional VDJ$_H$ knock-in allele with hypermutation in an AGCT-rich knock-in allele (9)—to test this prediction. *Ung*-knockout reduced C:G transversion mutation by 85% and 89% (relative to *Ung*$^{+/+}$*Msh2*$^{+/+}$ cells) in Chen *et al.*'s conventional and AGCT-rich target alleles, respectively. These reductions compare to the 85% drop observed in our UNG$^-$M$^+$ *SW*$_{HEL}$ cells. *Msh2*-knockout decreased C:G transversion mutation (relative to wild-type cells) in their conventional allele and our *SW*$_{HEL}$ allele by 44% and 40%, respectively. These reductions are also comparable, but contrast with the 29% reduction *Msh2*-knockout caused in Chen *et al.*'s AGCT-rich allele (our calculations, using Supplemental Data in (9)). This contrast is what we would predict from Figure 4D.

Based on our findings, we propose a refinement of the Neuberger hypermutation model (Figure 6), wherein AID-induced phase 1B mutation occurs by two distinct mechanisms: (1) In the near vicinity of AGCT motifs, phase 1B mutation depends on UNG2 and does not require MMR; (2) distal to AGCT motifs, phase 1B mutation is ∼90% co-dependent on UNG2 and MutSα. We propose that these distal mutations involve exposure of AP sites in Exo I excision patches. Critical to our model is the point that MMR could recruit ExoI to excise either the U-bearing or the G-bearing strand of U:G base pairs (albeit with a bias apparently towards ExoI entry into the top strand; see Discussion earlier), regardless of which strand was deaminated in the U:G base pair actually bound by MutSα. This creates the potential for uracils or AP-sites to be exposed in single-stranded MMR excision patches. If ExoI exposed AP-sites created by UNG2, or if UNG2 excised uracils after their exposure in ExoI excision patches, nicking by an AP-endonuclease or AP- lyase (20,35) would create a problematic double-strand break – likely leading to a deletion and/or cell death if nicking occurred in G1-phase (see Figure 6). AP-sites exposed within G1-phase ExoI excision patches would therefore require lesion bypass using a translesion polymerase to be resolved, potentially causing a phase 1B mutation (Figure 6). In the absence of MMR, or when MMR's access to U:G base pairs is overwhelmed by uracil *N*-glycosylase over-expression, AGCT-

distal uracils appear to be processed efficiently by UNG2, leading almost exclusively to faithful BER. We noticed that SMUG1 over-expression almost ablated transversions at C:G outside AGCW sites (Figure 5C). SMUG1 has been reported to bind the AP sites it creates more persistently than UNG2 (63). Bearing this in mind, Figure 6 can explain why excess SMUG1 prevents phase 1B mutation distal to AGCT motifs: if SMUG1, rather than UNG2, created an AP site within an ExoI excision patch, persistent binding of SMUG1 to its product AP site would block excision patch in-fill and force the creation of a double-strand break once an AP-endonuclease or AP-lyase was recruited.

Deaminations within AGCT sequences appeared to be more resistant to processing by UNG2 and were also resistant to faithful BER steps downstream of UNG2 regardless of MMR activity (see Figures 3D and 4). Since transversion mutation at C:G depends on G1-phase UNG2 regardless of proximity to AGCT motifs (19), we propose that slow G1-phase BER within and near AGCT motifs leads to the persistence of G1-phase AP sites into S-phase, where their entry into replication forks induces mutagenic lesion bypass (Figure 6), as we proposed before (19), and/or induces mutagenic processing by the MRN complex (20). In contrast, bypass of AP-sites exposed in Exo I excision patches is likely to occur entirely within G1-phase – without any need for AP sites to persist into S-phase (Figure 6). It's possible, therefore, that MMR-independent versus MMR-dependent phase 1B mutation occur in distinct cell cycle phases. Our model (Figure 6) posits that the same fraction of deaminations within AGCT motifs will undergo BER-induced lesion bypass regardless of MMR activity; in MMR-competent cells they will undergo lesion bypass by either of the phase 1B pathways shown in Figure 6, while in MMR-deficient cells, they will undergo lesion bypass *via* the classic Neuberger pathway (i.e. 'slow BER') only.

Why are deaminated AGCT motifs resistant to uracil-BER?

Why do deaminated AGCT motifs appear to be resistant to uracil BER? Option A: *In vivo,* deaminated AGCT sequences might innately form a structure inaccessible to UNG2 and other BER enzymes. Option B: AID or co-factors (e.g. 14-3-3 proteins, Spt5, RPA or RNA pol II) might preferentially accumulate at AGCT sites in hypermutating B cells (7,28,64–68), thus reducing access by UNG2 and subsequently by AP-endonucleases or AP-lyases. Since AGCT sites form an AID hotspot in both strands, it's conceivable that AID or co-factors continue to be recruited to the opposite strand even after one strand has been deaminated or processed by UNG2. This would enhance the likelihood that deaminations or AP sites created in AGCT motifs in G1-phase persisted into S-phase. Deaminated AGCT motifs were not more resistant to processing by MMR than other deamination sites (Figure 4A). This can be explained by the potential for ExoI to excise uracil(s) distal to the site of initial MutSα-binding, regardless of whether MutSα ever bound to the uracil(s) excised by ExoI. Figure 4D implies that the foot-print of BER-resistance at AGCT sites is about 10.6 bases (i.e. 2X the Figure 4D half-life, which represents one half of a bell curve footprint). This is consistent with the size of the DNA binding grooves in AID homologues APOBEC2 (69) and APOBEC3G (70) (Xiaojiang Chen, personal communication). However, it is also consistent with the size of the single-stranded DNA bubble formed by RNA pol II (~11 bases), which may be particularly prone to stalling at AGCT motifs in *Ig* genes independently of AID (64,68)].

CONCLUSION

In this study, we have refined the Neuberger model of antibody hypermutation, having provided evidence that UNG2-induced phase 1B mutation distal to AGCT hot-spots is MMR-dependent. Furthermore, our refined model potentially explains the previously enigmatic strand bias of phase II mutation. Our refinements are necessary to encompass the remarkable complexity of antibody hypermutation that has emerged since the Neuberger DNA deamination model was first developed.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We gratefully acknowledge the provision of cell sorting facilities and expertise by Sydney Cytometry (Centenary Institute), and the provision of mouse care and breeding by the Centenary Institute Animal Facility. All work was carried out with approval from and oversight by the University of Sydney and Royal Prince Alfred Hospital (RPAH) Animal Ethics/Welfare Committees and by the RPAH Institute Biosafety Committee, in accordance with NSW and Federal Australian legislation.

FUNDING

National Health and Medical Research Council (NHMRC) Project Grants [1067891, 1051820 and 1012291 to C.J.] and a Postgraduate Scholarship [to E.L.]; Australian Postgraduate Awards [to E.T. and G.S.]; Cancer Institute NSW Scholarship [to G.S.]. Funding for open access charge: Centenary Institute Foundation.
Conflict of interest statement. None declared.

REFERENCES

1. Gearhart,P.J. (1983) The effect of somatic mutation on antibody affinity. *Ann. N. Y. Acad. Sci.*, **418**, 171–176.
2. Muramatsu,M., Kinoshita,K., Fagarasan,S., Yamada,S., Shinkai,Y. and Honjo,T. (2000) Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*, **102**, 553–563.
3. Revy,P., Muto,T., Levy,Y., Geissmann,F., Plebani,A., Sanal,O., Catalan,N., Forveille,M., Dufourcq-Labelouse,R., Gennery,A. *et al.* (2000) Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell*, **102**, 565–575.
4. Wang,M., Rada,C. and Neuberger,M.S. (2010) Altering the spectrum of immunoglobulin V gene somatic hypermutation by modifying the active site of AID. *J. Exp. Med.*, **207**, 141–153.
5. Rogozin,I.B. and Diaz,M. (2004) Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. *J. Immunol.*, **172**, 3382–3384.

6. Jolly,C.J., Wagner,S.D., Rada,C., Klix,N., Milstein,C. and Neuberger,M.S. (1996) The targeting of somatic hypermutation. *Semin. Immunol.*, **8**, 159–168.
7. Xu,Z., Fulop,Z., Wu,G., Pone,E.J., Zhang,J., Mai,T., Thomas,L.M., Al-Qahtani,A., White,C.A., Park,S.R. *et al.* (2010) 14-3-3 adaptor proteins recruit AID to 5'-AGCT-3'-rich switch regions for class switch recombination. *Nat. Struct. Mol. Biol.*, **17**, 1124–1135.
8. Wei,L., Chahwan,R., Wang,S., Wang,X., Pham,P.T., Goodman,M.F., Bergman,A., Scharff,M.D. and MacCarthy,T. (2015) Overlapping hotspots in CDRs are critical sites for V region diversification. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E728–E737.
9. Chen,Z., Eder,M.D., Elos,M.T., Viboolsittiseri,S.S., Chen,X. and Wang,J.H. (2016) Interplay between target sequences and repair pathways determines distinct outcomes of AID-initiated lesions. *J. Immunol.*, **196**, 2335–2347.
10. Yeap,L.-S., Hwang,J.K., Du,Z., Meyers,R.M., Meng,F.-L., Jakubauskaitė,A., Liu,M., Mani,V., Neuberg,D., Kepler,T.B. *et al.* (2015) Sequence-intrinsic mechanisms that target AID mutational outcomes on antibody genes. *Cell*, **163**, 1124–1137.
11. Pérez-Durán,P., Belver,L., de Yébenes,V.G., Delgado,P., Pisano,D.G. and Ramiro,A.R. (2012) UNG shapes the specificity of AID-induced somatic hypermutation. *J. Exp. Med.*, **209**, 1379–1389.
12. Di Noia,J.M. and Neuberger,M.S. (2007) Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.*, **76**, 1–22.
13. Dianov,G.L. and Hubscher,U. (2013) Mammalian base excision repair: the forgotten archangel. *Nucleic Acids Res.*, **41**, 3483–3490.
14. Di Noia,J. and Neuberger,M.S. (2002) Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature*, **419**, 43–48.
15. Rada,C., Williams,G.T., Nilsen,H., Barnes,D.E., Lindahl,T. and Neuberger,M.S. (2002) Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr. Biol.*, **12**, 1748–1755.
16. Weill,J.-C. and Reynaud,C.-A. (2008) DNA polymerases in adaptive immunity. *Nat. Rev. Immunol.*, **8**, 302–312.
17. Daly,J., Bebenek,K., Watt,D.L., Richter,K., Jiang,C., Zhao,M.-L., Ray,M., McGregor,W.G., Kunkel,T.A. and Diaz,M. (2012) Altered Ig hypermutation pattern and frequency in complementary mouse models of DNA polymerase zeta activity. *J. Immunol.*, **188**, 5528–5537.
18. Kano,C., Hanaoka,F. and Wang,J.Y. (2012) Analysis of mice deficient in both REV1 catalytic activity and POLH reveals an unexpected role for POLH in the generation of C to G and G to C transversions during Ig gene hypermutation. *Int. Immunol.*, **24**, 169–174.
19. Sharbeen,G., Yee,C.W., Smith,A.L. and Jolly,C.J. (2012) Ectopic restriction of DNA repair reveals that UNG2 excises AID-induced uracils predominantly or exclusively during G1 phase. *J. Exp. Med.*, **209**, 965–974.
20. Larson,E.D., Cummings,W.J., Bednarski,D.W. and Maizels,N. (2005) MRE11/RAD50 cleaves DNA in the AID/UNG-dependent pathway of immunoglobulin gene diversification. *Mol. Cell*, **20**, 367–375.
21. Saribasak,H. and Gearhart,P.J. (2012) Does DNA repair occur during somatic hypermutation? *Semin. Immunol.*, **24**, 287–292.
22. Kadyrov,F.A., Dzantiev,L., Constantin,N. and Modrich,P. (2006) Endonucleolytic function of MutLalpha in human mismatch repair. *Cell*, **126**, 297–308.
23. Li,G.M. (2008) Mechanisms and functions of DNA mismatch repair. *Cell Res.*, **18**, 85–98.
24. Pena-Diaz,J. and Jiricny,J. (2012) Mammalian mismatch repair: error-free or error-prone? *Trends Biochem. Sci.*, **37**, 206–214.
25. Bardwell,P.D., Woo,C.J., Wei,K., Li,Z., Martin,A., Sack,S.Z., Parris,T., Edelmann,W. and Scharff,M.D. (2004) Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice. *Nat. Immunol.*, **5**, 224–229.
26. Langerak,P., Nygren,A.O., Krijger,P.H., van den Berk,P.C. and Jacobs,H. (2007) A/T mutagenesis in hypermutated immunoglobulin genes strongly depends on PCNAK164 modification. *J. Exp. Med.*, **204,** 1989-98
27. Roa,S., Avdievich,E., Peled,J.U., Maccarthy,T., Werling,U., Kuang,F.L., Kan,R., Zhao,C., Bergman,A., Cohen,P.E. *et al.* (2008) Ubiquitylated PCNA plays a role in somatic hypermutation and class-switch recombination and is required for meiotic progression. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 16248–16253.
28. Delbos,F., Aoufouchi,S., Faili,A., Weill,J.C. and Reynaud,C.A. (2007) DNA polymerase eta is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse. *J. Exp. Med.*, **204**, 17–23.
29. Pavlov,Y.I., Rogozin,I.B., Galkin,A.P., Aksenova,A.Y., Hanaoka,F., Rada,C. and Kunkel,T.A. (2002) Correlation of somatic hypermutation specificity and A-T base pair substitution errors by DNA polymerase eta during copying of a mouse immunoglobulin kappa light chain transgene. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 9954–9959.
30. Sharbeen,G., Cook,A.J., Lau,K.K., Raftery,J., Yee,C.W. and Jolly,C.J. (2010) Incorporation of dUTP does not mediate mutation of A:T base pairs in Ig genes in vivo. *Nucleic Acids Res.*, **38**, 8120–8130.
31. Peòa-Diaz,J., Bregenhorn,S., Ghodgaonkar,M., Follonier,C., Artola-Bor·n,M., Castor,D., Lopes,M., Sartori,A.A. and Jiricny,J. (2012) Noncanonical mismatch repair as a source of genomic instability in human cells. *Mol. Cell*, **47**, 669–680.
32. Dingler,F.A., Kemmerich,K., Neuberger,M.S. and Rada,C. (2014) Uracil excision by endogenous SMUG1 glycosylase promotes efficient Ig class switching and impacts on A:T substitutions during somatic mutation. *Eur. J. Immunol.*, **44**, 1925–1935.
33. Schanz,S., Castor,D., Fischer,F. and Jiricny,J. (2009) Interference of mismatch and base excision repair during the processing of adjacent U/G mispairs may play a key role in somatic hypermutation. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 5593–5598.
34. Zlatanou,A., Despras,E., Braz-Petta,T., Boubakour-Azzouz,I., Pouvelle,C., Stewart,G.S., Nakajima,S., Yasui,A., Ishchenko,A.A. and Kannouche,P.L. (2011) The hMsh2-hMsh6 complex acts in concert with monoubiquitinated PCNA and Pol eta in response to oxidative DNA damage in human cells. *Mol. Cell*, **43**, 649–662.
35. Stavnezer,J., Linehan,E.K., Thompson,M.R., Habboub,G., Ucher,A.J., Kadungure,T., Tsuchimoto,D., Nakabeppu,Y. and Schrader,C.E. (2014) Differential expression of APE1 and APE2 in germinal centers promotes error-prone repair and A:T mutations during somatic hypermutation. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 9217–9222.
36. Grigera,F., Wuerffel,R. and Kenter,A.L. (2016) MBD4 facilitates immunoglobulin class switch recombination. *Mol. Cell. Biol.*, doi:10.1128/MCB00316-16 [Epub ahead of print].
37. Rada,C., Ehrenstein,M.R., Neuberger,M.S. and Milstein,C. (1998) Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity*, **9**, 135–141.
38. Frey,S., Bertocci,B., Delbos,F., Quint,L., Weill,J.C. and Reynaud,C.A. (1998) Mismatch repair deficiency interferes with the accumulation of mutations in chronically stimulated B cells and not with the hypermutation process. *Immunity*, **9**, 127–134.
39. Jacobs,H., Fukita,Y., van der Horst,G.T., de Boer,J., Weeda,G., Essers,J., de Wind,N., Engelward,B.P., Samson,L., Verbeek,S. *et al.* (1998) Hypermutation of immunoglobulin genes in memory B cells of DNA repair- deficient mice. *J. Exp. Med.*, **187**, 1735–1743.
40. Rada,C., Di Noia,J.M. and Neuberger,M.S. (2004) Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol. Cell*, **16**, 163–171.
41. Shen,H.M., Tanaka,A., Bozek,G., Nicolae,D. and Storb,U. (2006) Somatic hypermutation and class switch recombination in Msh6(-/-)Ung(-/-) double-knockout mice. *J. Immunol.*, **177**, 5386–5392.
42. Di Noia,J.M., Rada,C. and Neuberger,M.S. (2006) SMUG1 is able to excise uracil from immunoglobulin genes: insight into mutation versus repair. *EMBO J.*, **25**, 585–595.
43. Nakamura,J. and Swenberg,J.A. (1999) Endogenous apurinic/apyrimidinic sites in genomic DNA of mammalian tissues. *Cancer Res.*, **59**, 2522–2526.
44. Cook,A.J.L., Oganesian,L., Harumal,P., Basten,A., Brink,R. and Jolly,C.J. (2003) Reduced switching in SCID B cells is associated with altered somatic mutation of recombined S regions. *J. Immunol.*, **171**, 6556–6564.
45. Refaeli,Y., Van Parijs,L., Alexander,S.I. and Abbas,A.K. (2002) Interferon gamma is required for activation-induced death of T lymphocytes. *J. Exp. Med.*, **196**, 999–1005.
46. Morita,S., Kojima,T. and Kitamura,T. (2000) Plat-E: an efficient and stable system for transient packaging of retroviruses. *Gene Ther.*, **7**, 1063–1066.

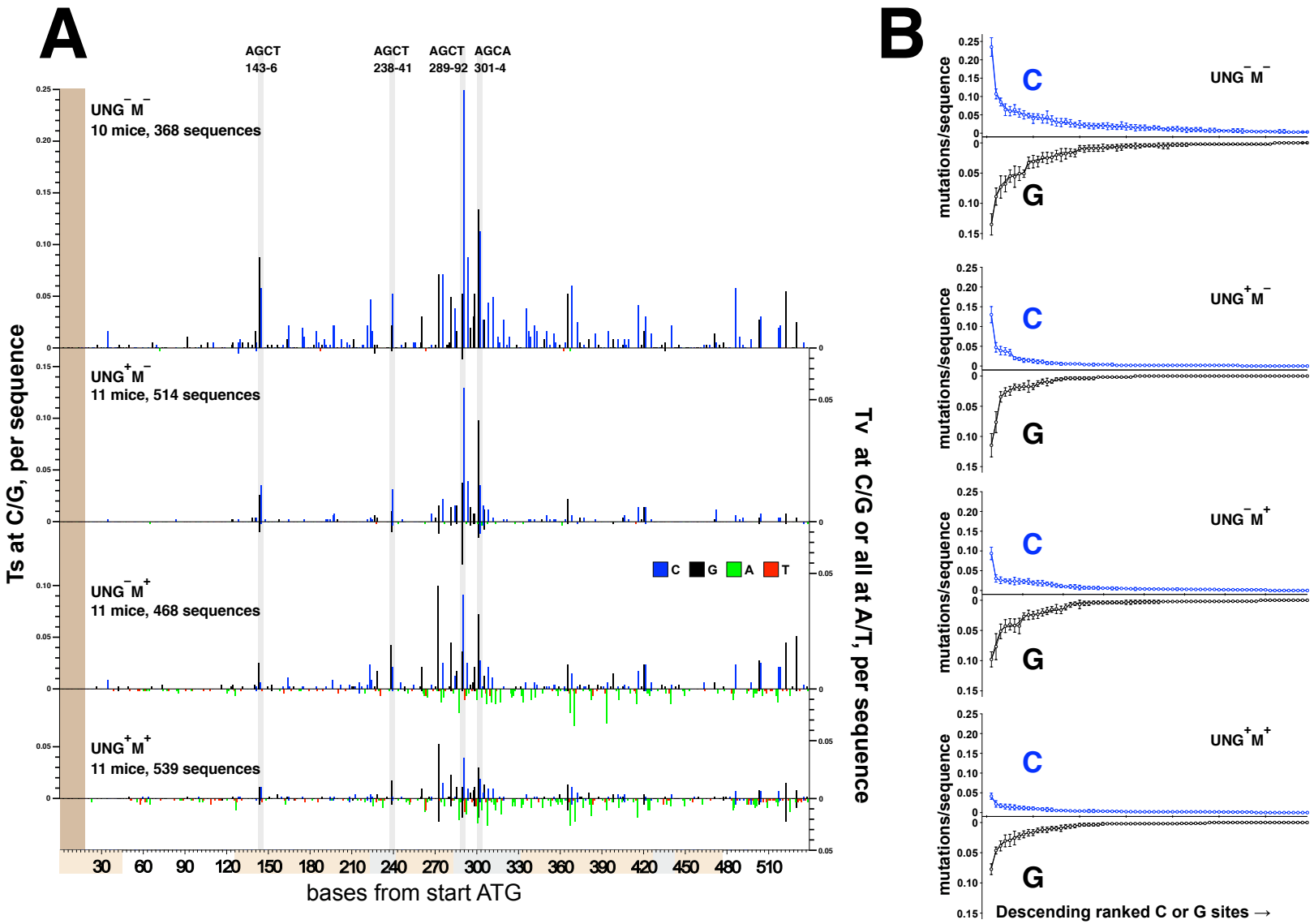
47. Phan,T.G., Amesbury,M., Gardam,S., Crosbie,J., Hasbold,J., Hodgkin,P.D., Basten,A. and Brink,R. (2003) B cell receptor-independent stimuli trigger immunoglobulin (Ig) class switch recombination and production of IgG autoantibodies by anergic self-reactive B cells. *J. Exp. Med.*, **197**, 845–860.
48. Mol,C.D., Arvai,A.S., Sanderson,R.J., Slupphaug,G., Kavli,B., Krokan,H.E., Mosbaugh,D.W. and Tainer,J.A. (1995) Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell*, **82**, 701–708.
49. de Wind,N., Dekker,M., Berns,A., Radman,M. and de Riele,H. (1995) Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination and predisposition to cancer. *Cell*, **82**, 321–330.
50. Phan,T.G., Gardam,S., Basten,A. and Brink,R. (2005) Altered migration, recruitment, and somatic hypermutation in the early response of marginal zone B cells to T cell-dependent antigen. *J. Immunol.*, **174**, 4567–4578.
51. Chan,T.D., Wood,K., Hermes,J.R., Butt,D., Jolly,C.J., Basten,A. and Brink,R. (2012) Elimination of germinal-center-derived self-reactive B cells is governed by the location and concentration of self-antigen. *Immunity*, **37**, 893–904.
52. Xue,K., Rada,C. and Neuberger,M.S. (2006) The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in msh2-/- ung-/- mice. *J. Exp. Med.*, **203**, 2085–2094.
53. Shibutani,S., Takeshita,M. and Grollman,A.P. (1997) Translesional synthesis on DNA templates containing a single abasic site. A mechanistic study of the "A rule". *J. Biol. Chem.*, **272**, 13916–13922.
54. Simonelli,V., Narciso,L., Dogliotti,E. and Fortini,P. (2005) Base excision repair intermediates are mutagenic in mammalian cells. *Nucleic Acids Res.*, **33**, 4404–4411.
55. Phung,Q.H., Winter,D.B., Cranston,A., Tarone,R.E., Bohr,V.A., Fishel,R. and Gearhart,P.J. (1998) Increased hypermutation at G and C nucleotides in immunoglobulin variable genes from mice deficient in the MSH2 mismatch repair protein. *J. Exp. Med.*, **187**, 1745–1751.
56. Frieder,D., Larijani,M., Collins,C., Shulman,M. and Martin,A. (2009) The concerted action of Msh2 and UNG stimulates somatic hypermutation at A . T base pairs. *Mol. Cell. Biol.*, **29**, 5148–5157.
57. Zivojnovic,M., Delbos,F., Girelli Zubani,G., Jule,A., Alcais,A., Weill,J.C., Reynaud,C.A. and Storck,S. (2014) Somatic hypermutation at A/T-rich oligonucleotide substrates shows different strand polarities in Ung-deficient or -proficient backgrounds. *Mol. Cell. Biol.*, **34**, 2176–2187.
58. Wiesendanger,M., Kneitz,B., Edelmann,W. and Scharff,M.D. (2000) Somatic hypermutation in MutS homologue (MSH)3-, MSH6-, and MSH3/MSH6-deficient mice reveals a role for the MSH2-MSH6 heterodimer in modulating the base substitution pattern. *J. Exp. Med.*, **191**, 579–584.
59. Li,Z., Zhao,C., Iglesias-Ussel,M.D., Polonskaya,Z., Zhuang,M., Yang,G., Luo,Z., Edelmann,W. and Scharff,M.D. (2006) The mismatch repair protein Msh6 influences the in vivo AID targeting to the Ig locus. *Immunity*, **24**, 393–403.
60. Kim,N. and Jinks-Robertson,S. (2010) Abasic sites in the transcribed strand of yeast DNA are removed by transcription-coupled nucleotide excision repair. *Mol. Cell. Biol.*, **30**, 3206–3215.
61. Krijger,P., Storb,U. and Jacobs,H. (2010) In: Fugmann,S, Diaz,M and Papavasiliou,N (eds.), *DNA Deamination and the Immune System: AID in Health and Disease*. World Scientific, pp. 97–126.
62. Wilson,T.M., Vaisman,A., Martomo,S.A., Sullivan,P., Lan,L., Hanaoka,F., Yasui,A., Woodgate,R. and Gearhart,P.J. (2005) MSH2-MSH6 stimulates DNA polymerase eta, suggesting a role for A:T mutations in antibody genes. *J. Exp. Med.*, **201**, 637–645.
63. Pettersen,H.S., Sundheim,O., Gilljam,K.M., Slupphaug,G., Krokan,H.E. and Kavli,B. (2007) Uracil-DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms. *Nucleic Acids Res.*, **35**, 3879–3892.
64. Ronai,D., Iglesias-Ussel,M.D., Fan,M., Li,Z., Martin,A. and Scharff,M.D. (2007) Detection of chromatin-associated single-stranded DNA in regions targeted for somatic hypermutation. *J. Exp. Med.*, **204**, 181–190.
65. Pavri,R., Gazumyan,A., Jankovic,M., Di Virgilio,M., Klein,I., Ansarah-Sobrinho,C., Resch,W., Yamane,A., Reina San-Martin,B., Barreto,V. *et al.* (2010) Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell*, **143**, 122–133.
66. Basu,U., Meng,F.-L., Keim,C., Grinstein,V., Pefanis,E., Eccleston,J., Zhang,T., Myers,D., Wasserman,C.R., Wesemann,D.R. *et al.* (2011) The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell*, **144**, 353–363.
67. Yamane,A., Resch,W., Kuo,N., Kuchen,S., Li,Z., Sun,H.W., Robbiani,D.F., McBride,K., Nussenzweig,M.C. and Casellas,R. (2011) Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat. Immunol*,. **12**, 62–69.
68. Maul,R.W., Cao,Z., Venkataraman,L., Giorgetti,C.A., Press,J.L., Denizot,Y., Du,H., Sen,R. and Gearhart,P.J. (2014) Spt5 accumulation at variable genes distinguishes somatic hypermutation in germinal center B cells from ex vivo-activated cells. *J. Exp. Med.*, **211**, 2297–2306.
69. Prochnow,C., Bransteitter,R., Klein,M.G., Goodman,M.F. and Chen,X.S. (2007) The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature*, **445**, 447–451.
70. Holden,L.G., Prochnow,C., Chang,Y.P., Bransteitter,R., Chelico,L., Sen,U., Stevens,R.C., Goodman,M.F. and Chen,X.S. (2008) Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature*, **456**, 121–124.

Supplementary Table 1

Supplementary to Table 1. Raw mutation numbers for each dataset.

bases/sequence	523	124	114	158	127
		MUTATION FROM:			
$UNG^{-}M^{-}$ 10 mice 368 seq's	MUTATION TO:	C	G	T	A
	C	0	3	2	0
	G	1	0	0	0
	T	626	5	0	2
	A	2	357	1	0
	999	629	365	3	2
$UNG^{+}M^{-}$ 11 mice 514 seq's	MUTATION TO:	C	G	T	A
	C	0	43	2	1
	G	8	0	0	5
	T	257	17	0	9
	A	6	179	0	0
	527	271	239	2	15
$UNG^{-}M^{+}$ 11 mice 468 seq's	MUTATION TO:	C	G	T	A
	C	0	5	33	62
	G	6	0	15	99
	T	270	7	0	97
	A	1	324	29	0
	948	277	336	77	258
$UNG^{+}M^{+}$ 11 mice 539 seq's	MUTATION TO:	C	G	T	A
	C	0	53	49	70
	G	25	0	27	111
	T	124	30	0	155
	A	19	166	38	0
	867	168	249	114	336

Supplementary Figure 1

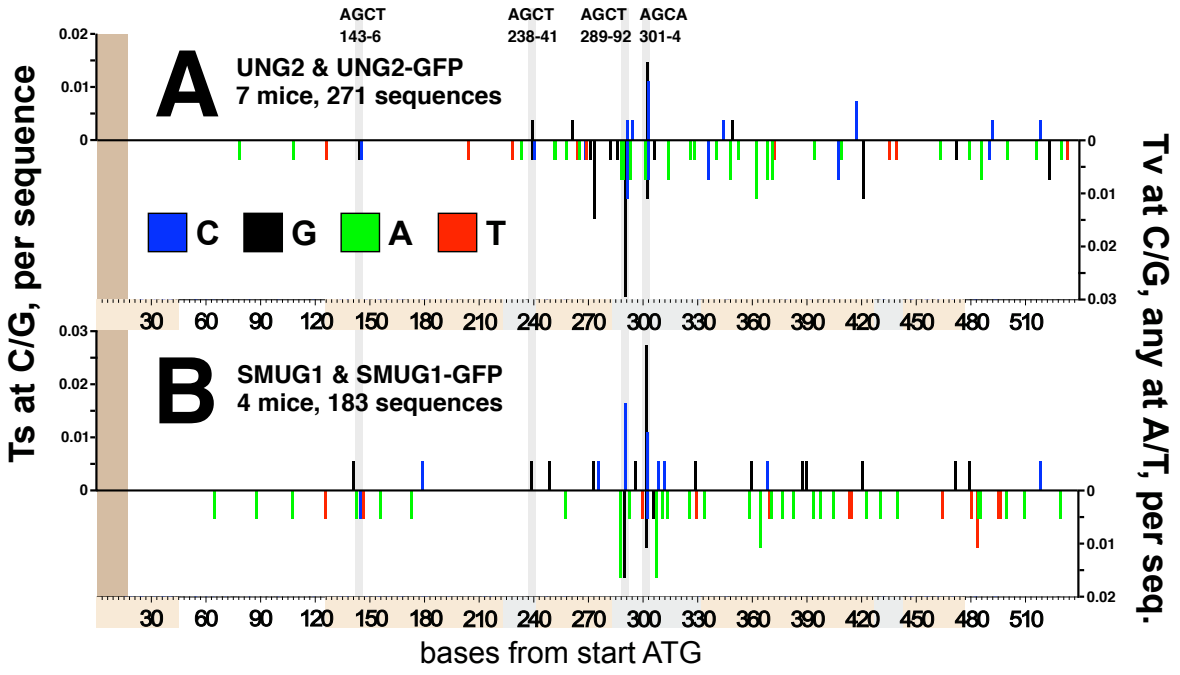


Supplementary Figure 1 (previous page)

Supplementary to Figure 2. (A) Histograms of mutation frequencies at each base in the sequenced window of 523bp. Transitions at C (blue) or G (black) are shown above the x-axis. Transversions at C (blue) or G (black) and mutations at A (green) or T (red) are shown below the x-axis. The brown stripe indicates a 5' region in which mutations were not collated because it over-lapped a secondary PCR primer. Grey dashed lines identify the four AGCW motifs present in the starting sequence (see Results and Discussion). Beige boxes at bottom indicate the L and VJ exons, while grey boxes indicate the three CDRs, according to IMGT (IMGT®, the international ImMunoGeneTics information system® http://www.imgt.org (founder and director: Marie-Paule Lefranc, Montpellier, France)). (B) Reproducibility of hotspot-targeting in the datasets. C (top, blue) or G (bottom, black) mutation sites were sorted as most to least mutated site (i.e. decendingly ranked) in each dataset. The variation between hosts in mean mutation at each sorted site is indicated by the error bars, which depict s.e.m.

Supplementary Figure 2 (below)

Supplementary to Figure 5. Mutation skylines for cells over-expressing (A) UNG2 or UNG2-GFP, or (B) SMUG1 or SMUG1-GFP. Data are presented as in Supplementary Figure 1A.



4.1 R2 overexpression reduces AID-induced mutations and increases retroviral and lentiviral transduction

Introduction

The current model of somatic hypermutation describes a multistep mutagenic process, initiated by AID deaminating cytosine residues to uracil (Frieder et al., 2009), which if left unrepaired, leads to transition mutations at C:G bases after replication. The majority of C deaminations are repaired faithfully by BER and MMR pathways (Thientosapol et al., 2017). However, error-prone BER of uracils that have been converted into apyrimidinic sites introduces transversion and transition mutations at deamination sites with assistance from error prone MMR (Krijger et al., 2009; Thientosapol et al., 2017). Error-prone MMR of U:G mismatches can also introduce mutations at A and T bases in proximity to the initial deamination site (Frieder et al., 2009).

In its usual role, during post replication repair, MMR repairs mismatches in an error-free manner, but is highly mutagenic during somatic mutation. Mutagenesis involves polymerase η, because *Polh* inactivation almost ablates mutations at A:T bases (Delbos et al., 2005; Zeng et al., 2004). Polymerase η is a translesional DNA polymerase that is able to bypass certain DNA lesions that stall replication forks, in an error-prone manner (reviewed in (Weill & Reynaud, 2008)). PCNA ubiquitination at lysine 164 is also involved in MMR-induced mutation, because $Pcna^{K164R/K164R}$ mice,

which are completely resistant to ubiquitination at lysine 164, have substantially decreased mutations at A:T bases (Langerak et al., 2005; Roa et al., 2008). PCNA becomes mono-ubiquitinated at K164 during replication fork stalling, which increases PCNA's affinity for polymerase η.

Why PCNA mono-ubiquitination occurs during AID-induced MMR, to then recruit DNA polymerase η, has not been revealed. It is probable that mutagenic MMR occurs in G1, for the following reasons. (1) Nuclear AID activity is restricted to G1 by nuclear-cytoplasmic shuttling (Wang et al., 2017). (2) The components of MMR known to be required for A:T mutation are also present in G1-phase B cells (Pena-Diaz et al., 2012; Pena-Diaz & Jiricny, 2012; Schrader et al., 2007). (3) Both UNG2-dependent *and* UNG2/MMR co-dependent mutations require UNG2 activity in G1-phase (Sharbeen et al., 2012). (4) In S-phase, when genome replication occurs, dNTP pools are in the range of 10-100pmoles for each dNTP per $x10^6$ cells; in G0/G1, dNTP pools are about 10-fold lower (Rampazzo et al., 2015). An optimal concentration of dNTPs is crucial for the fidelity of DNA synthesis during replication and post replication repair (Rampazzo et al., 2015), while depletion of dNTP pools causes S-phase replication fork stalling (Petermann et al., 2010). For these reasons, we hypothesised that dNTP paucity drives the stalling of the canonical MMR polymerase, polymerase δ, if AID-induced MMR occurs in G1-phase, and that it is this stalling that induces PCNA mono-ubiquitination and the recruitment of error-prone polymerase η (Sharbeen et al., 2010).

dNTP pools are regulated by an interplay of synthesis and degradation (Rampazzo et al., 2010). The rate limiting step in *de-novo* dNTP biosynthesis is catalyzed by the

enzyme ribonucleotide reductase (RNR) (Xu et al., 2008). Ribonucleotide reductase catalyzes the synthesis of dNTPs (Ayinde et al., 2012), by substitution of the 2'hydroxyl group of a ribonucleoside diphosphate (NDP) with hydrogen, resulting in a deoxynucleoside diphosphate (dNDP) (Nordlund & Reichard, 2006). RNR is a cytosolic enzyme maximally active in S-phase cells and consists of the R1 (90kDa) and R2 (45kDa) subunits (Niida et al., 2010). The R1 subunit is present throughout the cell cycle. However, the R2 subunit is undetectable during G1-phase (Bjorklund et al., 1990; Engstrom et al., 1985; Mann et al., 1988). This chapter describes experiments performed to determine whether the cell cycle regulation of the R2 subunit of RNR contributes to error-prone DNA repair downstream of AID in B cells.

Results

Experiments initiated by Daniel Bosnjak, an Honours student in our laboratory, used B cell transduction/adoptive transfer experiments to show that retroviral R2 overexpression in adoptive B cells significantly decreased *Ig* mutation at A:T base pairs (unpublished data). Three isoforms of ribonucleotide reductase R2 subunit (R2) were cloned for retroviral expression by Daniel. These were the canonical form (R2) encoded by mouse gene *Rrm2*, a cell cycle dysregulated R2 mutant (called $R2^{AAN}$), and a p53-inducible R2 paralog that is induced by DNA damage in a manner dependent on p53 (R2b, encoded by mouse gene *Rrm2b*) (Chabes et al., 2004; Tanaka et al., 2000). These R2 variants were co-expressed with GFP *via* an internal ribosome entry site (IRES) using the pMiG retroviral vector (Parijs et al., 1999).

A summary of mutation data collected by Daniel Bosnjak, then extended by this author, is shown in Figure 4.1. All three R2 isoforms significantly inhibited mutation at A:T bases, with normal R2, $R2^{AAN}$ and R2b decreasing mutation at A:T by 67% (p=0.0005), 74% (p=0.0003) and 48% (p=0.0023) respectively. The R2 constructs did not inhibit mutation at C:G. Rather, wild-type R2 significantly increased transitions at C:G by 51% (p=0.0473); the other isoforms did not increase these transitions significantly. There were no significant changes in frequencies of transversion mutations at C:G induced by any R2 isoform (Figure 4.1).

The R2-induced phenotype of increased transition mutation at C:G and reduced mutation at A:T base pairs is a somewhat similar phenotype to mismatch repair inhibition, because a 96% loss of hypermutation at A:T base pairs in adoptive *Msh2*$^{ko/ko}$ SW_{HEL} B cells is accompanied by a 58% increase in transitions at C:G base pairs (see (Thientosapol et al., 2017)), but also by a 39% drop in C:G transversions, which did not occur with R2 overexpression. A key "signature" of MMR inhibition in UNG-proficient B cells is hyper-focusing of AID-induced C:G transitions on AGCW hotspots (Delbos et al., 2007; Rada et al., 1998; Thientosapol et al., 2017). This occurs because the fidelity of uracil BER is markedly increased in the absence of MMR, except at AGCW motifs, which appear inherently resistant to faithful uracil BER regardless of MMR-proficiency. Using the analytical approach developed in the previous chapter (Thientosapol et al., 2017), we found that R2 over-expression did not produce this signature. In UNG-competent cells, mismatch repair deficiency significantly increased C:G mutation within, but not outside AGCW hotspots, as observed by others. On the other hand, R2-overexpression significantly increased C:G mutation outside, but not

within AGCW motifs in wild-type cells (see Figure 4.2). This analysis indicated that R2 over-expression suppressed MMR-induced mutation at A:T base pairs, without suppressing MMR itself.

Western blot data confirms R2 expression during G1 of primary splenocytes that have been transduced with R2-expressing constructs

Western blot was performed to investigate cell cycle regulation of RNR sub-units in B cells and to confirm over-expression of R2 in transduced cells. "*Fucci"* mouse B cells were cultured in 20 µg/mL *S. typhosa* LPS (Sigma-Aldrich) plus 20ng/mL mouse IL-4 (BD Biosciences) for 3 days to induce robust B cell proliferation and were sorted into a G1-phase fraction ($mKO2^{+}mAG^{-}$) and an S/G2/M-phase fraction ($mKO2^{-}mAG^{+}$) by flow cytometry. *Fucci* mice (Sakaue-Sawano et al., 2008) carry two separate transgenes for an orange or a green fluorescent protein that accumulates in a cell cycle-dependent manner as a result of fusion to "degrons", which are sequences encoding E3 ubiquitin ligase substrates. The "*fucci*-red" transgene encodes the mKO2 orange fluorescent protein fused to a degron from the human CDT1 protein (amino acids 30-120); mKO2 accumulates in G_0 and G1-phases and very early S-phase and is actively degraded in other phases. The "*fucci*-green" transgene encodes the green fluorescent protein mAG fused to a degron from the human geminin protein (the last 100 amino acids); mAG accumulates in S, G2 and M-phases and is actively degraded in G_0 and G1-phases. Western blot for R1 and R2 sub-units extracted from sorted cells

is shown in Figure 4.3. R2 levels were beneath detection in G1-phase B cells, but readily detectable in S/G2/M-phase cells.

To evaluate the expression of R2 in mouse B cells transduced with pMiG/R2, I harvested *Fucci*-red transgenic B cells from mice spleens and transduced these cells with pMiG/R2 or pMiG vector control retrovirus, and left cells to proliferate in culture for three days. $GFP^{+}mKO2^{+}$ and $GFP^{+}mKO2^{-}$ populations were then purified by flow cytometry. Note that I used mice that carried the *fucci*-red transgenes, but did not carry the *fucci*-green transgene.

The results in Figure 4.4 show that in G1-phase cells transduced with pMiG vector, R2 levels remained undetectable, but in G1-phase cells transduced with pMiG/R2 R2 was readily detectable. Transduction of B cells with pMiG/R2 retrovirus therefore increased R2 expression in G1-phase cells substantially – to levels that approximated R2 levels in S/G2/M-phase cells transduced with vector alone. In $mKO2^{-}$ cells (i.e. predominantly S/G2/M-phase cells, contaminated with some early G1-phase cells), transduction with pMiG/R2 retrovirus increased R2 expression substantially further, compared to vector-transduced controls. R1 levels across the cell cycle varied far less in B cells than R2 levels, and that over-expression of R2 did not appear to alter the levels of R1. The data indicated that retrovirally expressed R2 remained subject to post-transcriptional cell cycle regulation, but was able to enhance G1-phase R2 accumulation to levels approximating S-phase cells.

Retroviral R2 over-expression increases susceptibility to retroviral and lentiviral transduction

As increased dNTP pools are associated with increased retroviral transduction efficiency (Ravot et al., 2002; H. Zhang et al., 1995), we examined the transduction efficacy of R2 over-expressing cells when secondarily transduced with other retroviruses or lentiviruses.

Splenic B cells were transduced with either pMiG/R2 or empty pMiG vector, then secondarily transduced 48 hours later with retrovirus expressing mKO2. I then measured the fraction of GFP^+ cells (i.e. transduced in the first round of transduction) that were mKO^+ by flow cytometry. Primary R2 transduction significantly enhanced secondary retroviral transduction efficiency (2–way ANOVA p=0.027) (see Figure 4.5). R2 over-expression also enhanced secondary lentiviral transduction. NIH/3T3 cells were transduced with either pMiG/R2 or empty pMiG vector and sorted to greater than 90% GFP positivity, then secondarily transduced after expansion in culture with lentivirus expressing mCherry under confluent and pre-confluent conditions. R2 over-expression enhanced lentiviral transduction in confluent cells 2.5x better than in pre-confluent cells (that is, concurrent cells seeded at 2x lower concentration), (Figure 4.6, 2-way ANOVA $p<0.0001$).

R2 overexpression reduces MMR-independent A:T mutation in *Msh2*$^{-/-}$ B cells

Somatic mutation at *Ig* A:T base pairs is largely, but not entirely, dependent on mismatch repair. A small minority of A:T mutations are generated independently of mismatch repair; in adoptive SW_{HEL} cells this minority represents about 5% of A:T mutations (Thientosapol et al., 2017). They require UNG2 and Pol η and are therefore proposed to arise *via* Pol η-mediated mis-incorporation during LP-BER of deamination sites initiated by UNG2 (Delbos et al., 2007). To test whether MMR-independent A:T mutation could also be reduced by R2-overexpression, we used the SW_{HEL} transduction/adoptive transfer model (Sharbeen et al., 2010; Sharbeen et al., 2012) to measure the influence of R2-overexpression on *Ig* hypermutation on *Msh2*$^{-/-}$ B cells. SW_{HEL} mice carry a gene-targeted VDJ_H-rearrangement and a transgenic Ig_K rearrangement, which together confer B cell and Ig specificity for hen egg lysozyme (HEL) (Cook et al., 2003; Phan et al., 2003). Loss of RAG-recombinase activity in SW_{HEL} *Rag1*$^{-/-}$ mice prevents V(D)J receptor editing, ensuring that all B cells express identical HEL-specific receptors (Cook et al., 2003). *Msh2*$^{-/-}$ SW_{HEL} *Rag1*$^{-/-}$ splenocytes were transiently activated *via* CD40 and transduced with pMiG vectors *ex-vivo.* FACS-purified GFP$^+$ cells were then co-injected with antigen (HEL-conjugated sheep red blood cells, HEL-SRBC) into multiple C57BL/6 congenic hosts that had been primed with non-conjugated SRBC seven days earlier. This induces an immune response to HEL dominated by the adoptive SW_{HEL} B cells in the B cell compartment (see (Sharbeen et al., 2010)). Six days after adoptive transfer, GFP$^+$ HEL-binding cells were recovered from host spleens and sorted one cell per well into

96-well PCR plates. The single *SW$_{HEL}$* VDJ$_H$ allele (and 3'-intronic sequences) present in each well were then amplified by nested PCR. Single cell PCR products were then sequenced by automatic 96-well Sanger sequencing (Macrogen Inc, Korea) and mutations collated. Background mutation using this model is about one mutation per 94 sequences (Sharbeen et al., 2010). The sequencing primer used produced a data collection window of 560bp per sequence (Thientosapol et al., 2017).

Results from nine hosts bearing pMiG/R2-overexpressing *Msh2*$^{-/-}$ cells were compared to ten hosts bearing *Msh2*$^{-/-}$ cells transduced with empty vector (see Figure 4.7). The level of *Igh* mutation at T for day 6 adoptive Msh2$^{-/-}$ SWhel B cells is indistinguishable from the PCR background in our assay, whereas mutation at A remains significantly above background (Sharbeen et al., 2012) (see Figure 4.7). R2 overexpression produced a statistically significant decrease in mutation at A bases, and no significant change in mutation at T bases. R2 overexpression also produced no significant change in overall transition or transversion mutation at C or G (see Figure 4.7), which indicated that the increase in C:G transition mutations that R2 overexpression induces in Msh2$^{+/+}$ (see Figure 4.1) occured via mismatch repair.

It was previously demonstrated in wild-type cells that R2 overexpression significantly increased C:G mutation outside, but not within AGCW motifs, while mismatch repair deficiency significantly increased C:G mutations within, but not outside AGCW hotspots (see Figure 4.2). Analysis of R2 overexpression in *Msh2*$^{-/-}$ germinal centre B cells showed that *Msh2*$^{-/-}$ decreased C:G mutation outside AGCW motifs and increased C:G mutation within AGCW motifs regardless of R2 overexpression (see Figure 4.8).

Further analysis of these C:G mutations showed that *Msh2*$^{-/-}$ decreased C:G transversions outside AGCW and increased C:G transitions within AGCW regardless of R2 overexpression (see Figure 4.8). Also, R2 overexpression suppressed C:G transversions still further in *Msh2*$^{-/-}$ cells, but only outside AGCW motifs.

Discussion

The experiments presented in this chapter showed that R2 levels are cell cycle-regulated in activated B cells; R2 levels were substantially lower in G1-phase B cells than in S/G2/M-phase B cells. R1 levels varied far less across the cell cycle. When splenocytes were transduced with pMiG/R2 retrovirus, R2 was detected in G1, while vector control splenocytes transduced with pMiG had an absence of detectable R2 in G1. Transduction with pMiG/R2 also substantially increased R2 protein in post G1-phase compared to vector control. R1 protein levels were unchanged in G1-phase and post G1-phase in cells transduced with R2 compared to vector controls. Therefore, overexpression of R2 presumably increased RNR levels in G1-phase cells. It's possible that pMiG/R2 also increased RNR levels outside G1-phase, but only if endogenous R1 exists in excess to endogenous R2 in S/G2/M-phase B cells, which seems unlikely.

It has been shown previously that retroviral and lentiviral transduction improves with an increased availability of dNTPs (Ravot et al., 2002; H. Zhang et al., 1995). We used retroviral and lentiviral transduction susceptibility therefore as a proxy to measure dNTP biological availability in cells overexpressing R2. Our experiments indicated that R2 over-expression enhanced retroviral/lentiviral transduction, and that this effect was greater in confluent than in non-confluent cell cultures. Given the known function

of R2 as a subunit of RNR, it is likely therefore that R2 over-expression increased the availability of dNTPs to retroviral/lentiviral reverse transcriptases and DNA polymerases in primary mouse B cells, especially in G0/G1-phase. However, we cannot rule out the possibility that R2 overexpression enhances transduction by some other mechanism.

R2 over-expression significantly reduced mutation at A in *Msh2*$^{ko/ko}$ *SW*$_{HEL}$ B cells (p=0.036), but we were unable to measure an impact on mutation at T, which is already extremely low in *Msh2*$^{ko/ko}$ *SW*$_{HEL}$ B cells, because of background noise. The data suggests that dNTP paucity drives mis-incorporation during UNG-induced LP-BER, producing a small number of UNG2-dependent mutations at A:T base pairs and also at C:G base pairs outside AGCW motifs, in addition to strongly influencing mis-incorporation at both A:T and C:G base pairs during AID-induced MMR. Combined with previous data produced by Daniel Bosnjak (see Figure 4.2) and analysis of mutation within versus outside AGCW hotspots (Figure 4.8), these data suggest that the mechanism by which R2 over-expression suppresses mutation at A:T base pairs is not by direct inhibition of MMR. As R2 over-expression did not increase C:G transition mutation in *Msh2*$^{ko/ko}$ cells, it is unlikely that R2 over-expression inhibits uracil BER, because inhibition of uracil BER substantially increases the frequency of C:G to T:A transition mutation in *Ig* genes (Thientosapol et al., 2017).

In total, our data indicate, but do not prove that reduced production of dNTPs in G1-phase drives much of the error-prone synthesis at A:T base pairs, and to a lesser extent at C:G base pairs in germinal centre B cells, whether mutation is *via* nan-canonical

MMR or long patch BER.

4.2 shRNAs against *SAMHD1* decrease *Ig* mutation at A and T bases in vivo

If over-expression of R2 influenced *Ig* hypermutation *via* an increase in dNTP supply, as proposed in the previous chapter, then treatments that reduce dNTP degradation should have a similar impact. dNTPs are actively degraded in vertebrate cells by the dNTP triphosphate hydrolase SAMHD1 (Goldstone et al., 2011; Powell et al., 2011). The restriction of dNTP supply by SAMHD1 is potently anti-viral, although SAMHD1 also possesses an exonuclease activity that additionally may be important in restricting viruses (Beloglazova et al., 2013; Ryoo et al., 2014). SAMHD1 has also been found to be associated with the RNA exosome (Lim et al., 2015) and participates in HR-mediated DNA repair (Daddacha et al., 2017). Although SAMHD1 is not degraded in a cell cycle-dependent manner akin to R2, it's activity is nonetheless cell-cycle regulated by post-translational modification (Pauls et al., 2014c).

To generate preliminary data on the role of SAMHD1 in *Ig* mutation, we chose to retrovirally express shRNAs targeting mouse *Samhd1* in adoptive SW_{HEL} B cells, as a means to generate data directly comparable to our data on the influence of R2 on *Ig* mutation. *Samhd1*-knockout mice were published after this shRNA project commenced (Rehwinkel et al., 2013). Using these mice would have required importation, rederivation and crossing with SW_{HEL} mice to produce data directly comparable to our existing hypermutation data.

SAMHD1 is present in primary B lymphocytes

SAMHD1 expression was confirmed in primary B cells. Splenocytes from *Fucci* mice (Sakaue-Sawano et al., 2008) were cultured *in-vitro* in the presence of bacterial LPS. G1-phase fraction ($mKO2^{+}mAG^{-}$) and S/G2/M-phase fractions ($mKO2^{-}mAG^{+}$) were obtained by flow cytometric sorting and western blotting for SAMHD1 expression (see Methods: Preparation of cell extracts, western blot analysis). Using the same blot that was used for R1 and R2 western blot (Figures 4.3), SAMHD1 was readily detectable in both G1 and post G1 cell cycle phases (Figure 4.9).

Selection of potent shRNAs targeting *Samhd1* in NIH/3T3 cell lines

Potential 21-mer shRNA constructs targeting SAMHD1 mRNA (Ensemble cDNA ENSMUST000000057725) were designed using the protocol described in (Dow et al., 2012). Identification of potential candidates used a three-step process. An online siRNA prediction tool 'Designer of Small Interfering RNAs-DSIR' (Vert et al., 2006) identified potential sequences. These sequences were then cross-checked against the BLAST database to exclude sequences matching off-target genes. Sequences were then filtered again through a series of seven Sensor exclusion criteria (Dow et al., 2012).

11 *Samhd1* shRNA sequences were synthesized as DNA gene blocks (Integrated DNA Technologies) and cloned into context of natural miRNA mIR30 in the retroviral expression plasmid pLMP (Zhou et al., 2008), along with an additional four commercial

shRNA sequences targeting mouse *Samhd1* and two control sequences (Table 4.1) (see Methods: Creation and validation of shRNA against SAMHD1). Retroviral supernatants were created from these LMP vectors and used to transduce NIH/3T3 cells (see Methods: Creation and validation of shRNA against SAMHD1).

Semi-quantitative western blot analysis for SAMHD1 was performed using extracts from LMP-transduced NIH/3T3 cells (see Methods: Preparation of cell extracts, western blot analysis). Total protein blotting was measured using Memcode protein stain; anti-SAMHD1 signal was then normalized to Memcode signal using Imagelab analysis software (Bio-Rad). Estimated percentage knockdowns of *Samhd1* knockdown is listed in Table 4.1. Knockdown ranged between 8-98% based on this analysis. (see Figure 4.10).

Using these data, western blot was repeated for the most potent shRNAs (#1,2,8 and the commercial shRNA TRCN0000099803) and for a low potency shRNA (#4). This time, closer approximation of equal protein loading was performed, as confirmed by Memcode staining (see Figure 4.10). Percentage knockdowns were similar to the first-round results. From this, the candidates selected for *in vivo* experiments were *Samhd1* shRNA 1 (96-98% knockdown) and TRCN0000099800 (91-93% knockdown).

ID	*Samhd1* shRNA sequence	DSIR score (Higher is better)	Knockdown (%)	Knockdown (Repeat) (%)
1	CACGAGAUAGUGUCUAACAAGA	98.2	98	96
2	CUCGUGUUCACUUCUAUUGUAA	95	93	86
3	ACCAUAAGGUCAUAUAAUUCAA	93.3	78	
4	CUCACUUCUAUUGUAAGAGCAA	93.3	8	78
5	AUGGAACUGUUUGAAAGGAUUA	92.7	47	
6	AGAAGAAGAUGAUAGAAUGUAU	92.7	89	
7	AACCAAUGACCUUUAUUUAGUA	91.9	83	
8	CCACAAUUAAGUUAGGAAAUAU	91.5	90	85
9	AAAAGCGUGAGAAGAUUAGGAA	88.4	N/A	
10	CUCACAAUUAAGUUAGGAAAUA	88.1	86	
11	CUGAACUCAAACUUGUCAUGAA	88	95	
TRCN0000099803	CAAGAAGCAUUAUCUGGUGUAU		73	
TRCN0000099804	CAGAAGACAUUACCUUUAUCAA		31	
TRCN0000099800	AUGGACAUCUUCAGAGACAAUA		91	93
TRCN0000099801	CGAGAAGAUUAGGAAGGAAGAA		75	
CJ171	ACAACAAGATGAAGAGCACCAA		Control	
ET06	ATCTCGCTTGGGCGAGAGTAAG		Control	

Table 4.1: 11 shRNA constructs (IDs 1-11) against *Samhd1* were created as per protocol described in (Dow et al., 2012) and cloned into the mIR30 retroviral expression plasmid LMP1066, along with an additional four commercial shRNA sequences (TRCN0000099800,1,3,4) and two control sequences (CJ171, ET06) (construct 9 was not successfully cloned). Retroviral supernatants were created and used to transduce NIH/3T3 cells. Western blot analyses for Samhd1 were performed from extracts of transduced NIH/3T3 cells cultured for 3 days in LPS. *Samhd1* knockdown were calculated from western blot signal compared to controls (CJ171 and ET06).

Knockdown of *Samhd1* decreases mutations at A and T bases

LMP-derived retroviruses producing *Samhd1* shRNA1 and TRCN0000099800 shRNAs were used in the SW_{HEL} transduction/adoptive transfer model. Two separate SW_{HEL} transduction/adoptive transfer experiments were performed, consisting of three mice each for *Samhd1* shRNA1 and TRCN0000099800 constructs, with a total of 134 and 127 germinal centre SW_{HEL} *Igh* sequences obtained, respectively. Three separate SW_{HEL} transduction/adoptive transfer experiments consisting of four mice for each control shRNA (CJ171 and ET06) were performed and yielded a total of 176 and 162 sequences respectively.

Mutations at A and T bases, which mostly represent nucleotide substitutions due to MMR were decreased in cells expressing *Samhd1* shRNA1 and TRCN0000099800, by 23 and 15%, respectively, relative to the control sequences (see Figure 4.11). Using a non-parametric test (Mann-Whitney), these results were not statistically significant. On a per mouse basis, the Mann Whitney test shows a two-tailed p value of 0.0546. The remaining nucleotide substitutions were also analysed. There was no difference in transition mutations at C:G or transversion mutations at C:G between cells expressing *Samhd1* shRNAs versus control shRNAs.

Discussion

In previously published experiments, SAMHD1 expression was detected by western blot in resting phase human fibroblasts only and was not detected in proliferating human fibroblasts (Franzolin et al., 2013). In a later paper, SAMHD1 was detected in

both G1 and S-phases in human lymphocytes (Yan et al., 2015). In agreement with the latter report, we found that SAMHD1 is expressed consistently throughout the cell cycle in mouse B cells.

T592 phosphorylation of SAMHD1 occurs in S-phase human lymphocytes (Yan et al., 2015); this inhibits SAMHD1 dNTPase activity by preventing the formation of SAMHD1 tetramers. We did not measure SAMHD1 T592 phosphorylation. An antibody specific for T592 phosphorylated human SAMHD1 exists (White et al., 2013b), but it's use on western blots would not have quantified non-phosphorylated dNTPase-active SAMHD1.

Samhd1 knockdown produced a modest decrease in mutation at A:T that was not statistically significant. There were no significant changes in transitions at C:G or tranversions at C:G. These results were modest compared to the results from R2 overexpression, which showed a much more pronounced decrease in mutations at A:T and a significant increase in transitions at C:G. These results suggested that analysis of hypermutation in *Samhd1*-knockout B cells would be worthwhile nonetheless, because the likelihood that incomplete silencing of *Samhd1* produces a strongly "leaky" phenotype. Even low levels of Samhd1 are able to change dNTP pools (Arnold et al., 2015; Hansen et al., 2014), so the residual SAMHD1 we detected in silenced cells may still cleave dNTPs sufficiently to drive high levels of *Igh* A:T mutation. The negative correlation between *Samhd1* shRNA efficacy and levels of A:T mutation in adoptive B cells (see Figure 4.11) also indicated that investigation of *Ig* hypermutation in *Samhd1*-knockout cells would be worthwhile. We proceeded to use CRISPR/Cas9 to induce

frameshift deletions in *Samhd1* in SW_{HEL} embryos in collaboration with Professor Robert Brink.

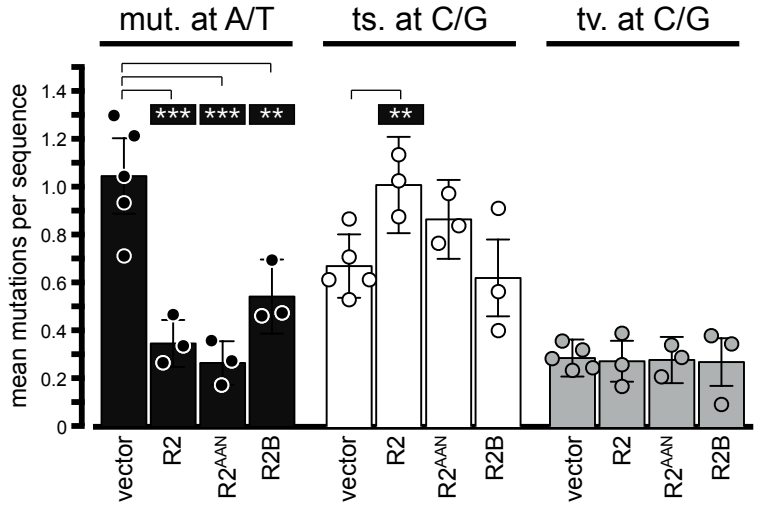


Figure 4.1: Day 6 *Igh* hypermutation in adoptive SW_{HEL} germinal centre B cells retrovirally over-expressing one of three distinct ribonucleotide reductase R2 isoforms compared to pMiG vector controls. Mutations are divided into mutations at **A/T**, transition mutations (**ts**) at C/G and transversion mutations (**tv**) at C/G. Each circle represents the mean mutation (+/- SEM) per sequence · for one mouse. Histograms represent the means (+/- SEM) of mouse means: $^{**}P<0.01$,$^{***}P<0.001$.

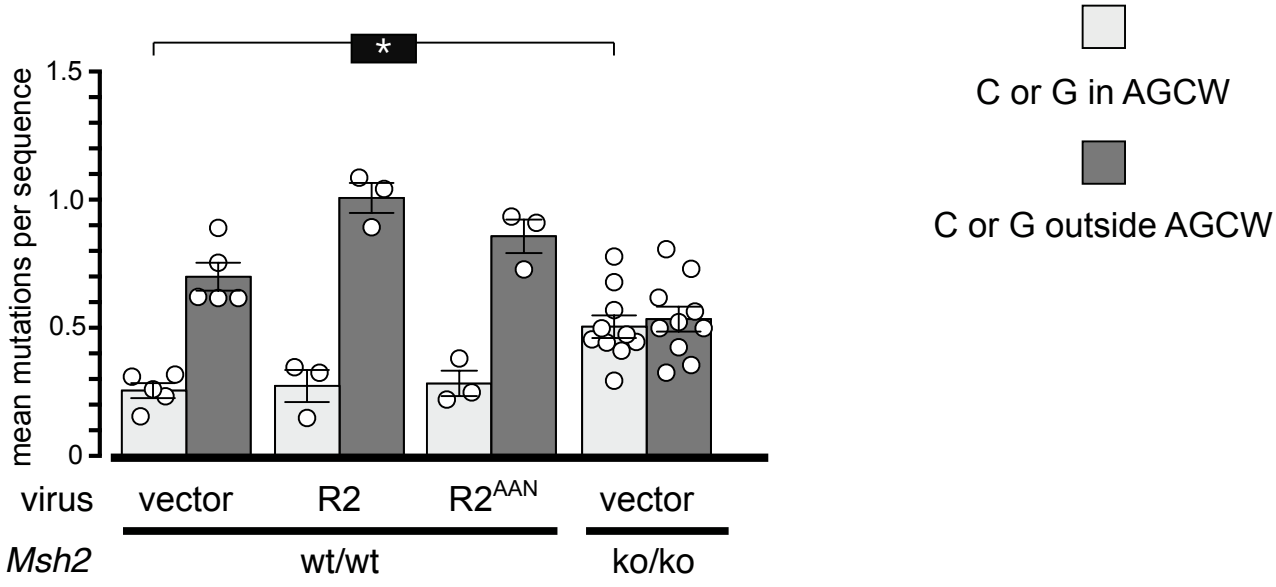


Figure 4.2: Mutation focussing to AGCW hotspots in *Igh* in germinal centre B cells over-expressing R2 or R2^{AAN} in *Msh2*$^{+/+}$ (wt) mice compared to vector transduced *Msh2*$^{-/-}$ (ko) B cells. Each circle represents the mean mutation (+/- SEM) per sequence for one mouse. Histograms represent the means for mice: *P<0.05.

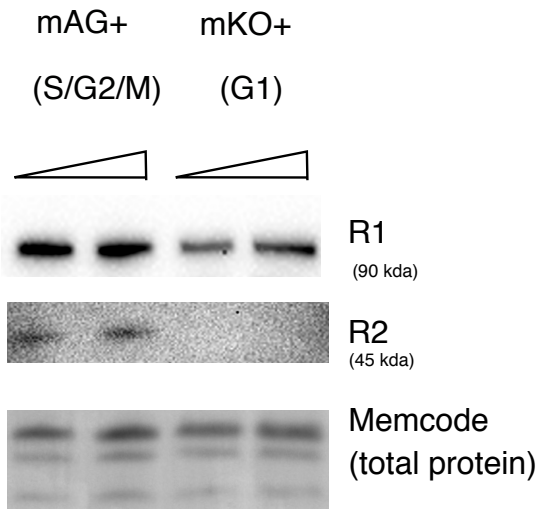


Figure 4.3: Western blot for the R1 (90kda) and R2 (45kda) subunits of ribonucleotide reductase extracted from *Fucci* splenic B cells cultured for 3 days in the presence of bacterial LPS. Cells were sorted for $mKO^{+}mAG^{-}$ (G1-phase) and $mKO^{-}mAG^{+}$ (S/G2/M-phase) populations. R1, anti-R1 signal; R2, anti-R2 signal; Memcode total protein stain (10-20kda bands).

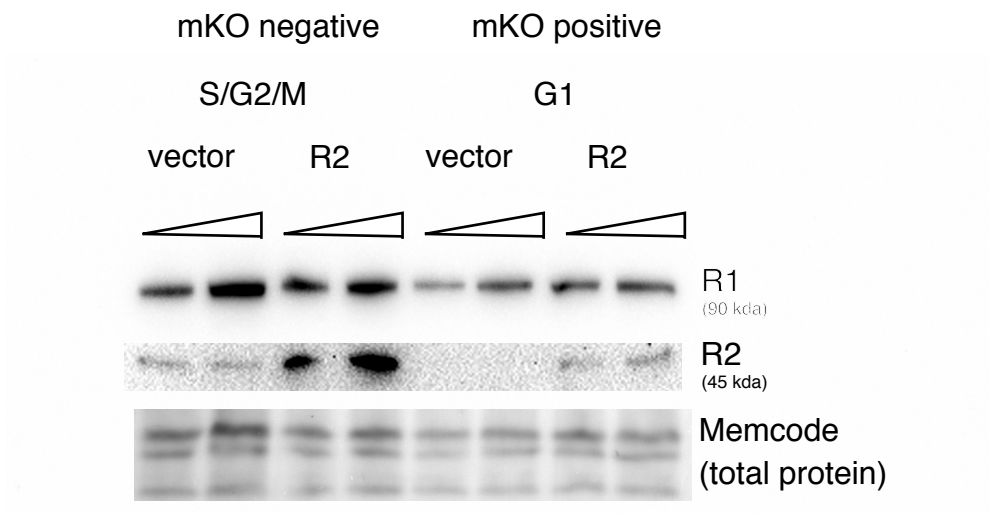


Figure 4.4: Western blot for the R1 (90kda) and R2 (45kda) subunits of ribonucleotide reductase extracted from *Fucci-red* splenic B cells transduced with retrovirus co-expressing GFP and R2 subunit or GFP reporter only virus and cultured for 3 days in the presence of bacterial LPS. Cells were sorted into $GFP^{+}mKO^{-}$ and $GFP^{+}mKO^{+}$ populations. R1, anti-R1 signal; R2, anti-R2 signal; Memcode total protein stain (10-20kda bands).

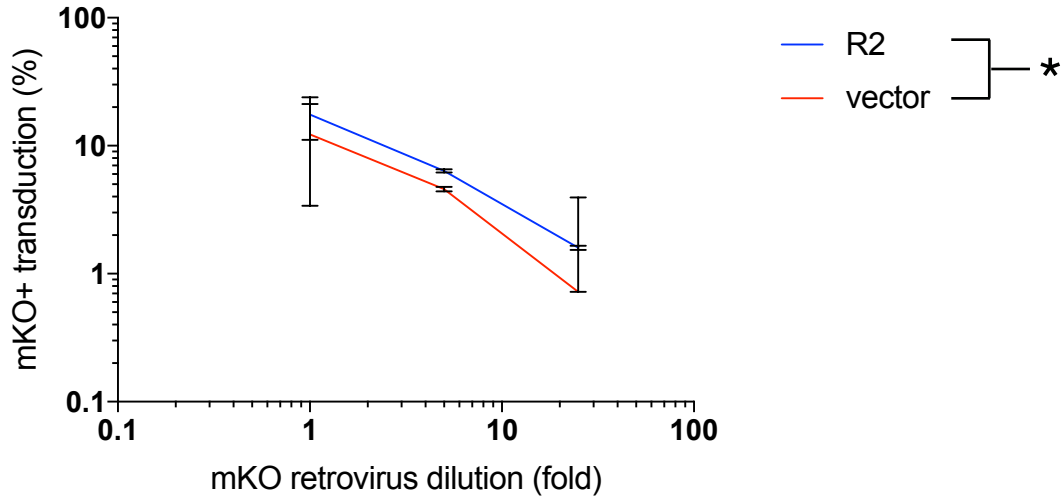


Figure 4.5: Retroviral transducibility conferred by R2 overexpression. Primary mouse splenocytes were transduced with retrovirus with pMiG/R2 or MiG vector to express R2 plus GFP or GFP alone, respectively, then secondarily transduced 48 hours later with serially diluted retrovirus to express mKO. Plots show the frequency of mKO^+ cells amongst GFP^+ cells 48h after secondary transduction, as determined by flow cytometry:(*P=0.027 two-way ANOVA, error bars = 95%CI).

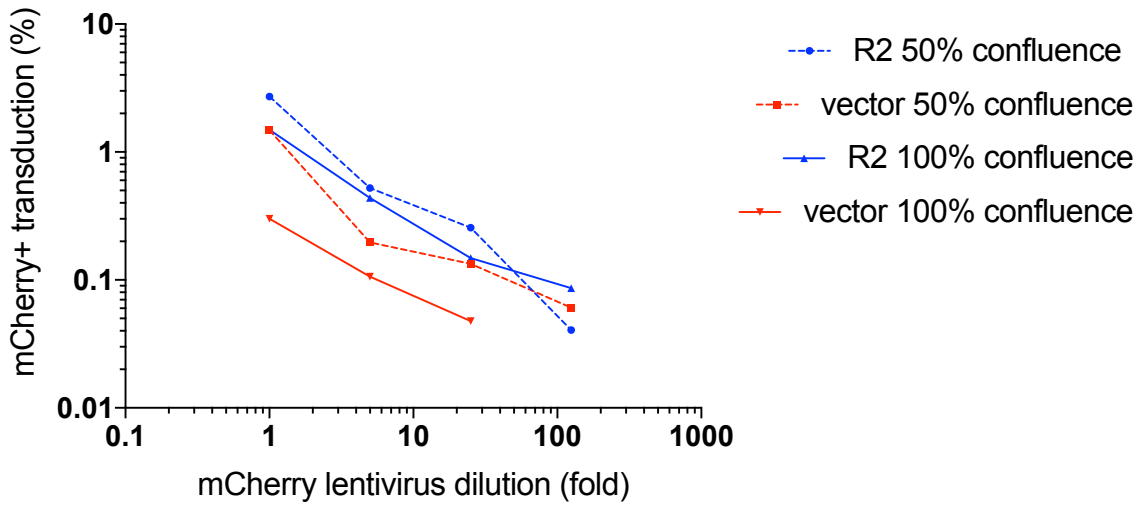


Figure 4.6: Lentiviral transducibility conferred by R2 overexpression in confluent and pre-confluent NIH/3T3 cells. NIH/3T3 cells were transduced with retrovirus to co-express GFP and R2 or GFP alone and flow-cytometry sorted to be >90% GFP^+. After expansion in duplicate cultures at 2-fold different starting concentrations, they were secondarily transduced with serially diluted lentivirus expressing mCherry; lentiviral transduction occurred at the point of confluence for one culture ("100% confluence") and prior to confluence for the other ("50% confluence"). Plots show the frequency of $mCherry^+$ cells amongst GFP^+ cells 48h after secondary transduction, as determined by flow cytometry. R2 over-expression increased the frequency of subsequent lentiviral transduction (i.e. frequency of $mCherry^+$ cells) to a 2.5x greater extent in confluent cells than in pre-confluent cells (two-way ANOVA $P<0.0001$).

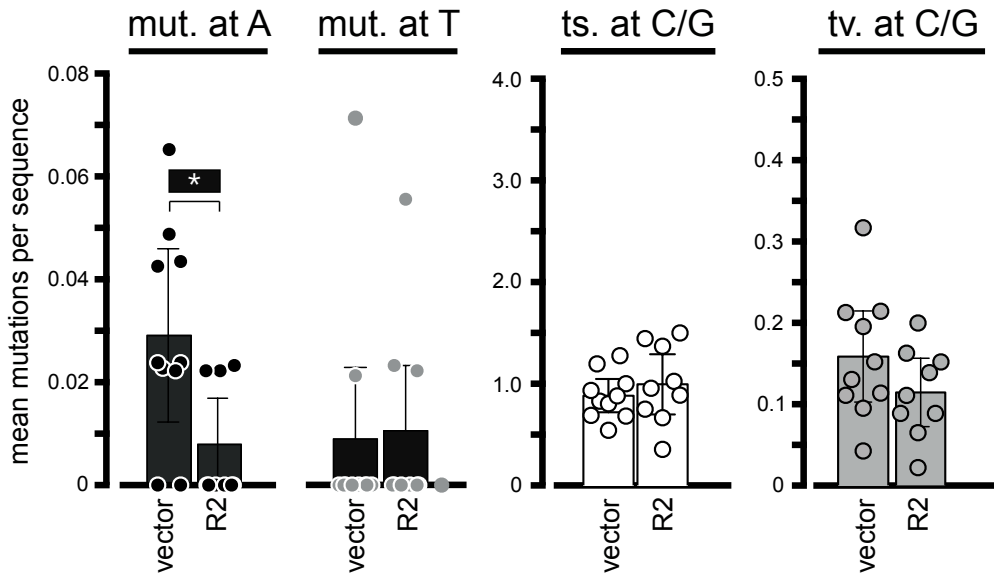


Figure 4.7: Day 6 *Igh* hypermutation in adoptive SW_{HEL} $Msh2^{-/-}$ germinal centre B cells transduced with retrovirus to overexpress R2 subunit of ribonucleotide reductase or transduced with empty vector. Results are divided into mutations at **A**, mutations at **T**, transition mutations (**ts**) **at C/G** and transversion mutations (**tv) at C/G**. Each circle represents the mean mutation (+/-SEM) for sequences for one mouse. Histograms represent the means (+/-SEM) of mouse means: **P*<0.05.

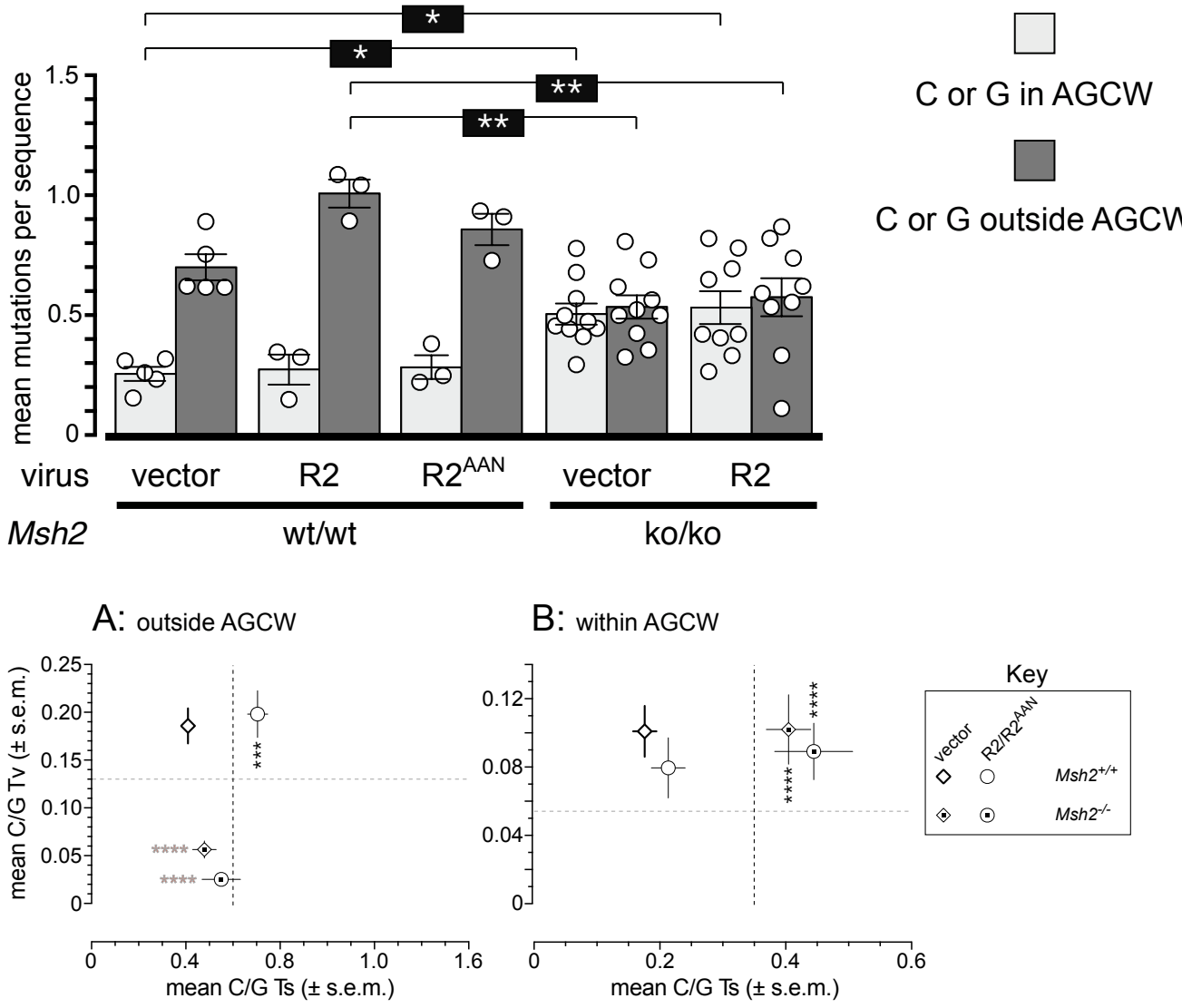


Figure 4.8: Focusing of C/G mutation to AGCW motifs in adoptive *SW$_{HEL}$Msh2$^{-/-}$* germinal centre B cells transduced to express the R2 subunit of ribonucleotide reductase or with empty vector. The top figure shows all forms of mutation at C:G within or outside AGCW motifs. Each circle represents the mean mutation per sequence for one mouse. Histograms represent the means (±s.e.m.) of mouse means: **P*<0.05, ***P*<0.01. The bottom figure shows the same data simplified in a 2D plot. x-axes: Transition (**Ts**) mutations at C/G outside (left) or within (right) AGCW motifs. y-axes: Transversion (**Tv**) mutations at C/G outside (left) or within (right) AGCW motifs. Significant differences to the vector-transduced Msh2$^{+/+}$ means are indicated with asterisks (one-way ANOVA, Holm–Sidak's post hoc tests): vertical and black for C/G transitions; horizontal and gray for C/G transversions. All means lying to the right of the vertical dotted line represent significantly elevated C/G transition frequencies relative to the Msh2$^{+/+}$ data. Data from cells overexpressing R2 or R2^{AAN} were pooled. ***P<0.001, ****P<0.0001.

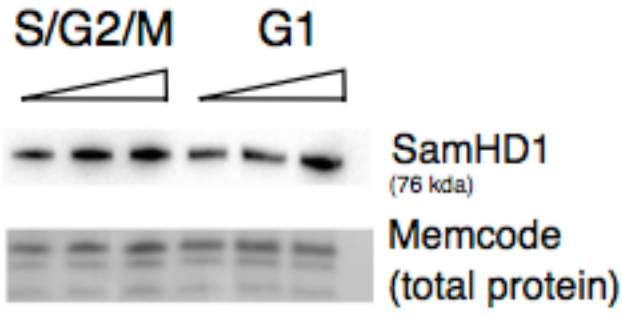
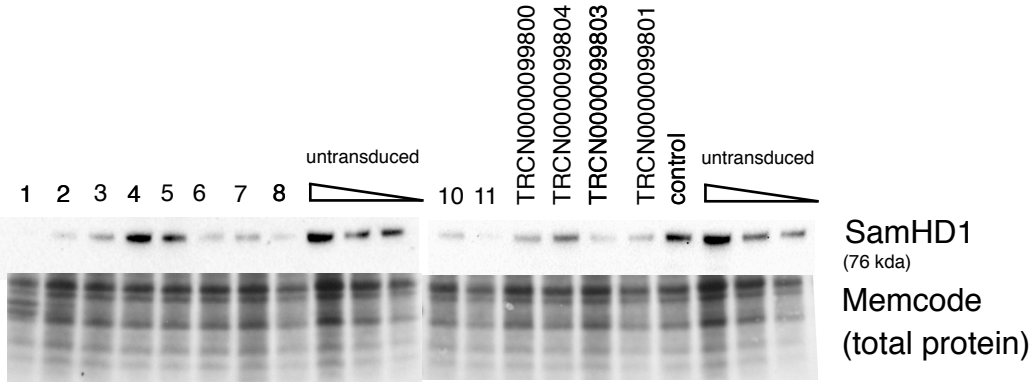


Figure 4.9: Western blot for SAMHD1 (76kda) extracted from *Fucci* splenic B cells cultured for 3 days in the presence of bacterial LPS. Cells were sorted for $mKO^{+}mAG^{-}$ (G1-phase) and $mKO^{-}mAG^{+}$ (S/G2/M-phase) populations. SamHD1, anti-SAMHD1 signal; Memcode total protein stain (10-20kda bands).

A



B

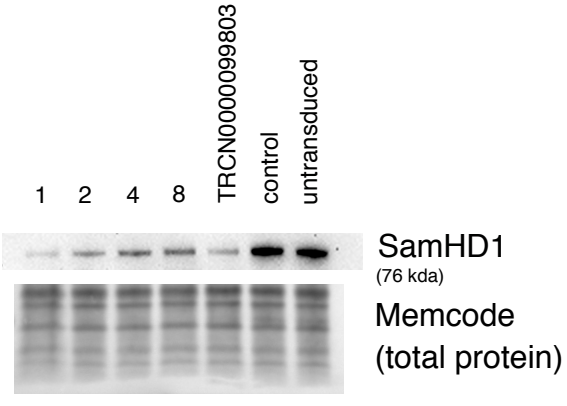


Figure 4.10 (A) Western blot for Samhd1 (76kda) extracted from NIH/3T3 cells transduced with LMP retrovirus expressing shRNAs against *Samhd1.* NIH/3T3 cells were cultured for 3 days after transduction. The shRNAs expressed by vectors 1-11 were designed in our laboratory, TRCN 00009980-4 were designed by the shRNA consortium and published by Sigma-Aldrich. Memcode staining (10-30kda window) represents protein loading. **(B) Repeat Western blot with better approximation of equal protein loading using selected samples.**

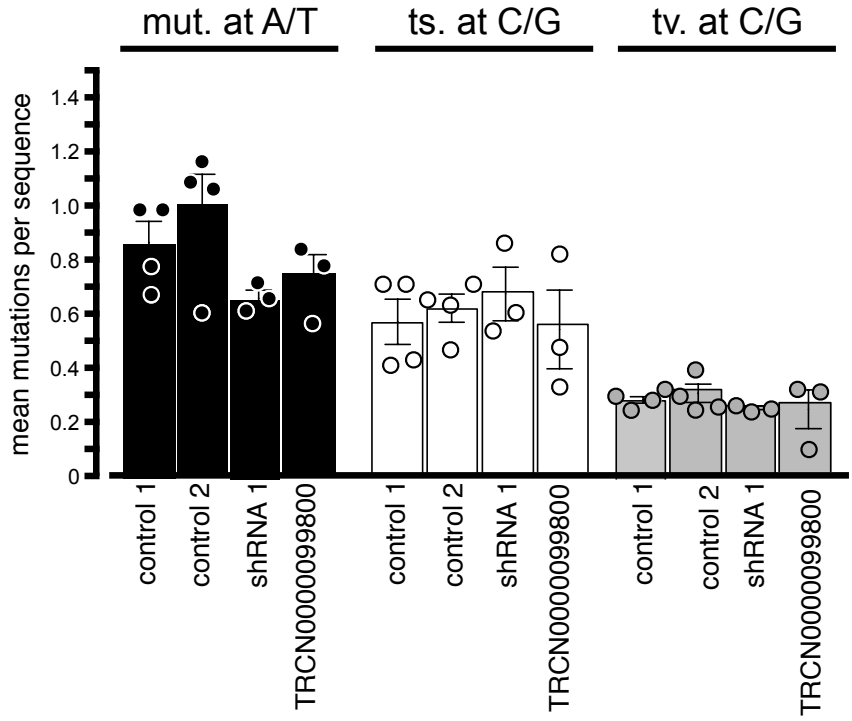


Figure 4.11: Day6 *Igh* hypermutation in adoptive SW_{HEL} germinal centre B cells transduced with retrovirus expressing shRNAs against *Samhd1.* Results are divided into mutations at **A/T**, transition mutations (**ts**) **at C/G** and transversion mutations (**tv) at C/G**. Each circle represents the mean mutation for sequences for one mouse. Histograms represent the means (+/-SEM) of means for all mice (control 1, non-sliencing shRNA; control 2, luciferase control).

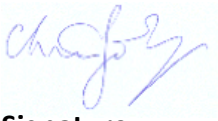
5. SAMHD1 enhances immunoglobulin hypermutation by promoting transversion mutation

Thientosapol E S, Bosnjak D, Durack T, Stevanovski I, van Geldermalsen M, Holst J, Jahan Z, Shephard C, Weninger W, Kim B, Brink R, Jolly C J. SAMHD1 enhances immunoglobulin hypermutation by promoting transversion mutation. *Proceedings of the National Academy of Sciences of the United States of America* May 8, 2018. 115 (19) 4921-4926; published ahead of print April 18, 2018. https://doi.org/10.1073/pnas.1719771115.

Author Contributions:

C.J.J. designed research; Experiments from Figure 1A, 1B, 2B and 3 were performed by E.S.T. with assistance by D.B., T.D., I.S., M.v.G., Z.J. and C.J.J., Experiments from 1C were performed by C.S.; J.H., W.W., B.K., and R.B. contributed new reagents/analytic tools; E.S.T. and C.J.J. analyzed data; and E.S.T. and C.J.J. wrote the paper.

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Chris Jolly		27/6/2018
Supervisor Name	**Signature**	**Date**

SAMHD1 enhances immunoglobulin hypermutation by promoting transversion mutation

Eddy Sanchai Thientosapol[a,b], Daniel Bosnjak[a,b], Timothy Durack[a,b], Igor Stevanovski[a,b], Michelle van Geldermalsen[a,b], Jeff Holst[a,b], Zeenat Jahan[a,b], Caitlin Shepard[c], Wolfgang Weninger[a,b], Baek Kim[c,d], Robert Brink[e], and Christopher J. Jolly[a,b,1]

[a]Centenary Institute, Camperdown NSW 2050, Australia; [b]Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia; [c]Emory School of Medicine, Emory University, Atlanta, GA 30322; [d]School of Pharmacy, Kyung Hee University, 02447 Seoul, South Korea; and [e]Garvan Institute of Medical Research, Darlinghurst, Sydney, NSW 2010, Australia

Edited by David G. Schatz, Howard Hughes Medical Institute and Yale University School of Medicine, New Haven, CT, and accepted by Editorial Board Member Robert L. Coffman March 28, 2018 (received for review November 15, 2017)

Activation-induced deaminase (AID) initiates hypermutation of *Ig* genes in activated B cells by converting C:G into U:G base pairs. G_1-phase variants of uracil base excision repair (BER) and mismatch repair (MMR) then deploy translesion polymerases including REV1 and Pol η, which exacerbates mutation. dNTP paucity may contribute to hypermutation, because dNTP levels are reduced in G_1 phase to inhibit viral replication. To derestrict G_1-phase dNTP supply, we CRISPR-inactivated SAMHD1 (which degrades dNTPs) in germinal center B cells. *Samhd1* inactivation increased B cell virus susceptibility, increased transition mutations at C:G base pairs, and substantially decreased transversion mutations at A:T and C:G base pairs in both strands. We conclude that SAMHD1's restriction of dNTP supply enhances AID's mutagenicity and that the evolution of *Ig* hypermutation included the repurposing of antiviral mechanisms based on dNTP starvation.

B cells | mutation | DNA repair | dNTPs | deamination

Somatic hypermutation, in combination with clonal selection, creates high-affinity antibodies (Ig) in response to viral and bacterial infection. Activation-induced deaminase (AID, gene *Aicda*) initiates mutation by deaminating genomic cytosines, in either strand of *Ig* V(D)J-regions, which creates a U:G mismatch in the DNA (Fig. S1). If replication proceeds without uracil excision, this produces a C:G to T:A transition mutation in one daughter cell, because A is incorporated opposite U (reviewed in ref. 1).

Most deaminations are excised by the base excision repair (BER) enzyme UNG2, a uracil-specific enzyme that converts U into an apyrimidinic (AP) site (Fig. S1). The enzymes SMUG1 and TDG can substitute for UNG2, but in UNG2-proficient cells, their contribution to *Ig* hypermutation is probably minor (2–5). Mismatch repair (MMR) is also recruited (Fig. S1), presumably by the U:G mismatches deamination induces. Processing via BER or MMR predominantly restores C in place of U (6–9) but is also error-prone. Error-prone BER converts deamination sites into transition or transversion mutations at C/G and occasionally produces mutations at bases flanking deamination sites, including A/T bases (1). Mutation does not require the canonical BER polymerase, Pol β (10); instead, C/G to G/C transversion mutation requires the translesion polymerases REV1 and Pol η (11, 12). Bypass of persistent AP sites using translesion polymerases is therefore thought to be the main mechanism of UNG2-dependent *Ig* hypermutation (12, 13).

Mutation at *Ig* A:T base pairs is substantially reduced by inactivation of the MMR genes *Msh2, Msh6*, or *Exo1,* inactivation of *Polh*, or by mutation of the K164 ubiquitination site in mouse PCNA (14–19). This has led to a model in which A/T mutations are introduced by translesion Pol η, predominantly recruited by noncanonical MMR (1, 20), with minor recruitment during UNG2-driven long patch (LP) BER (21). In this model (Fig. S1), activated MutSα recruits ExoI to create excision patches. Monoubiquitinated PCNA then recruits error-prone Pol η during excision patch in-fill, which introduces mutations biased toward A:T base pairs. Until recently, it was unknown how the DNA was nicked to allow ExoI entry, but it is now clear that APE1, APE2, and MutLα produce nicks semiredundantly (4, 5, 22) (Fig. S1).

A major question raised by this model is why PCNA becomes ubiquitinated to recruit Pol η during AID-induced MMR. Only G_1-phase activities of AID or UNG2 are mutagenic in mouse B cells (5, 6, 23). Strong evidence for mutagenic interaction between BER and MMR implies that mutagenic MMR also occurs in G_1 phase (2, 3, 8, 9, 12, 21). dNTP levels were reported to be lowest in G_1 phase, especially in the nucleus (24). We proposed, therefore, that dNTP pools may be inadequate to support long-patch DNA synthesis by conventional DNA polymerases during G_1-phase MMR; polymerase stalling could then induce PCNA ubiquitination (ref. 25; also see ref. 20). The rate-limiting enzyme for dNTP production is ribonucleotide reductase (RNR), which is downregulated in G_1-phase cells via turnover of the R2 subunit (26). However, the major regulator of dNTP levels is not in fact RNR activity, but the deoxynucleotide triphosphohydrolase SAMHD1; dNTPs increase more than 10-fold in resting and proliferating *Samhd1*$^{-/-}$ cells (27, 28). SAMHD1 is a potent restrictor of virus

Significance

Antibody affinity rises during immune responses to viruses via antibody gene somatic hypermutation and Darwinian selection of mutated B cells—in a time frame of days. The enzyme activation-induced deaminase (AID) initiates hypermutation by deaminating genomic cytosines. Mutation is exacerbated by noncanonical G_1-phase DNA repair pathways that deploy error-prone polymerases, including Pol η (gene *Polh*). In G_1 phase, dNTP levels are restricted to inhibit viral replication. We derestricted G_1-phase dNTP supply in hypermutating B cells, which increased virus susceptibility in vitro and caused changes in antibody hypermutation in vivo akin to *Polh* inactivation. We conclude that G_1-phase dNTP paucity contributes to antibody hypermutation and that the evolution of antibody hypermutation included the repurposing of intracellular antivirus mechanisms based on dNTP starvation.

Author contributions: C.J.J. designed research; E.S.T., D.B., T.D., I.S., M.v.G., Z.J., C.S., R.B., and C.J.J. performed research; J.H., W.W., B.K., and R.B. contributed new reagents/analytic tools; E.S.T. and C.J.J. analyzed data; and E.S.T. and C.J.J. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. D.G.S. is a guest editor invited by the Editorial Board.

Data deposition: The somatic *Igh V*-region sequences reported in this paper have been deposited in the GenBank database (accession nos. MH198782–MH200576).

[1]To whom correspondence should be addressed. Email: c.jolly@centenary.org.au.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1719771115/-/DCSupplemental.

Published online April 18, 2018.

replication (27, 28), and its activity is highest in G_1 phase (29). We quantified *Ig* hypermutation in mouse germinal center B cells lacking SAMHD1. We observed an increase in transitions at C/G and major decreases in transversions at A/T and C/G, which suggests that SAMHD1-induced dNTP paucity contributes to AID-induced mutagenesis via both MMR and uracil BER.

Results

RNR and SAMHD1 Expression in Mouse B Cells. Ribonucleotide reductase (RNR) consists of R1 and R2 subunits, which form a catalytic site at R1/R2 interfaces (reviewed in ref. 30). Consistent with reports for other cell types (30), R2 levels were substantially lower in G_1-phase B cells [i.e., in $mKO2^{+ve}mAG^{-ve}$ (orange) *Fucci*-transgenic B cells; ref. 31] than in the rest of the cell cycle [i.e., in $mKO2^{-ve}mAG^{+ve}$ (green) *Fucci*- transgenic B cells; Fig. 1*A*], implying that B cells carry out far less de novo deoxynucleotide synthesis in G_1 phase than in S phase. SAMHD1 levels were not obviously cell cycle regulated in mouse B cells (Fig. 1*A*), which was consistent with posttranslational modification being the principal regulator of SAMHD1 activity (29).

***Samhd1* Inactivation Substantially Decreased Transversions in *Ig* Genes.** We CRISPR-targeted *Samhd1* exon 1 in C57BL/6 mouse embryos. *Samhd1* alleles carrying 25- or 41-bp frame-shift deletions ("Δ") in exon 1 (Fig. S2) appeared to be knockout alleles (Fig. 1*A*). CRISPR mutation of *Samhd1* enhanced retroviral transduction ($P = 0.0042$; Fig. 1*B*) and substantially increased the levels of dNTPs in resting and blasting B cells ($P < 0.0001$, $P = 0.0003$, respectively) and in G_1-phase cells (Fig. 1*C*). Purines (dATP and dGTP) were increased more than pyrimidines (dTTP and dCTP; Fig. 1*C*). These transduction and dNTP data closely reproduced data generated using conventional $Samhd1^{-/-}$ mice (28, 32, 33).

We four times back-crossed $Samhd1^{\Delta}$ alleles to *Ig*-transgenic "SW_{HEL}" mice, which have a C57BL/6 background, then bred $Samhd1^{\Delta 25/\Delta 25}SW_{HEL}$ mice (and later, $Samhd1^{\Delta 41/\Delta 41}SW_{HEL}$ mice). To quantify *Ig* hypermutation, we ex vivo transduced SW_{HEL} B cells, which are specific for hen egg lysozyme (HEL), to express the PSB2 uracil glycosylase inhibitor and/or GFP then adoptively transferred GFP-positive cells into congenic hosts and immunized with HEL conjugated to SRBC, as outlined in Fig. 2*A* and described previously (6, 9, 25). Adoptive cells are T-dependent and concentrate in germinal centers soon after transfer (25, 34). *Igh* mutations were collated in a 560-bp window from ≤47 single HEL-binding GFP^{+ve} cells per host, sorted from spleens 6 d after adoptive transfer, using a minimum of three donors and hosts per treatment (*Methods*). This procedure results in reproducible hypermutation with minimal impact on the mutation spectrum by antigen selection, because the BCR analyzed starts with high affinity for HEL and the duration of hypermutation (i.e., 6 d) is short (9). Our model predicted that loss of SAMHD1 would decrease mutation at A:T base pairs, and we indeed found this to be the case ($P < 0.0001$ for mutation at A; $P = 0.029$ for mutation at T; Fig. S3*A*). Specifically, *Samhd1* inactivation reduced the frequency of transversion mutations at A:T base pairs (hereon called "A/T transversions"), relative to wild-type cells ($P < 0.0001$; Fig. 2 *B*, *i*), but barely changed the frequency of A/T transitions (Fig. 2 *B*, *ii*). *Samhd1* inactivation also substantially reduced C/G transversions ($P < 0.0001$; Fig. 2 *B*, *iii*) and nearly doubled C/G transitions ($P = 0.0041$; Fig. 2 *B*, *iv*). This increase in C/G transitions partly compensated for the loss of A/T and C/G transversion mutations; total *Ig* mutation in $Samhd1^{\Delta/\Delta}$ B cells was lower than in wild-type cells (Table 1), but not significantly so ($P = 0.07$ by one-way ANOVA). See Table 1 and Figs. S3 and S4 for overviews of the mutation data. *Samhd1* inactivation reduced all classes of C/G transversions. That is C/G to G/C transversions, which are introduced by REV1 via BER and by Pol η via MMR (11, 12, 35), and C/G to A/T transversions, which are introduced by a polymerase or polymerases currently unknown (Table 1 and Fig. S3*A*).

To test whether the impact of *Samhd1* inactivation on *Ig* hypermutation was an artifact of antigen selection, we quantified mutation in intron sequences immediately 3′ to the VDJ_H exon (Fig. S3*B*). Our 560-bp sequence window included the 5′ 101 bp of the J-C_H intron (Fig. S4); this region is analogous to that commonly analyzed in Peyers patch B cells (5, 36)—albeit, with the SW_{HEL} intron carrying a deletion that tags gene-targeted cells (34). *Samhd1* inactivation caused significant reductions in transversions at A/T ($P = 0.0089$) and at C/G ($P = 0.0001$) in the J-C_H intron and increased C/G transitions, albeit not significantly ($P = 0.065$; Fig. S3*B*), relative to wild-type cells. We conclude that the impact of *Samhd1* inactivation on *Ig* hypermutation—in particular, reduced transversion mutation—was not an artifact of antigen selection.

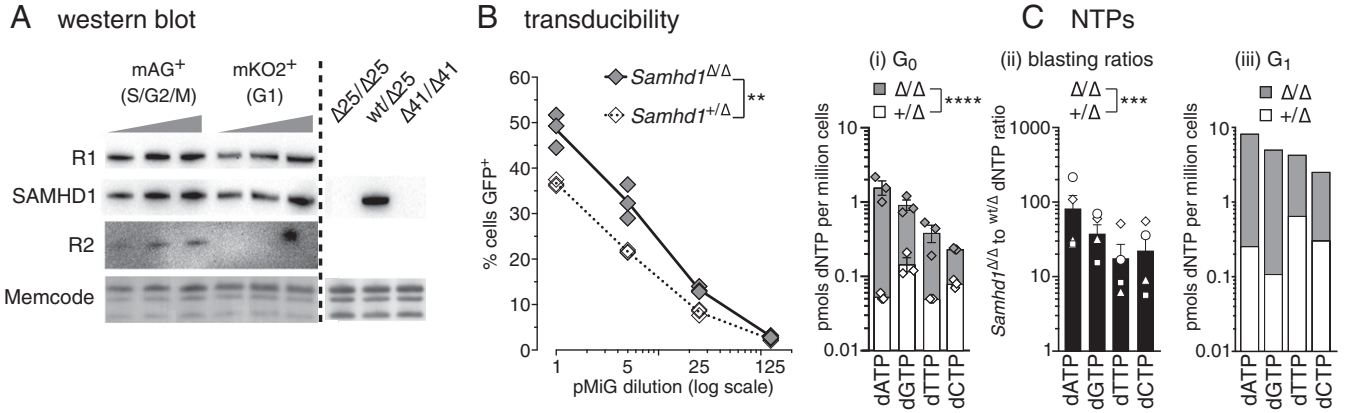


Fig. 1. (*A*) Western blot for RNR subunits (R1 = 90 kDa; R2 = 45 kDa) or SAMHD1 (76 kDa) extracted from splenic B cells after culture for 3 d in LPS. *Fucci*-red/*Fucci*-green transgenic B cells were sorted into green (mAG^+mKO2^-) and orange (mAG^-mKO2^+) populations. Memcode staining (~10- to 20-kDa region) indicates protein loading and transfer. In *Fucci* cells, $mKO2^+$ cells are in G_1 phase and mAG^+ cells are in S, G_2, and M phases (6, 60). (*B*) Transducibility of $Samhd1^{\Delta/\Delta}$ B cells. Splenocytes were cultured with LPS for 1 d to activate B cells, then transduced in triplicates with serially diluted pMiG retrovirus. Frequencies of GFP^+ cells were determined by cytometry 2 d later. $^{**}P = 0.0042$, two-way ANOVA. (*C*) dNTPs measured in whole-cell extracts. (*C*, *i*) dNTPs per 10^6 resting splenocytes, measured in three independent experiments. $^{****}P < 0.0001$, two-way ANOVA. (*C*, *ii* and *iii*) Splenocytes were activated with LPS, and dNTPs were extracted from live cells purified by sedimentation over Histopaque 1083 2 d after activation (*ii*). The ratio of $Samhd1^{\Delta/\Delta}$ dNTPs to $Samhd1^{w.t./\Delta}$ dNTPs is shown for four independent experiments. $^{***}P = 0.0003$, two-way ANOVA. (*C*, *iii*) One day after LPS activation, cells were transduced to express mKO2-cdt fusion protein, which tagged $G_{0/1}$ cells with orange fluorescence (6). Blasting $mKO2^+$ cells were purified by flow cytometry 2 d later.

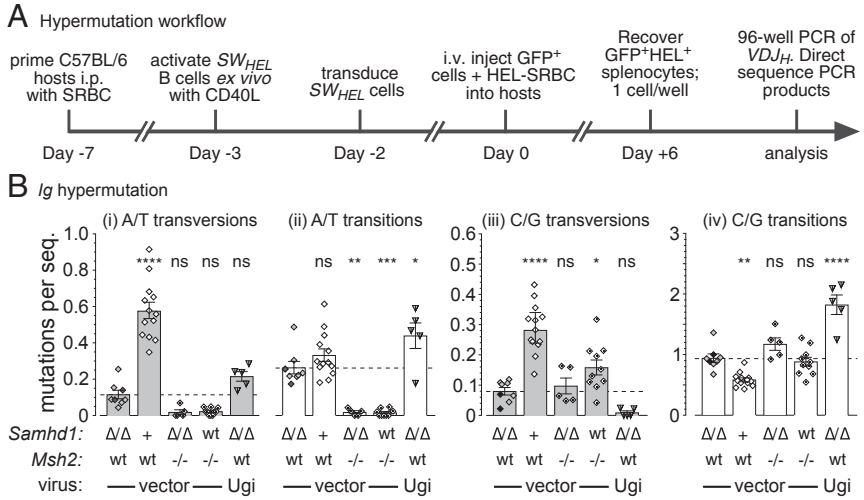


Fig. 2. *Igh* hypermutation in transduced germinal center B cells. (*A*) Experiment design. (*B*) Each symbol represents the mean *Igh* mutations (in a 560-bp sequence window) for ~45 transduced cells sorted from one host mouse. Histograms represent the means (±SEM) of hosts. Mutations are divided into transversions at A:T (*i*), transitions at A:T (*ii*), transversions at C:G (*iii*), and transitions at C:G (*iv*). Symbols: white, $Samhd1^{wt}$ (thin border) or $Samhd1^{+/\Delta 25}$ (bold border); gray, $Samhd1^{\Delta 25/\Delta 25}$; black; $Samhd1^{\Delta 41/\Delta 41}$; black dot, $Msh2^{-/-}$. Transduced with: diamonds, pMiG; triangles, pM/ugi-GFP. Significant differences (one-way ANOVA with Holm-Sidak's post hoc tests) to the vector-transduced $Samhd1^{\Delta 25/\Delta 25}$ means (i.e., the dashed line) are indicated: *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$; ns, not significant.

***Samhd1* Inactivation Does Not Phenocopy MMR Inhibition.** Like *Samhd1* inactivation, *Msh2* knockout decreases A/T mutations and C/G transversions in SW_{HEL} B cells (9) (Fig. 2, Table 1, and Fig. S3). This raised the possibility that *Samhd1* inactivation acted by partially inhibiting AID-induced MMR. A hallmark feature of MMR knockout is the focusing of AID-induced mutation on hotspot motifs, especially AGCW motifs (see ref. 9 and Table S1). We therefore examined our mutation data for hotspot focusing (Fig. 3) and measured hypermutation in adoptive $Msh2^{-/-}Samhd1^{\Delta 25/\Delta 25}$ SW_{HEL} B cells (Figs. 2*B* and 3, Table 1, and Figs. S3 and S4). In summary, we found that *Samhd1* inactivation did not focus AID-induced mutation on AGCW hotspots, an outcome quite different to loss of MMR (see overviews in Table S1).

The strand bias of mutation changes further distinguished $Samhd1^{\Delta/\Delta}$ from $Msh2^{-/-}$ cells. As described before (e.g., ref. 9), $Msh2^{-/-}$ altered mutations at C nearly twice as much as at G. In contrast, *Samhd1* inactivation altered mutation at C and at G to comparable extents (Fig. S3*A*).

***Samhd1* Inactivation Does Not Inhibit Uracil Repair via UNG2.** Reduced C/G transversion in $Samhd1^{\Delta/\Delta}$ B cells raised the obvious possibility that uracil excision by UNG2 might partially depend on SAMHD1, via some unexpected mechanism. To measure UNG2-mediated uracil repair, we blockaded UNG2 activity in adoptive $Samhd1^{\Delta/\Delta}$ cells using retroviral expression of the uracil glycosylase inhibitor (ugi) from bacteriophage PSB2 (37). Previous blockades of UNG activity in wild-type SW_{HEL} B cells with ugi increased *Igh* C/G transitions 2.3-fold, in a strand-unbiased manner, presumably by increasing "ignorant" replication opposite persistent uracils (6, 9). In $Samhd1^{\Delta/\Delta}$ B cells, retroviral expression of ugi-GFP increased C/G transitions by 1.9-fold ($P < 0.0001$), relative to expression of GFP alone (Fig. 2 *B*, *iv*), and this increase was strand-unbiased (Table 1 and Fig. S3*A*). This demonstrated that UNG2 mediates reversion of similar frequencies of U:G base pairs into C:G base pairs in hypermutating *Igh* genes regardless of SAMHD1 activity, so it is unlikely SAMHD1 alters uracil excision rates by UNG2.

Discussion

***Samhd1* Inactivation Partially Phenocopies Genotypes That Block Pol η Recruitment.** We, and later others, proposed polymerase stalling induced by dNTP imbalances as a G_1-phase mechanism that could explain PCNA ubiquitination, Pol η recruitment, and nucleotide misincorporation during AID-induced MMR (20, 25). Our finding here of reduced A/T mutation in $Samhd1^{\Delta/\Delta}$ cells (Fig. S3*A*) was predicted by this model but does not prove it. The unequal impact of *Samhd1* inactivation on A/T transversions versus A/T transitions (Fig. 2*B* and Fig. S3*B*) was not expected but might be explained by the greater impact of *Samhd1* inactivation on purine dNTPs over pyrimidine dNTPs (see below and Fig. 4). *Samhd1*-CRISPR also decreased C/G transversions

Table 1. Raw mutation data

						No. of mutations											
						From G (124 bases) to:			From C (131 bases) to:			From A (135 bases) to:			From T (170 bases) to:			TOTAL
Cells	Protein	No. of Hosts	No. of Sequences	No. of Deletions*	No. of Insertions*	A	C	T	T	G	A	G	T	C	C	A	G	Per sequence
WT or $Samhd1^{+/\Delta}$	GFP	13	565	9,2,1	0	190	74	35	142	26	27	125	170	87	64	44	28	1.79
$Samhd1^{\Delta/\Delta}$	GFP	8	354	1	1	191	12	4	144	10	2	60	15	7	32	11	8	1.40
$Msh2^{-/-}Samhd1^{\Delta/\Delta}$	GFP	5	213	1,2,15,5,1	0	129	11	3	122	2	5	3	1	2	1	1	0	1.31
$Msh2^{-/-}$	GFP	10	446	10,3	0	183	40	18	210	5	7	5	5	3	2	2	0	1.08
$Samhd1^{\Delta/\Delta}$	ugi-GFP	5	218	0	0	235	0	1	163	0	1	61	20	12	34	6	8	2.48

*The size (bp) of all deletions or insertions observed is listed, separated by commas.

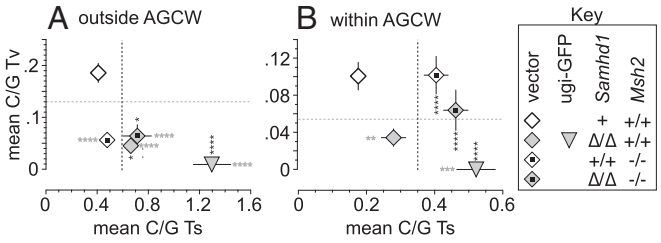


Fig. 3. Occurrence of C/G transitions (*x* axes) or transversions (*y* axes) outside or within AGCW motifs. The mean (±SEM) number of C/G transitions or transversions per 560-bp sequence were calculated on a per host basis, as in Fig. 2, for C/G located outside AGCW motifs (*A*) or C/G located inside AGCW motifs (*B*). Significant differences to the vector-transduced *Samhd1*$^{+}$ means are indicated with asterisks (one-way ANOVA, Holm–Sidak's post hoc tests): vertical and black for C/G transitions; horizontal and gray for C/G transversions. All means lying to the right of the vertical dotted line represent significantly elevated C/G transition frequencies, and all means lying below the gray dotted line represent significantly reduced C/G transversion frequencies, relative to the *Samhd1*$^{+}$*Msh2*$^{+/+}$ data.

and increased C/G transitions in a strand unbiased manner that was not focused on AGCW hotspots (Figs. 2 and 3, Figs. S3 and S4, and Table S1). We largely ruled out reduced uracil excision via MMR or UNG2 as the cause of this. Reduced A/T mutations decoupled from an increase in focus on AGCW hotspots was observed previously in *Polh*$^{-/-}$ and *Pcna*$^{K164R/K164R}$ mice (21, 38), and contrasts with MMR deficiency, which couples these two mutation characteristics, regardless of *Samhd1, Pcna*, or *Polh* genotype (Fig. 3; see also phenotypes summarized in Table S1). The contrast in hotspot focusing among *Samhd1*$^{\Delta/\Delta}$, *Polh*$^{-/-}$, and *Pcna*$^{K164R/K164R}$ on the one hand, and *Msh2*$^{-/-}$, *Msh6*$^{-/-}$, and *Exo1*$^{-/-}$ on the other, can be explained by the likelihood that the former genetic manipulations do not prevent mismatch excision and only impact excision patch in-fill, while the latter manipulations prevent both. We interpret the similarities in the hypermutation phenotypes of *Polh*$^{-/-}$, *Pcna*$^{K164R/K164}$, and *Samhd1*$^{\Delta/\Delta}$ B cells as consistent with *Samhd1* inactivation reducing PCNA ubiquitination and Pol η recruitment during AID-induced MMR, without reducing the production of mismatch excision patches. Nonetheless, our data do not prove this model. Differences in the precise frequencies of mutation subclasses in *Polh*$^{-/-}$, *Pcna*$^{K164R/K164}$, and *Samhd1*$^{\Delta/\Delta}$ B cells (summarized in Table S1) might be ascribed to differences in the polymerases recruited downstream of AID, as well as the unique dNTP milieu in *Samhd1*$^{\Delta/\Delta}$ cells. For instance, Pol δ and Pol κ are likely to be recruited to in-fill MMR excision patches in *PCNA*$^{K164R/K164R}$ and *Polh*$^{-/-}$ B cells, respectively (18).

Unlike the *Msh2*$^{-/-}$, *PCNA*$^{K164R/K164R}$, and *Polh*$^{-/-}$ mutations, *Samhd1* inactivation did not suppress A/T transitions, although it suppressed A/T transversions (Fig. 2*B* and Fig. S3*B*). This may reflect the fact that *Samhd1* inactivation skewed the dNTP pool toward purines (Fig. 1*C*), which would increase bias toward incorporation of dA and dG, regardless of the polymerase used. Since mismatch excision downstream of AID is heavily skewed to the top strand (5, 9), this would result in bias toward incorporation of dA and dG opposite bottom-strand dT during AID-induced MMR (Fig. 4). Thus, the A/T mutation phenotype of *Samhd1*$^{\Delta/\Delta}$ B cells can theoretically be explained by altered polymerase recruitment, as a consequence of better dNTP supply, combined with bias toward top-strand purine incorporation during MMR. Would A/T mutation disappear altogether in a hypothetical situation where G_1-phase dNTP pools perfectly mimicked S phase? We doubt it: The corequirement for UNG2 and MMR for almost half of *Ig* C/G transversions implies that UNG2-derived AP sites are encountered by polymerases during AID-induced MMR (9, 39, 40). This could induce a baseline of polymerase stalling and PCNA ubiquitination regardless of dNTP supply. The role of Ub-PCNA in AID-induced MMR might therefore be explained by a combination of both polymerase encounter with AP sites and SAMHD1-induced dNTP paucity.

The hypermutation phenotypes of *Ung*$^{-/-}$, *Rev1*$^{-/-}$, and *Polb*$^{-/-}$ B cells suggest that most C/G transversions are created by translesion bypass of AP sites, with a minority (~15%) created by simple misincorporation at C/G during MMR, long-patch BER, or single-nucleotide BER (10, 13, 35). We speculate that the substantially reduced C/G transversion and increased C/G transition mutation in *Samhd1*$^{\Delta/\Delta}$ cells is again most simply explained by bias toward purine misincorporation over pyrimidine misincorporation—this time during UNG2-induced lesion bypass events, as illustrated in Fig. 4. It is possible that SAMHD1's exonuclease activity and/or its capacity to recruit the endonuclease CtIP (41, 42) influences *Ig* mutation, rather than its dNTPase activity. The production of double-strand breaks (DSB) downstream of AID is critical for *Ig* class switching (1), but DSB formation plays a minor role in V-region hypermutation (43–45). In our experiments, insertions and deletions, which likely arise from DSB, were not discernably affected by *Samhd1* inactivation (Table 1) and clustered near AGCW motifs regardless of *Samhd1* genotype (Fig. S4).

An Entirely G_1-Phase Model of *Ig* Hypermutation. We suggested previously that MMR-independent lesion bypass might occur when G_1-phase AP sites persist into S phase and are replicated (6; see Fig. S1). This model is not explicitly contradicted by the *Samhd1*$^{\Delta/\Delta}$ phenotype we report here, because *Samhd1* inactivation produced supraoptimal dNTPs in all cell cycle phases. However, we recently identified an ~20-bp footprint flanking AGCW sites within which C/G transversions occurred independently of MMR (9). In remarkable convergence with this footprint size, Woodrick et al. (46) showed that LP-BER produces excision patches centered on the excised AP site that are ≤20 nucleotides long; they further showed that LP-BER predominates over single-nucleotide BER in living cells. AGCW motifs are the only hotspots where AID can deaminate either DNA strand with comparable efficiency in vivo (8, 9, 47). This leads us to update the *Ig* hypermutation model and propose that MMR-independent C/G transversions occur via simultaneous G_1-phase LP-BER of deaminations that have accumulated in

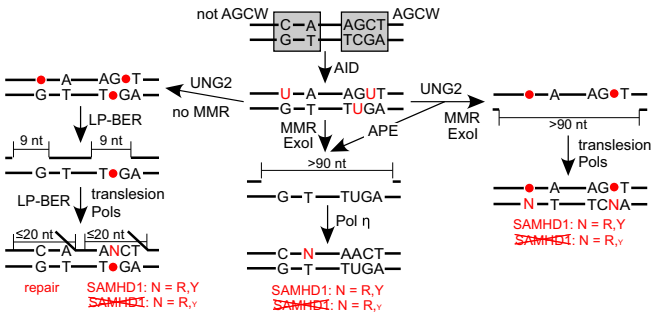


Fig. 4. A completely G_1-phase model of *Ig* hypermutation that incorporates SAMHD1 and a major role for LP-BER. If both strands of an AGCW motif are deaminated, then LP-BER in one strand can expose an AP site in the other, whereas single-nucleotide BER cannot. At deamination sites >10 bases away from AGCW motifs, the chance that excision during LP-BER will expose another AP site in the opposite strand is much lower. Recruitment of ExoI via MutSα produces much longer excision patches (>90 bases long; ref. 61) than LP-BER's 20-base excision patch (46). In UNG2^{+} cells, these long excision patches have a roughly equal chance of exposing an AP site in the opposite strand regardless of where the deamination site that recruited MutSα lies. SAMHD1 activity determines the dNTP milieu, thus dictating the misincorporation spectrum for both MMR and LP-BER. This model explains (*i*) the hyperfocusing of *Ig* mutation on AGCW motifs that occurs only in UNG2^{+} MMR-deficient B cells and not in UNG2/MMR double-deficient cells (9, 21), (*ii*) the near complete MMR dependence of C/G transversion mutation at sites >10 bases from AGCW motifs (9), (*iii*) increased C/G transitions, and (*iv*) substantially decreased A/T and C/G transversions in *Samhd1*$^{\Delta/\Delta}$ B cells (this work).

opposite strands of AGCW sites, as illustrated in Fig. 4, without any requirement for AP sites to persist into S phase. A key strength of this model is its ability to explain why UNG2 is at its most error-prone in regions where AGCW motifs are concentrated.

Similarities Between *Ig* Hypermutation and Intracellular Virus Restriction Mechanisms. Cell cycle turnover of RNR and the dNTPase activity of SAMHD1 are intracellular antiviral mechanisms that evolved long before the appearance of adaptive immunity (48). The APOBEC3 family of AID paralogues also restrict viral infection by deaminating viral genomes; UNG2 and APE activities downstream of APOBEC3 can contribute to viral restriction (49, 50). AID/APOBEC-like deaminases occur across metazoa, dictyosteliida, and even algae, and may have evolved in an arms race with viruses and retro elements (51). Here, we've shown that SAMHD1 is another virus restriction factor that contributes to the mutagenicity of AID—by promoting transversion mutation downstream of AID. Transition mutation is the most common form of point mutation in the genome. Codon degeneracy has evolved to accommodate this, which means that transversion mutations are generally more mutagenic than transition mutations (52), implying that *Ig* point mutations likely to cause the greatest changes in antibody binding characteristics are promoted by SAMHD1. For example, the generation of high-affinity anti-phenyloxazalone antibodies in BALB/c mice predominantly involves H34N and Y36F mutation of $V_\kappa Ox1/J_\kappa 5$ rearrangements: a C to A transversion and an A to T transversion, respectively (53). It's conceivable that G_1-phase MMR can also act as an antiviral mechanism, because repeated recruitment of ExoI to viral genomes that misincorporated G_1-phase dNTPs would inhibit viral replication. Given that G_1-phase nicking by MutLα in B cells does not discriminate between mutated and nonmutated DNA strands (5), it is further conceivable that G_1-phase MMR evolved to attack viral genomes and trigger apoptosis, rather than to repair errors, and was then coopted by *Ig* hypermutation, along with UNG2 and SAMHD1, as part of a preexisting suite of ancient intracellular antivirus mechanisms.

Materials and Methods

Mice. Male C57BL/6 host mice were initially purchased from Animal Resources Centre (Canning Vale, Western Australia) and later from Australian BioResources (Moss Vale, New South Wales), and were used in experiments when 8–16 wk old. *Msh2*^{ko}, *Rag1*^{ko}, *Igh*-knock-in/*Igk*-transgenic SW_{HEL}, and *Fucci*-red/*Fucci*-green mice (25, 31, 34, 54) were interbred and maintained on a C57BL/6 background under SPF conditions in the Centenary Institute Animal Facility.

Samhd1$^{\Delta/\Delta}$ mice were produced by the Mouse Engineering Garvan/ABR (MEGA) Facility (Moss Vale and Sydney, Australia) by CRISPR/Cas9 gene targeting in C57BL/6J mouse embryos following established molecular and animal husbandry techniques (55). To minimize off-target genome modifications, the double-nicking approach employing the single-strand cleaving mutant of the Cas9 endonuclease (Asp10Ala = Cas9n) was employed (56). For this purpose, paired single-guide RNAs (sgRNAs) were designed to target within the first coding exon (Exon 1) of *Samhd1*. The sites used were CTCCAAGGGTGCGCTCTGCATGG and GCTAAGCGACCCCGCTGCGATGG (protospacer-associated motifs = PAMs underlined).

To target *Samhd1*, a solution consisting of the two sgRNAs (15 ng/μL each) and full-length, polyadenylated Cas9n mRNA (30 ng/μL) was prepared and microinjected into the nucleus and cytoplasm of SW_{HEL} × C57BL/6 zygotes. Microinjected embryos were cultured overnight and introduced into pseudopregnant foster mothers. Pups were screened by PCR across the two target sites and Sanger sequencing to detect those with modifications to *Samhd1*. Independent founders carrying 25- and 41-bp frame-shift deletions within Exon 1 were selected, backcrossed to wild-type SW_{HEL} (C57BL/6J) mice four times, and the progeny was intercrossed to derive homozygous *Samhd1*$^{\Delta 25/\Delta 25}$ and *Samhd1*$^{\Delta 41/\Delta 41}$ lines.

All work was carried out with approval from and oversight by the Royal Prince Alfred Hospital Animal Welfare and Biosafety Committees, in accordance with NSW and Federal Australian legislation and the *Australian Code for the Care and Use of Animals for Scientific Purposes* (57).

Plasmids and Retroviruses. pMiG-based retroviruses expressing mKO2-cdt fusion protein (whose accumulation is restricted to G_0/G_1-phases by the degron from human Cdt1), EGFP, or ugi-EGFP fusion protein have been described (6). All cDNAs were cloned into the pMiG vector (58) using conventional techniques and verified by Sanger sequencing (Macrogen Inc). Ecotropic viral supernatants were produced by calcium phosphate transfection of the Platinum-E retroviral packaging cell line (Cell Biolabs, Inc.) as described (9). Culture medium (DMEM, 10% bovine calf serum, 1 mM sodium pyruvate, 2 mM L-alanyl-L-glutamine dipeptide, penicillin/streptomycin) was replaced at 12 h, then retroviral supernatants were harvested at 60 h, 0.45-μm filtered, snap frozen in liquid nitrogen, and stored at −70 °C.

Transduction and Adoptive Transfer of SW_{HEL} Splenocytes and Mutation Analysis. Activation of mouse splenocytes with recombinant CD40L for 24 h, followed by transduction in culture for a further 48 h has been described (9). GFP$^+$ cells were sorted using a BD FACS Aria II or BD influx sorter, mixed with sheep red blood cells (SRBC; Applied Biological Products Management) conjugated to hen egg lysozyme (HEL; Sigma-Aldrich) in B cell medium (54) and injected into host tail veins as a bolus of $\leq 10^4$ HEL-binding B cells plus 10^8 HEL-SRBC into 8- to 12-wk-old male hosts that had been primed i.p. 7 d prior with 10^8 SRBC in PBS, as described (34). Six days after adoptive transfer, host spleens were harvested, depleted of RBC using Histopaque 1083 fractionation (Sigma-Aldrich), and GFP$^+$ HEL-binding cells were sorted using a BD FACS Aria II or BD influx, one cell per well into 96-well PCR plates (4titude) containing 15 μL per well of ice-cold Mg-free 1× Taq DNA polymerase buffer (Promega "Go-Taq" buffer) supplemented to include 0.1 mM EDTA, 1% Tween-20 detergent (Astral Scientific) and 0.25 mg/mL proteinase K (Roche Diagnostics). Single-cell nested PCR of the singular gene-targeted SW_{HEL} VDJ_H rearrangement present in each cell was performed as described (25). Single allele mutation data were collected from up to 47 single cells using direct Sanger sequencing (Macrogen) of single-cell PCR products, as described (25), except that the sequencing primer (jol27: 5′-ACTC CACC AACA CCAT CACA C-3′) was positioned further 3′ to the start codon than previously, enlarging the sequence window from 523 bases to 560 bases. Each treatment involved at least three individual donors and hosts. Mutation data were analyzed by one-way ANOVA and Holm–Sidak's post hoc multiple comparisons tests with Prism 7 for MacOS (GraphPad Software, Inc), using the mean mutation of each hosts' sampled cells as a single data point (9).

Western Blot Analysis. B cells from *Fucci*-red/*Fucci*-green double-transgenic mice (31) were cultured in 20 μg/mL *S. typhosa* LPS (Sigma-Aldrich) plus 20 ng/mL mouse IL-4 (BD Biosciences) for 3 d to induce robust B cell proliferation. B cells from *Fucci* mice express an orange fluorescent fusion protein ("mKO2-cdt") in G_1 phase of the cell cycle or a green fluorescent fusion protein ("mAG-gmnn") in S, G_2, and M phases (31). Transduced cells were sorted into mKO2$^+$mAG$^-$ (G_1 phase) and mKO2$^-$mAG$^+$ ($S/G_2/M$ phase) fractions using a BD Influx sorter. Proteins were extracted from cells by vortexing cell pellets at 2×10^7 cells per mL in Laemmli sample buffer (0.1% 2-mercaptoethanol, 2% SDS, 10% glycerol, 0.0005% bromophenol blue, 63 mM Tris·Cl, pH 6.8), then incubating for 30 min at 20 °C with 250 units/mL Benzonase nuclease (Millipore). Extracts were electrophoresed using 4–12% Bolt Bis-Tris Plus polyacrylamide gels (Invitrogen) and transferred to nitrocellulose membranes using a Novex semi-dry blotter (Invitrogen) according to manufacturer protocols. Membranes were blocked for 1 h with 5% (wt/vol) skim milk powder in TBST [0.9% (wt/vol) NaCl, 0.1% (vol/vol) Tween-20 and 10 mM Tris·HCl, pH 7.4] at room temperature, then incubated overnight at 4 °C with mouse anti-SAMHD1 IgG2b monoclonal OTI1A1 (Abcam), goat anti-R2 IgG polyconal E-16 (Santa Cruz Biotechnology), or rabbit anti-R1 IgG monoclonal EPR8483 (Abcam) in skim milk/TBST. Washed membranes were incubated with species-appropriate secondary antibodies conjugated to horseradish peroxidase (Santa Cruz Biotechnology) for 1 h at 20 °C in skim milk/TBST, then washed again and incubated with SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher). Luminescence was recorded using a ChemiDoc MP (Bio-Rad).

dNTP Quantitation. *Samhd1*$^{\Delta 25/\Delta 25}$ or *Samhd1*$^{+/\Delta 25}$ splenocytes were cultured with LPS plus IL-4, for 24 h as above. Activated cells were then transduced with fresh retroviral supernatants of pMiG-derived retroviruses that expressed mKO2-cdt; cells expressing mKO2-cdt fluoresce red when they are in G_1 phase (6, 31). Two days later, orange-fluorescent (i.e., G_1-phase) cells were sorted using a BD Influx sorter into ice-cold bovine serum. Total dNTPs were extracted from PBS-washed sorted cells using ice-cold 60% aqueous methanol, as described (59). Resting or nonfractionated proliferating *Samhd1*$^{\Delta 25/\Delta 25}$ or *Samhd1*$^{+/\Delta 25}$ splenocytes depleted of RBC and dead cells using density gradient centrifugation over Histopaque 1083 (Sigma-Aldrich) were also extracted. dNTPs were then quantified using an HIV reverse transcriptase primer extension assay, as described (59).

ACKNOWLEDGMENTS. We thank Sydney Cytometry (Centenary Institute/The University of Sydney) for the provision of cell sorting facilities and expertise, and Australian BioResources and the Centenary Institute Animal Facility for the provision of mouse care and breeding. This research was supported by National Health and Medical Research Council Project Grants 1067891 and 1051820 (to C.J.J.), an Australian Postgraduate Award (to E.S.T.), NIH Grants AI049781 and GM104198 (to B.K.), and a Centenary Foundation grant (to C.J.J.).

1. Methot SP, Di Noia JM (2017) Molecular mechanisms of somatic hypermutation and class switch recombination. *Adv Immunol* 133:37–87.
2. 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* 16:163–171.
3. Shen HM, Tanaka A, Bozek G, Nicolae D, Storb U (2006) Somatic hypermutation and class switch recombination in Msh6(-/-)Ung(-/-) double-knockout mice. *J Immunol* 177: 5386–5392.
4. Dingler FA, Kemmerich K, Neuberger MS, Rada C (2014) Uracil excision by endogenous SMUG1 glycosylase promotes efficient Ig class switching and impacts on A:T substitutions during somatic mutation. *Eur J Immunol* 44:1925–1935.
5. Girelli Zubani G, et al. (2017) Pms2 and uracil-DNA glycosylases act jointly in the mismatch repair pathway to generate Ig gene mutations at A-T base pairs. *J Exp Med* 214:1169–1180.
6. Sharbeen G, Yee CW, Smith AL, Jolly CJ (2012) Ectopic restriction of DNA repair reveals that UNG2 excises AID-induced uracils predominantly or exclusively during G1 phase. *J Exp Med* 209:965–974.
7. Pérez-Durán P, et al. (2012) UNG shapes the specificity of AID-induced somatic hypermutation. *J Exp Med* 209:1379–1389.
8. Chen Z, et al. (2016) Interplay between target sequences and repair pathways determines distinct outcomes of AID-initiated lesions. *J Immunol* 196:2335–2347.
9. Thientosapol ES, et al. (2017) Proximity to AGCT sequences dictates MMR-independent versus MMR-dependent mechanisms for AID-induced mutation via UNG2. *Nucleic Acids Res* 45:3146–3157.
10. Esposito G, et al. (2000) Mice reconstituted with DNA polymerase beta-deficient fetal liver cells are able to mount a T cell-dependent immune response and mutate their Ig genes normally. *Proc Natl Acad Sci USA* 97:1166–1171.
11. Kano C, Hanaoka F, Wang JY (2012) Analysis of mice deficient in both REV1 catalytic activity and POLH reveals an unexpected role for POLH in the generation of C to G and G to C transversions during Ig gene hypermutation. *Int Immunol* 24:169–174.
12. Krijger PH, et al. (2013) Rev1 is essential in generating G to C transversions downstream of the Ung2 pathway but not the Msh2+Ung2 hybrid pathway. *Eur J Immunol* 43:2765–2770.
13. Weill J-C, Reynaud C-A (2008) DNA polymerases in adaptive immunity. *Nat Rev Immunol* 8:302–312.
14. Rada C, Ehrenstein MR, Neuberger MS, Milstein C (1998) Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity* 9:135–141.
15. Frey S, et al. (1998) Mismatch repair deficiency interferes with the accumulation of mutations in chronically stimulated B cells and not with the hypermutation process. *Immunity* 9:127–134.
16. Zeng X, et al. (2001) DNA polymerase eta is an A-T mutator in somatic hypermutation of immunoglobulin variable genes. *Nat Immunol* 2:537–541.
17. Bardwell PD, et al. (2004) Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice. *Nat Immunol* 5:224–229.
18. Delbos F, et al. (2005) Contribution of DNA polymerase eta to immunoglobulin gene hypermutation in the mouse. *J Exp Med* 201:1191–1196.
19. Langerak P, Nygren AO, Krijger PH, van den Berk PC, Jacobs H (2007) A/T mutagenesis in hypermutated immunoglobulin genes strongly depends on PCNAK164 modification. *J Exp Med* 204:1989–1998.
20. Peña-Diaz J, et al. (2012) Noncanonical mismatch repair as a source of genomic instability in human cells. *Mol Cell* 47:669–680.
21. Delbos F, Aoufouchi S, Faili A, Weill JC, Reynaud CA (2007) DNA polymerase eta is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse. *J Exp Med* 204:17–23.
22. Stavnezer J, et al. (2014) Differential expression of APE1 and APE2 in germinal centers promotes error-prone repair and A:T mutations during somatic hypermutation. *Proc Natl Acad Sci USA* 111:9217–9222.
23. Wang Q, et al. (2017) The cell cycle restricts activation-induced cytidine deaminase activity to early G1. *J Exp Med* 214:49–58.
24. Skoog L, Bjursell G (1974) Nuclear and cytoplasmic pools of deoxyribonucleoside triphosphates in Chinese hamster ovary cells. *J Biol Chem* 249:6434–6438.
25. Sharbeen G, et al. (2010) Incorporation of dUTP does not mediate mutation of A:T base pairs in Ig genes in vivo. *Nucleic Acids Res* 38:8120–8130.
26. Rampazzo C, et al. (2010) Regulation by degradation, a cellular defense against deoxyribonucleotide pool imbalances. *Mutat Res* 703:2–10.
27. Franzolin E, et al. (2013) The deoxynucleotide triphosphohydrolase SAMHD1 is a major regulator of DNA precursor pools in mammalian cells. *Proc Natl Acad Sci USA* 110:14272–14277.
28. Behrendt R, et al. (2013) Mouse SAMHD1 has antiretroviral activity and suppresses a spontaneous cell-intrinsic antiviral response. *Cell Rep* 4:689–696.
29. Ballana E, Este JA (2015) SAMHD1: At the crossroads of cell proliferation, immune responses, and virus restriction. *Trends Microbiol* 23:680–692.
30. Nordlund P, Reichard P (2006) Ribonucleotide reductases. *Annu Rev Biochem* 75: 681–706.
31. Aiba Y, et al. (2010) Preferential localization of IgG memory B cells adjacent to contracted germinal centers. *Proc Natl Acad Sci USA* 107:12192–12197.
32. Lahouassa H, et al. (2012) SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat Immunol* 13:223–228.
33. Rehwinkel J, et al. (2013) SAMHD1-dependent retroviral control and escape in mice. *EMBO J* 32:2454–2462.
34. Phan TG, et al. (2003) B cell receptor-independent stimuli trigger immunoglobulin (Ig) class switch recombination and production of IgG autoantibodies by anergic self-reactive B cells. *J Exp Med* 197:845–860.
35. Jansen JG, et al. (2006) Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. *J Exp Med* 203:319–323.
36. Jolly CJ, Klix N, Neuberger MS (1997) Rapid methods for the analysis of immunoglobulin gene hypermutation: Application to transgenic and gene targeted mice. *Nucleic Acids Res* 25:1913–1919.
37. Mol CD, et al. (1995) Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: Protein mimicry of DNA. *Cell* 82:701–708.
38. Krijger PH, Langerak P, van den Berk PC, Jacobs H (2009) Dependence of nucleotide substitutions on Ung2, Msh2, and PCNA-Ub during somatic hypermutation. *J Exp Med* 206:2603–2611.
39. Schanz S, Castor D, Fischer F, Jiricny J (2009) Interference of mismatch and base excision repair during the processing of adjacent U/G mispairs may play a key role in somatic hypermutation. *Proc Natl Acad Sci USA* 106:5593–5598.
40. 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: 5148–5157.
41. Ryoo J, et al. (2014) The ribonuclease activity of SAMHD1 is required for HIV-1 restriction. *Nat Med* 20:936–941.
42. Daddacha W, et al. (2017) SAMHD1 promotes DNA end resection to facilitate DNA repair by homologous recombination. *Cell Rep* 20:1921–1935.
43. Sale JE, Neuberger MS (1998) TdT-accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line. *Immunity* 9:859–869.
44. Cook AJ, et al. (2007) DNA-dependent protein kinase inhibits AID-induced antibody gene conversion. *PLoS Biol* 5:e80.
45. Kepler TB, et al. (2014) Immunoglobulin gene insertions and deletions in the affinity maturation of HIV-1 broadly reactive neutralizing antibodies. *Cell Host Microbe* 16: 304–313.
46. Woodrick J, et al. (2017) A new sub-pathway of long-patch base excision repair involving 5′ gap formation. *EMBO J* 36:1605–1622.
47. 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* 203:2085–2094.
48. Ayinde D, Casartelli N, Schwartz O (2012) Restricting HIV the SAMHD1 way: Through nucleotide starvation. *Nat Rev Microbiol* 10:675–680.
49. 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: 367–377.
50. Harris RS, Dudley JP (2015) APOBECs and virus restriction. *Virology* 479–480:131–145.
51. Krishnan A, Iyer LM, Holland SJ, Boehm T, Aravind L (2018) Diversification of AID/APOBEC-like deaminases in metazoa: Multiplicity of clades and widespread roles in immunity. *Proc Natl Acad Sci USA* 115:E3201–E3210.
52. Watson JD, et al. (2013) *Molecular Biology of the Gene* (Pearson, Boston), 7th Ed.
53. Griffiths GM, Berek C, Kaartinen M, Milstein C (1984) Somatic mutation and the maturation of immune response to 2-phenyl oxazolone. *Nature* 312:271–275.
54. Cook AJL, et al. (2003) Reduced switching in SCID B cells is associated with altered somatic mutation of recombined S regions. *J Immunol* 171:6556–6564.
55. Yang H, Wang H, Jaenisch R (2014) Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nat Protoc* 9:1956–1968.
56. Ran FA, et al. (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154:1380–1389.
57. National Health and Medical Research Council (2013) *Australian Code for the Care and Use of Animals for Scientific Purposes* (National Health and Medical Research Council, Canberra, Australia), 8th Ed.
58. Refaeli Y, Van Parijs L, Alexander SI, Abbas AK (2002) Interferon gamma is required for activation-induced death of T lymphocytes. *J Exp Med* 196:999–1005.
59. Diamond TL, et al. (2004) Macrophage tropism of HIV-1 depends on efficient cellular dNTP utilization by reverse transcriptase. *J Biol Chem* 279:51545–51553.
60. Sakaue-Sawano A, et al. (2008) Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* 132:487–498.
61. Zivojnovic M, et al. (2014) Somatic hypermutation at A/T-rich oligonucleotide substrates shows different strand polarities in Ung-deficient or -proficient backgrounds. *Mol Cell Biol* 34:2176–2187.

Supporting Information

Thientosapol et al. 10.1073/pnas.1719771115

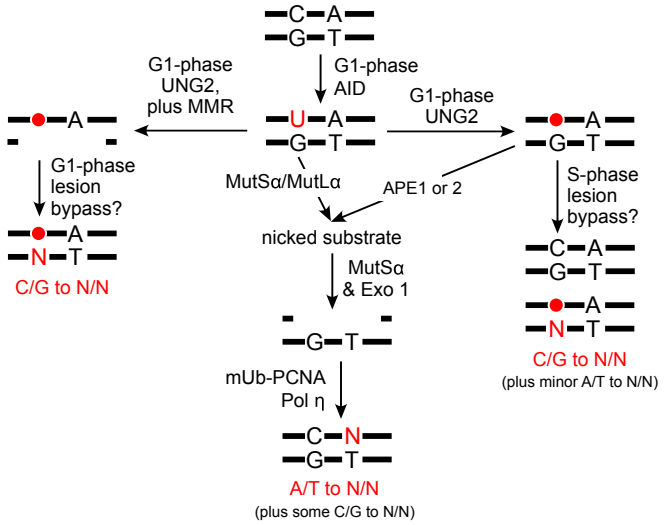


Fig. S1. Overview of the current *Ig* hypermutation model (based on refs. 1–3). •, unnicked apyrimidinic site.

1. Sharbeen G, Yee CW, Smith AL, Jolly CJ (2012) Ectopic restriction of DNA repair reveals that UNG2 excises AID-induced uracils predominantly or exclusively during G1 phase. *J Exp Med* 209:965–974.
2. Girelli Zubani G, et al. (2017) Pms2 and uracil-DNA glycosylases act jointly in the mismatch repair pathway to generate Ig gene mutations at A-T base pairs. *J Exp Med* 214:1169–1180.
3. Methot SP, Di Noia JM (2017) Molecular mechanisms of somatic hypermutation and class switch recombination. *Adv Immunol* 133:37–87.


```
Multiple sequence alignment

3 Sequences Aligned           Alignment Score = 0
Gaps Inserted = 2             Conserved Identities = 247

wt_SamHD1_exon_   1 ATGGACTCACTTTTGGGGTGTGGTGTCAGTGCGGCTGCGCGCGAGCCGGTGCCGCGGTAC  60
25bp_del          1 ATGGACTCACTTTTGGGGTGTGGTGTCAGTGCGGCTGCGCGCGAGCCGGTGCCGCGGTAC  60
41bp_del          1 ATGGACTCACTTTTGGGGTGTGGTGTCAGTGCGGCTGCGCGCGAGCCGGTGCCGCGGTAC  60
                    M  D  S  L  L  G  C  G  V  S  A  A  A  R  E  P  V  P  R  Y

wt_SamHD1_exon_  61 TTAACTTCGCAGCCCCGGGTGTCGGAGGTAGCCATGCAGAGCGCACCCTTGGAGCAGCCA 120
25bp_del         61 TTAACTTCGCAGCCCCGGGTGTCGGAGGTAGCCATGC----------------------  97
41bp_del         61 TTAACTTCGCAGCCCCGGGTGTCGGAGGTAGCCATGCAGAGCGCACCCAAGGA------- 113
                    L  T  S  Q  P  R  V  S  E  V  A  M  Q  S  A  P  L  E  Q  P

wt_SamHD1_exon_ 121 GCTAAGCGACCCCGCTGCGATGGCAGCCCAAGGACGCCACCGAGCACTCCTCCTGCAACA 180
25bp_del         98 --TAAGCGACCCCGCTGCGATGGCAGCCCAAGGACGCCACCGAGCACTCCTCCTGCAACA 155
41bp_del        114 ----------------------------------CGCCACCGAGCACTCCTCCTGCAACA 139
                     A  K  R  P  R  C  D  G  S  P  R  T  P  P  S  T  P  P  A  T

wt_SamHD1_exon_ 181 GCTAATCTGTCTGCAGACGACGACTTCCAAAACACCGACCTCCGAACCTGGGAACCGGAG 240
25bp_del        156 GCTAATCTGTCTGCAGACGACGACTTCCAAAACACCGACCTCCGAACCTGGGAACCGGAG 215
41bp_del        140 GCTAATCTGTCTGCAGACGACGACTTCCAAAACACCGACCTCCGAACCTGGGAACCGGAG 199
                     A  N  L  S  A  D  D  D  F  Q  N  T  D  L  R  T  W  E  P  E

wt_SamHD1_exon_ 241 GACGTGTGCTCCTTCTTAGAGAATCGTGGTTTCCGAGAGAAGAAAGTGCTGGACATCTTC 300
25bp_del        216 GACGTGTGCTCCTTCTTAGAGAATCGTGGTTTCCGAGAGAAGAAAGTGCTGGACATCTTC 275
41bp_del        200 GACGTGTGCTCCTTCTTAGAGAATCGTGGTTTCCGAGAGAAGAAAGTGCTGGACATCTTC 259
                     D  V  C  S  F  L  E  N  R  G  F  R  E  K  K  V  L  D  I  F


wt_SamHD1_exon_ 301 AGAG 304
25bp_del        276 AGAG 279
41bp_del        260 AGAG 263
                     R
```

Fig. S2. Alignment of the Exon 1 *Samhd1* sequences encoded by *Samhd1*$^{w.t.}$, *Samhd1*$^{\Delta 25}$, and *Samhd1*$^{\Delta 41}$ alleles, with the WT amino acid sequence indicated at the bottom. The guide RNA binding sites used for CRISPR mutation are bold underlined.

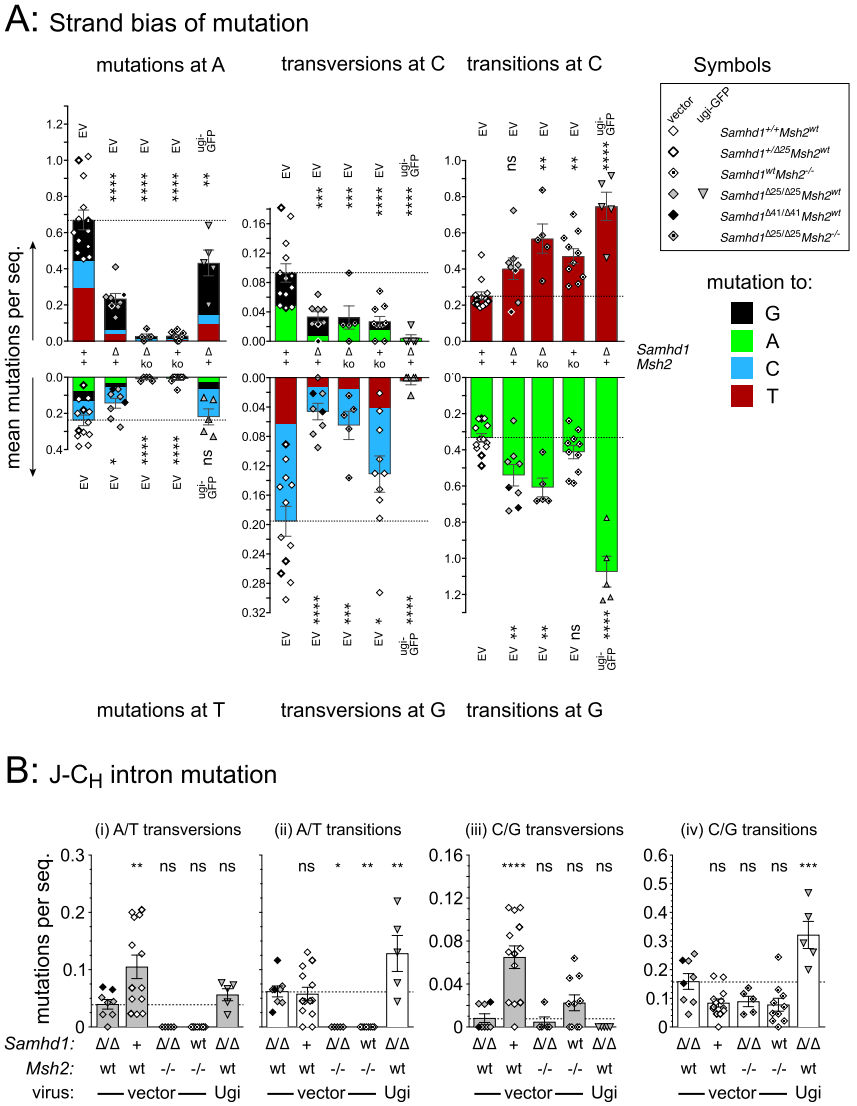


Fig. S3. (*A*) Mutation data from Fig. 2*B* and Table 1 broken down into mutations from A or T, transversions from C or G, and transitions from C or G, referring to the top strand. Each symbol is the mean number of mutations per sequence within one host. Histograms are the mean of hosts (±SEM), with each histogram segmented into mutations to G, A, C, or T (colors, without any variance indicated). (*B*) Hypermutation within the J-C_H intron, analyzed exactly as in Fig. 2*B*. Significant differences (one-way ANOVA, Holm–Sidak's post hoc multiple comparisons tests) to the vector-transduced wild-type means (*A*) or the vector-transduced *Samhd1*$^{\Delta/\Delta}$*Msh2*$^{+/+}$ means (*B*) are indicated: *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$; ns, not significant.

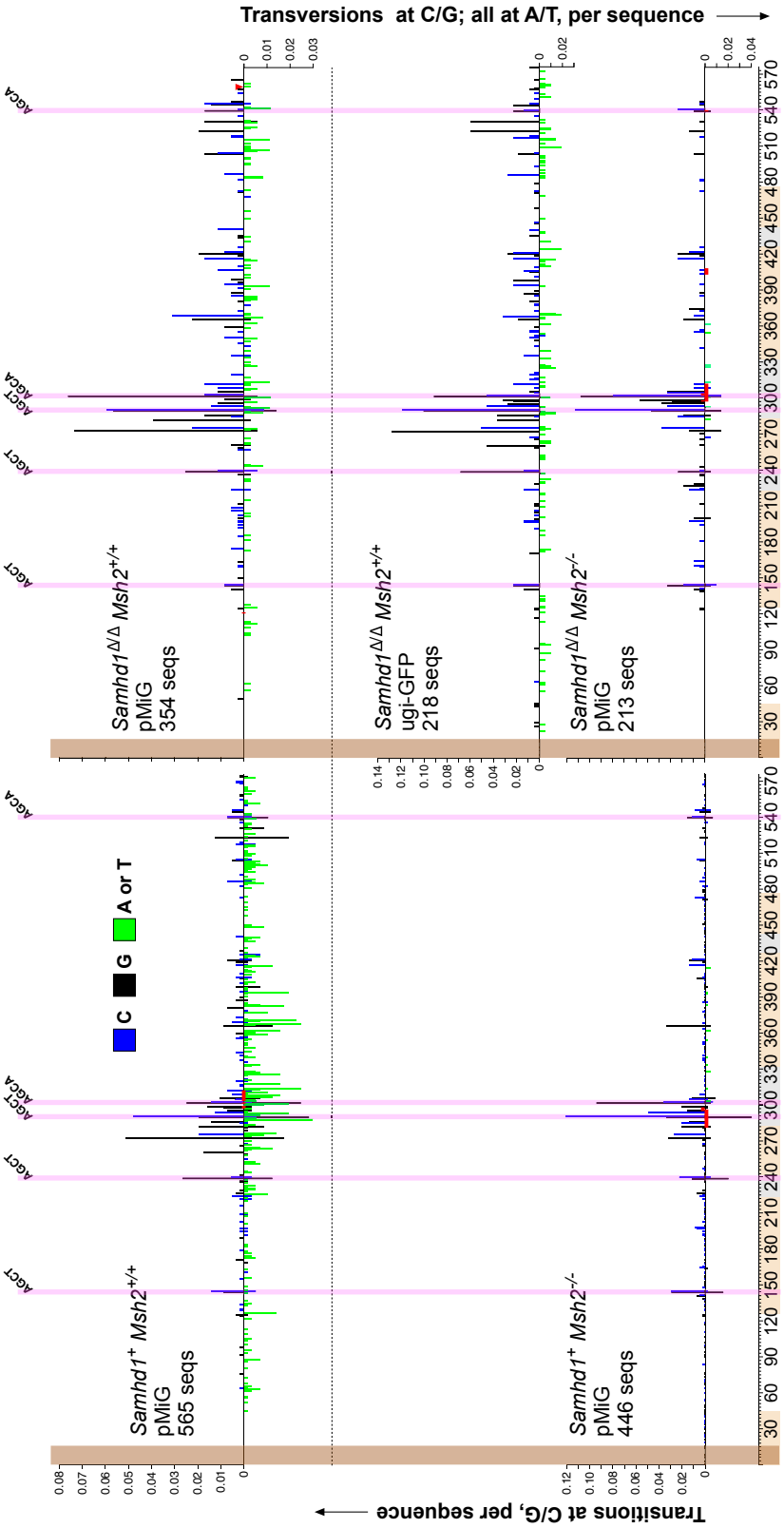


Fig. S4. Mutation frequencies (mutations per sequence) across the 560-bp window sequenced in the SW_{HEL} VDJ and 3′ region. Transitions at C (blue) or G (black) are shown above the *x* axis. Transversions at C (blue) or G (black) and mutations at A or T (green) are shown below the *x* axis. Red bars indicate deletions and the red triangle indicates a 1-bp insertion in the vector-transduced *Samhd1*$^{\Delta/\Delta}$ *Msh2*$^{+/+}$ dataset. The brown stripe indicates a 16-bp region in which mutations were not collated because of overlap with a secondary PCR primer. Pink stripes identify the five AGCW motifs present in the starting sequence (*Results* and *Discussion*). Beige boxes at the bottom indicate the L and VDJ exons, while gray boxes indicate the three CDRs, according to IMGT (the international ImMunoGeneTics information system www.imgt.org).

Table S1. Overview of mutation phenotypes discussed in this paper

	C/G mutation		A/T mutation		Hotspot-focusing	
Dataset	%C mutations = Tv	%G mutations = Tv	%mutations at A:T	%A/T mutations = Tv	% C/G mutations at WGCW* or AGCW†	Source
Peyer's patch VDJ_{H4}-C_H intron (495 bp)						1
Control	32	32	51	50	11*	
Msh2$^{-/-}$	18	24	10	64	48*	
Polh$^{-/-}$	41	40	16	81	26*	
Polh$^{-/-}$*Msh2*$^{-/-}$	13	17	1	–	57*	
Peyer's patch VDJ_{H4}-C_H intron (510 bp)						2
Control	37	45	55	54	19†,‡	
PCNA$^{K164R/K164R}$	26	42	5	73	16†,‡	
Ung$^{-/-}$	1	3	44	45	22†,‡	
Msh2$^{-/-}$	22	25	12	67	29†,‡	
PCNA$^{K164R/K164R}$ *Msh2*$^{-/-}$	11	27	2	67	32†,‡	
SW_{HEL} allele (560 bp)						This paper
Control	27	36	51	64	32†	
Msh2$^{-/-}$	5	24	4	59	48†	
Samhd1$^{\Delta/\Delta}$	8	8	27	31	31†	
Samhd1$^{\Delta/\Delta}$, *Msh2*$^{-/-}$	5	10	3	50	41†	
ugi+ *Samhd1*$^{\Delta/\Delta}$	1	0	26	33	11†	
SW_{HEL} VDJ-C_H intron (101 bp)						This paper
Control	32	52	52	65	19†	
Msh2$^{-/-}$	19	26	0	–	34†	
Samhd1$^{\Delta/\Delta}$	5	5	38	39	12†	
Samhd1$^{\Delta/\Delta}$, *Msh2*$^{-/-}$	0	9	0	–	40†	
ugi+ *Samhd1*$^{\Delta/\Delta}$	0	0	36	30	10†	

*Focusing to WGCW motifs.
†Focusing to AGCW motifs.
‡Focusing estimated using mutation skylines in figure 4 of ref. 2 and numbers in figure 2 of ref. 2.

1. Delbos F, Aoufouchi S, Faili A, Weill JC, Reynaud CA (2007) DNA polymerase eta is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse. *J Exp Med* 204:17–23.
2. Krijger PH, Langerak P, van den Berk PC, Jacobs H (2009) Dependence of nucleotide substitutions on Ung2, Msh2, and PCNA-Ub during somatic hypermutation. *J Exp Med* 206: 2603–2611.

6. General Discussion

In 2002, the laboratory of Michael Neuberger proposed the first comprehensive model of *Ig* somatic hypermutation (Di Noia & Neuberger, 2002; Petersen-Mahrt et al., 2002; Rada et al., 2002). This model incorporated the then recently discovered requirement for AID and the older discovery of a role for MMR in somatic hypermutation with new evidence that AID directly targeted DNA and that UNG2 diversified transition mutations at C/G to transversions. Supporting data was originally produced from *E. coli* (Petersen-Mahrt et al., 2002), then from the DT40 chicken lymphoblast cell line (Di Noia & Neuberger, 2002) and later from mouse B cells (Rada et al., 2002). The Neuberger model has since become the paradigm (Maul & Gearhart, 2014). While this model has been refined in the years since, this thesis attempts to answer some remaining important questions. Why do mutations hyper-focus on AID hotspots in MMR-deficient mouse B cells? Also, why are AID-induced deamination events not accurately repaired by DNA repair pathways, and are instead intensely mutagenic (Maul & Gearhart, 2014)?

Our laboratory developed the mouse SW_{HEL} transduction/adoptive transfer system as a quantitative model that can bypass embryo transgenesis (Sharbeen et al., 2010; Sharbeen et al., 2012), and potentially embryo gene-knockout, to investigate why DNA repair pathways are mutagenic during B cell somatic hypermutation. The transduction/adoptive transfer system was first used, in 2010 (Sharbeen et al., 2010), to test and refute Neuberger's "dUTP incorporation" model of AID-induced A:T mutation (Neuberger et al., 2005). Neubereger's hypothesis proposed that during G1 phase, when the dUTP/dTTP ratio rises as a result of depressed dUTPase enzyme

activity, dUTP becomes incorporated opposite A during re-synthesis of DNA in MMR generated excision patches. Incorporated uracils would then be excised by one of several redundant DNA glycosylases to generate AP sites. Polymerase η would then be recruited to bypass these AP sites in later rounds of AID-induced MMR, leading to permanent mutations opposite dA during in-fill of the MMR excision patches. In 2012, the SW_{HEL} model was used to produce definitive evidence that processing of AID-induced deaminations by UNG2 occurred predominantly in G1 and was both faithful and mutagenic in this cell cycle phase (Sharbeen et al., 2012). This thesis further develops the SW_{HEL} transduction/adoptive transfer model to gain new insights into somatic hypermutation, focusing on the role of cell cycle in turnover of all dNTPs (as opposed to dUTP solely), and testing whether RNA-silencing in adoptive B cells can be used as a substitute for time-consuming embryonic gene-knockout to definitively determine the role of novel factors in antibody hypermutation.

Using the largest number of biological replicates per dataset ever published, Chapter three quantifies the dependence of AID-induced hypermutation on UNG2 and MMR and mathematically models the focusing of mutations on AGCT/AGCA (AGCW) hotspots in *Msh2* knockout mouse B cells. Various mechanisms have been proposed to explain this phenomenon of hotspot-focusing. Scharff's group suggested that MMR was necessary to recruit AID to deaminate Cs outside of hotspot WRC motifs (Li et al., 2006). They based this conclusion on analysis of $Msh6^{+/+}$, $Msh6^{ko/ko}$ mice and mice with a *Msh6* knock-in ($Msh6^{TD/TD}$) that produced Msh6 protein, but ablated the ability of MMR to repair mismatches. They found that the $Msh6^{TD/TD}$ mice lacked A/T mutation, similar to the $Msh6^{ko/ko}$ mice. AID hotspot focusing was also present in $Msh6^{TD/TD}$ mice,

but displayed a different pattern of mutation at G:C compared to *Msh6*$^{ko/ko}$ and *Msh6*$^{+/+}$ mice. They concluded that these differences meant that Msh6 recruited AID to motifs outside of WRC. They did not perform experiments incorporating the effects of UNG2 deficiency and used small datasets, with low statistical power. Given the finding in Chapter 3 that MMR and UNG2 double-deficiency produced abundant mutation outside hotspots, rather than focusing mutation on hotspots, plus our use of very large datasets, it is unlikely that AID targeting *via* MSH6 is the reason for mutation focusing on hotspots in MMR deficiency. Instead, in agreement with Delbos *et al.* (Delbos et al., 2007), our data were consistent with the notion that mutation focuses in MMR-deficient B cells to motifs that are most resistant to faithful UNG2-mediated repair, and that it is this resistance to UNG2-mediated repair that explains hyper-focusing of mutation to hotspots in MMR-deficient cells.

Chapter three also refines the Neuberger model by showing that transversion mutations at C:G, already known to be largely dependent on UNG2, arise independently of MMR if they are in or near AGCT and AGCA hotspots, but are increasingly dependent (mathematically described by an exponential decay equation) on UNG2 plus MMR as distance from AGCT hotspots increases. This result was consistent with independent data generated by Chen *et al* (Chen et al., 2016), which we partially re-analysed in Chapter 3. Their data measured mutations in a conventional VDJ_H knock-in allele (analogous to our SW_{HEL} allele) and in an allele carrying a knocked-in AGCT-rich sequence. Chen *et al.*'s data revealed an 85% (conventional) and 89% (AGCT-rich) reduction in transversions at C:G with UNG2 deficiency and a 44% (conventional) and 29% (AGCT-rich) reduction in transversions

at C:G with *Msh2* knockout. These findings of a substantially lower reduction in MMR-dependent transversions in the AGCT-rich allele compared to the conventional allele, are consistent with our conclusions.

Chapters four and five investigate the role of dNTP paucity in G1 phase during somatic hypermutation. The original hypothesis that dNTP paucity (as opposed to dUTP/dTTP ratios) might contribute to somatic hypermutation was generated in our laboratory as a consequence of testing Neuberger's dUTP-incorporation hypothesis. Our laboratory's experiments showed that constitutive over-expression of mouse or EBV dUTPase in adoptive SW_{HEL} B cells reduced dUTP levels (because these enzymes convert dUTP to dUMP), but did not decrease mutations at A/T. Indeed, the opposite occurred: a small increase in mutations at A/T. It was proposed that excess dUTPase may have perturbed the dNTP pool generally (because dTTP is synthesized *via* dUMP), and this may have led to increased mis-incorporation of dNTPs during infill of MMR excision patches.

To test whether dNTP supply determines the frequencies of MMR-dependent *Ig* mutations, we have assessed the effect of constitutive R2 overexpression, and of *Samhd1* silencing and CRISPR-knockout in adoptive SW_{HEL} B cells. R2 overexpression and *Samhd1* CRISPR both caused nearly 50% reductions in A/T mutations, which was consistent with our dNTP paucity hypothesis. R2 overexpression and *Samhd1* CRISPR, however, also increased transition mutations at C/G by about 50%, and in *Samhd1* CRISPR-knockout cells there was a >70% decrease in C/G tranversion mutations. These additional findings were not initially predicted by our hypothesis.

The recurrent combination of decreased mutation at A/T and increased transition at C/G in our manipulated cells appeared to partially phenocopy MMR inhibition. However, using the hotspot-focused mutation signature described for MMR-deficiency in Chapter 3, we showed that *Samhd1*-inactivation and R2 overexpression did not induce hyper-focusing of mutation at AGCW motifs, unlike MMR deficiency. Therefore, it was unlikely that *Samhd1* inactivation or R2 overexpression inhibited MMR activity. *Polh*$^{-/-}$ (Delbos et al., 2007) and *Pcna*$^{K164R/K164R}$ (Krijger et al., 2009) mice also lose A/T mutations without concomitant hyper-focusing of mutation on AGCW hotspots (Chapter 5 Discussion). These similarities between *Samhd1* inactivation, R2 over-expression, *Polh*-inactivation and *Pcna*$^{K164R/K164R}$ mutation are consistent with our initial hypothesis: loss of R2 expression and increased *Samhd1* activity during G1 phase lead to nucleotide paucity-driven recruitment of polymerase η during AID-induced MMR, but do not prove it.

In *Samhd1*$^{\Delta/\Delta}$ B cells, increased C/G transition mutations combined with decreased C/G transversions can theoretically be explained by the greater increase in purine dNTPs relative to the pyrimidine dNTPs. To explicate: incorporation of purines opposite UNG2-induced AP sites by lesion bypass produces a transition mutation if dA is incorporated, or faithful bypass if dG is incorporated; incorporation of pyrimidines dC or dT produces a transversion mutation. Markedly increased bias towards transition mutation was also seen when analysing the A/T mutations remaining in *Samhd1*$^{\Delta/\Delta}$ B cells. Because mismatch excision downstream of AID appears to be heavily skewed to the top strand (Girelli Zubani et al., 2017), favoured incorporation

of dA and dG opposite dT can theoretically explain the preferential loss of transversion mutations at A:T base pairs (Chapter 5).

Although there was a similar and significant decrease in mutation at A/T induced by R2 overexpression and *Samhd1* inactivation, neither produced a complete loss in mutation at A/T. It is unlikely that in our models we achieved an exact S-phase level of dNTPs. This may explain why we saw an incomplete loss of A/T mutation, although it is also probable that there are multiple drivers of polymerase η recruitment in hypermutating B cells. In Chapter three we noted that C/G transversion mutations distal to AGCW motifs were dependent on both UNG2 and MMR. This implied that UNG2-derived AP sites are encountered by polymerases during infill of MMR excision patches. Encounters with AP sites in the intact strand may cause polymerases to stall during MMR infill, providing an additional driver for polymerase η recruitment.

In total, this thesis significantly refines the Neuberger model of somatic hypermutation, providing insights into how sequence context affects downstream processing by UNG2 and the BER pathway and also how G1-phase dNTP paucity could contribute to mutagenic DNA repair (see Figure 4 from Chapter 5). Combining these results with our laboratory's previous results, showing that transversions at C/G depend on G1 phase UNG2 activity, we now show that mutagenic processing of AID deaminations by uracil BER to create C/G transversions (phase 1B mutation) and by MMR to create A/T mutations (phase 2 mutations) are both at least partly driven by G1 phase dNTP paucity, which evolved prior to the appearance of antibodies as an ancient intra-cellular anti-virus mechanism.

Future Directions

Since writing this thesis, more recent journal articles have been published which raise different possibilities in which SAMHD1 may enhance immunoglobulin hypermutation.

Recently, SAMHD1 has been shown to interact with the DNA repair protein MRE11 at stalled replication forks (Coquel et al., 2018). While MRE11 is involved in DNA double-strand break repair (Paull, 2018), it has recently been found that initiation of somatic hypermutation and class switch recombination leads to the formation of a complex between many DNA repair proteins, including MRE11 (Kumar et al., 2018) and also proteins involved in somatic hypermutation, such as PCNA, MSH2, UNG, APE1 and polymerase η. SAMHD1 could possibly contribute to mutagenic DNA repair during somatic hypermutation by direct activity on the *Igh* locus or interactions with other DNA repair proteins, rather than its effect on G1 phase dNTP pools.

To further investigate these possibilities, we will measure the changes in somatic hypermutation in murine B cells with *Samhd1* mutants for either dNTPase activity or DNA repair protein binding ability.

7. References

Ahn, J. (2016). Functional organization of human SAMHD1 and mechanisms of HIV-1 restriction. *Biol Chem, 397*(4), 373-379. doi:10.1515/hsz-2015-0260

Ahn, J., Hao, C., Yan, J., DeLucia, M., Mehrens, J., Wang, C., . . . Skowronski, J. (2012). HIV/simian immunodeficiency virus (SIV) accessory virulence factor Vpx loads the host cell restriction factor SAMHD1 onto the E3 ubiquitin ligase complex CRL4DCAF1. *J Biol Chem, 287*(15), 12550-12558. doi:10.1074/jbc.M112.340711

Alvarez-Prado, A. F., Perez-Duran, P., Perez-Garcia, A., Benguria, A., Torroja, C., de Yebenes, V. G., & Ramiro, A. R. (2018). A broad atlas of somatic hypermutation allows prediction of activation-induced deaminase targets. *J Exp Med, 215*(3), 761-771. doi:10.1084/jem.20171738

Amie, S. M., Bambara, R. A., & Kim, B. (2013a). GTP is the primary activator of the anti-HIV restriction factor SAMHD1. *J Biol Chem, 288*(35), 25001-25006. doi:10.1074/jbc.C113.493619

Amie, S. M., Daly, M. B., Noble, E., Schinazi, R. F., Bambara, R. A., & Kim, B. (2013b). Anti-HIV host factor SAMHD1 regulates viral sensitivity to nucleoside reverse transcriptase inhibitors via modulation of cellular deoxyribonucleoside triphosphate (dNTP) levels. *J Biol Chem, 288*(28), 20683-20691. doi:10.1074/jbc.M113.472159

Andres, S. N., Vergnes, A., Ristic, D., Wyman, C., Modesti, M., & Junop, M. (2012). A human XRCC4-XLF complex bridges DNA. *Nucleic Acids Res, 40*(4), 1868-1878. doi:10.1093/nar/gks022

Andrulis, E. D., Werner, J., Nazarian, A., Erdjument-Bromage, H., Tempst, P., & Lis, J. T. (2002). The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila. *Nature, 420*(6917), 837-841. doi:10.1038/nature01181

Arnold, L. H., Groom, H. C., Kunzelmann, S., Schwefel, D., Caswell, S. J., Ordonez, P., . . . Bishop, K. N. (2015). Phospho-dependent Regulation of SAMHD1 Oligomerisation Couples Catalysis and Restriction. *PLoS Pathog, 11*(10), e1005194. doi:10.1371/journal.ppat.1005194

Ayinde, D., Casartelli, N., & Schwartz, O. (2012). Restricting HIV the SAMHD1 way: through nucleotide starvation. *Nat Rev Microbiol, 10*(10), 675-680. doi:10.1038/nrmicro2862

Bachl, J., Carlson, C., Gray-Schopfer, V., Dessing, M., & Olsson, C. (2001). Increased transcription levels induce higher mutation rates in a hypermutating cell line. *J Immunol, 166*(8), 5051-5057.

Baldauf, H. M., Pan, X., Erikson, E., Schmidt, S., Daddacha, W., Burggraf, M., . . . Keppler, O. T. (2012). SAMHD1 restricts HIV-1 infection in resting CD4(+) T cells. *Nature Medicine, 18*(11), 1682-1687. doi:10.1038/nm.2964

Ballana, E., & Este, J. A. (2015). SAMHD1: at the crossroads of cell proliferation, immune responses, and virus restriction. *Trends Microbiol, 23*(11), 680-692. doi:10.1016/j.tim.2015.08.002

Bardwell, P. D., Woo, C. J., Wei, K., Li, Z., Martin, A., Sack, S. Z., . . . Scharff, M. D. (2004). Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice. *Nat Immunol, 5*(2), 224-229. doi:10.1038/ni1031

Baroni, A., Buommino, E., De Gregorio, V., Ruocco, E., Ruocco, V., & Wolf, R. (2012). Structure and function of the epidermis related to barrier properties. *Clin Dermatol, 30*(3), 257-262. doi:10.1016/j.clindermatol.2011.08.007

Basu, U., Meng, F. L., Keim, C., Grinstein, V., Pefanis, E., Eccleston, J., . . . Alt, F. W. (2011). The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell, 144*(3), 353-363. doi:10.1016/j.cell.2011.01.001

Behrendt, R., Schumann, T., Gerbaulet, A., Nguyen, L. A., Schubert, N., Alexopoulou, D., . . . Roers, A. (2013). Mouse SAMHD1 has antiretroviral activity and suppresses a spontaneous cell-intrinsic antiviral response. *Cell Rep, 4*(4), 689-696. doi:10.1016/j.celrep.2013.07.037

Beloglazova, N., Flick, R., Tchigvintsev, A., Brown, G., Popovic, A., Nocek, B., & Yakunin, A. F. (2013). Nuclease activity of the human SAMHD1 protein implicated in the Aicardi-Goutieres syndrome and HIV-1 restriction. *J Biol Chem, 288*(12), 8101-8110. doi:10.1074/jbc.M112.431148

Berger, A., Sommer, A. F., Zwarg, J., Hamdorf, M., Welzel, K., Esly, N., . . . Flory, E. (2011). SAMHD1-deficient CD14+ cells from individuals with Aicardi-Goutieres syndrome are highly susceptible to HIV-1 infection. *PLoS Pathog, 7*(12), e1002425. doi:10.1371/journal.ppat.1002425

Bjorklund, S., Skog, S., Tribukait, B., & Thelander, L. (1990). S-phase-specific expression of mammalian ribonucleotide reductase R1 and R2 subunit mRNAs. *Biochemistry, 29*(23), 5452-5458.

Bonilla, F. A., & Oettgen, H. C. (2010). Adaptive immunity. *J Allergy Clin Immunol, 125*(2 Suppl 2), S33-40. doi:10.1016/j.jaci.2009.09.017

Brandariz-Nunez, A., Valle-Casuso, J. C., White, T. E., Laguette, N., Benkirane, M., Brojatsch, J., & Diaz-Griffero, F. (2012). Role of SAMHD1 nuclear localization in restriction of HIV-1 and SIVmac. *Retrovirology, 9*, 49. doi:10.1186/1742-4690-9-49

Bransteitter, R., Pham, P., Scharff, M. D., & Goodman, M. F. (2003). Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc Natl Acad Sci U S A, 100*(7), 4102-4107. doi:10.1073/pnas.0730835100

Burnet, F. (1957). A Modifcation of Jerne's Theory of Antibody Production using the Concept of Clonal Selection. *The Australian Journal of Science, 20*(3), 67-69.

Cambier, J. C., Gauld, S. B., Merrell, K. T., & Vilen, B. J. (2007). B-cell anergy: from transgenic models to naturally occurring anergic B cells? *Nat Rev Immunol, 7*(8), 633-643. doi:10.1038/nri2133

Cannon, W. B. (1929). Organization for physiological homeostasis. *Physiological reviews, 9*(3), 399-431.

Chabes, A. L., Bjorklund, S., & Thelander, L. (2004). S Phase-specific transcription of the mouse ribonucleotide reductase R2 gene requires both a proximal repressive E2F-binding site and an upstream promoter activating region. *J Biol Chem, 279*(11), 10796-10807. doi:10.1074/jbc.M312482200

Chang, H. H. Y., Pannunzio, N. R., Adachi, N., & Lieber, M. R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol, 18*(8), 495-506. doi:10.1038/nrm.2017.48

Chaudhuri, J., Tian, M., Khuong, C., Chua, K., Pinaud, E., & Alt, F. W. (2003). Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature, 422*(6933), 726-730. doi:10.1038/nature01574

Chen, Z., Eder, M. D., Elos, M. T., Viboolsittiseri, S. S., Chen, X., & Wang, J. H. (2016). Interplay between Target Sequences and Repair Pathways Determines Distinct Outcomes of AID-Initiated Lesions. *J Immunol, 196*(5), 2335-2347. doi:10.4049/jimmunol.1502184

Chen, Z., Zhu, M., Pan, X., Zhu, Y., Yan, H., Jiang, T., . . . Shen, Y. (2014). Inhibition of Hepatitis B virus replication by SAMHD1. *Biochem Biophys Res Commun, 450*(4), 1462-1468. doi:10.1016/j.bbrc.2014.07.023

Choi, J. H., Wang, K. W., Zhang, D., Zhan, X., Wang, T., Bu, C. H., . . . Beutler, B. (2017). IgD class switching is initiated by microbiota and limited to mucosa-associated lymphoid tissue in mice. *Proc Natl Acad Sci U S A, 114*(7), E1196-E1204. doi:10.1073/pnas.1621258114

Clifford, R., Louis, T., Robbe, P., Ackroyd, S., Burns, A., Timbs, A. T., . . . Schuh, A. (2014). SAMHD1 is mutated recurrently in chronic lymphocytic leukemia and is involved in response to DNA damage. *Blood, 123*(7), 1021-1031. doi:10.1182/blood-2013-04-490847

Cook, A. J., Oganesian, L., Harumal, P., Basten, A., Brink, R., & Jolly, C. J. (2003). Reduced switching in SCID B cells is associated with altered somatic mutation of recombined S regions. *J Immunol, 171*(12), 6556-6564.

Cooper, S. J. (2008). From Claude Bernard to Walter Cannon. Emergence of the concept of homeostasis. *Appetite, 51*(3), 419-427. doi:10.1016/j.appet.2008.06.005

Coquel, F., Silva, M. J., Techer, H., Zadorozhny, K., Sharma, S., Nieminuszczy, J., . . . Pasero, P. (2018). SAMHD1 acts at stalled replication forks to prevent interferon induction. *Nature, 557*(7703), 57-61. doi:10.1038/s41586-018-0050-1

Cribier, A., Descours, B., Valadao, A. L., Laguette, N., & Benkirane, M. (2013). Phosphorylation of SAMHD1 by cyclin A2/CDK1 regulates its restriction activity toward HIV-1. *Cell Rep, 3*(4), 1036-1043. doi:10.1016/j.celrep.2013.03.017

Crow, Y. J. (1993). Aicardi-Goutieres Syndrome. In R. A. Pagon, M. P. Adam, H. H. Ardinger, S. E. Wallace, A. Amemiya, L. J. H. Bean, T. D. Bird, C. T. Fong, H. C. Mefford, R. J. H. Smith, & K. Stephens (Eds.), *GeneReviews(R)*. Seattle (WA).

Crow, Y. J., & Manel, N. (2015). Aicardi-Goutieres syndrome and the type I interferonopathies. *Nat Rev Immunol, 15*(7), 429-440. doi:10.1038/nri3850

Crow, Y. J., & Rehwinkel, J. (2009). Aicardi-Goutieres syndrome and related phenotypes: linking nucleic acid metabolism with autoimmunity. *Hum Mol Genet, 18*(R2), R130-136. doi:10.1093/hmg/ddp293

Daddacha, W., Koyen, A. E., Bastien, A. J., Head, P. E., Dhere, V. R., Nabeta, G. N., . . . Yu, D. S. (2017). SAMHD1 Promotes DNA End Resection to Facilitate DNA Repair by Homologous Recombination. *Cell Rep, 20*(8), 1921-1935. doi:10.1016/j.celrep.2017.08.008

Davis, & Chen, D. J. (2013). DNA double strand break repair via non-homologous end-joining. *Transl Cancer Res, 2*(3), 130-143. doi:10.3978/j.issn.2218-676X.2013.04.02

Davis, Havener, J., & Ramsden, D. (2008). End-bridging is required for pol mu to efficiently promote repair of noncomplementary ends by nonhomologous end joining. *Nucleic Acids Res, 36*(9), 3085-3094. doi:10.1093/nar/gkn164

de Bono, B., Madera, M., & Chothia, C. (2004). VH gene segments in the mouse and human genomes. *J Mol Biol, 342*(1), 131-143. doi:10.1016/j.jmb.2004.06.055

de Wind, N., Dekker, M., Berns, A., Radman, M., & te Riele, H. (1995). Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. *Cell, 82*(2), 321-330.

Dedeoglu, F., Horwitz, B., Chaudhuri, J., Alt, F. W., & Geha, R. S. (2004). Induction of activation-induced cytidine deaminase gene expression by IL-4 and CD40 ligation is dependent on STAT6 and NFkappaB. *Int Immunol, 16*(3), 395-404.

Delbos, F., Aoufouchi, S., Faili, A., Weill, J. C., & Reynaud, C. A. (2007). DNA polymerase eta is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse. *J Exp Med, 204*(1), 17-23. doi:10.1084/jem.20062131

Delbos, F., De Smet, A., Faili, A., Aoufouchi, S., Weill, J. C., & Reynaud, C. A. (2005). Contribution of DNA polymerase eta to immunoglobulin gene hypermutation in the mouse. *J Exp Med, 201*(8), 1191-1196. doi:10.1084/jem.20050292

Di Noia, J. M., & Neuberger, M. S. (2002). Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature, 419*(6902), 43-48. doi:10.1038/nature00981

Di Noia, J. M., Rada, C., & Neuberger, M. S. (2006). SMUG1 is able to excise uracil from immunoglobulin genes: insight into mutation versus repair. *EMBO J, 25*(3), 585-595. doi:10.1038/sj.emboj.7600939

Dickerson, S. K., Market, E., Besmer, E., & Papavasiliou, F. N. (2003). AID mediates hypermutation by deaminating single stranded DNA. *J Exp Med, 197*(10), 1291-1296. doi:10.1084/jem.20030481

Dingler, F. A., Kemmerich, K., Neuberger, M. S., & Rada, C. (2014). Uracil excision by endogenous SMUG1 glycosylase promotes efficient Ig class switching and impacts on A:T substitutions during somatic mutation. *Eur J Immunol, 44*(7), 1925-1935. doi:10.1002/eji.201444482

Dow, L. E., Premsrirut, P. K., Zuber, J., Fellmann, C., McJunkin, K., Miething, C., . . . Lowe, S. W. (2012). A pipeline for the generation of shRNA transgenic mice. *Nat Protoc, 7*(2), 374-393. doi:10.1038/nprot.2011.446

Dudley, D. D., Chaudhuri, J., Bassing, C. H., & Alt, F. W. (2005). Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences. *Adv Immunol, 86*, 43-112. doi:10.1016/S0065-2776(04)86002-4

Dunkelberger, J. R., & Song, W. C. (2010). Complement and its role in innate and adaptive immune responses. *Cell Res, 20*(1), 34-50. doi:10.1038/cr.2009.139

Dussaix, E., Lebon, P., Ponsot, G., Huault, G., & Tardieu, M. (1985). Intrathecal synthesis of different alpha-interferons in patients with various neurological diseases. *Acta Neurol Scand, 71*(6), 504-509.

Elias, P. M. (2007). The skin barrier as an innate immune element. *Semin Immunopathol, 29*(1), 3-14.

Engstrom, Y., Eriksson, S., Jildevik, I., Skog, S., Thelander, L., & Tribukait, B. (1985). Cell cycle-dependent expression of mammalian ribonucleotide reductase. Differential regulation of the two subunits. *J Biol Chem, 260*(16), 9114-9116.

Epelman, S., Lavine, K. J., & Randolph, G. J. (2014). Origin and functions of tissue macrophages. *Immunity, 41*(1), 21-35. doi:10.1016/j.immuni.2014.06.013

Faili, A., Stary, A., Delbos, F., Weller, S., Aoufouchi, S., Sarasin, A., . . . Reynaud, C. A. (2009). A backup role of DNA polymerase kappa in Ig gene hypermutation only takes place in the complete absence of DNA polymerase eta. *J Immunol, 182*(10), 6353-6359. doi:10.4049/jimmunol.0900177

Franzolin, E., Pontarin, G., Rampazzo, C., Miazzi, C., Ferraro, P., Palumbo, E., . . . Bianchi, V. (2013). The deoxynucleotide triphosphohydrolase SAMHD1 is a major regulator of DNA precursor pools in mammalian cells. *Proc Natl Acad Sci U S A, 110*(35), 14272-14277. doi:10.1073/pnas.1312033110

Frey, S., Bertocci, B., Delbos, F., Quint, L., Weill, J. C., & Reynaud, C. A. (1998). Mismatch repair deficiency interferes with the accumulation of mutations in chronically stimulated B cells and not with the hypermutation process. *Immunity, 9*(1), 127-134.

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. doi:10.1128/MCB.00647-09

Fugmann, S. D., Lee, A. I., Shockett, P. E., Villey, I. J., & Schatz, D. G. (2000). The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu Rev Immunol, 18*, 495-527. doi:10.1146/annurev.immunol.18.1.495

Ganz, T. (2003). Defensins: antimicrobial peptides of innate immunity. *Nat Rev Immunol, 3*(9), 710-720. doi:10.1038/nri1180

Girelli Zubani, G., Zivojnovic, M., De Smet, A., Albagli-Curiel, O., Huetz, F., Weill, J. C., . . . Storck, S. (2017). Pms2 and uracil-DNA glycosylases act jointly in the mismatch repair pathway to generate Ig gene mutations at A-T base pairs. *J Exp Med, 214*(4), 1169-1180. doi:10.1084/jem.20161576

Gitlin, A. D., Mayer, C. T., Oliveira, T. Y., Shulman, Z., Jones, M. J., Koren, A., & Nussenzweig, M. C. (2015). HUMORAL IMMUNITY. T cell help controls the speed of the cell cycle in germinal center B cells. *Science, 349*(6248), 643-646. doi:10.1126/science.aac4919

Gitlin, A. D., Shulman, Z., & Nussenzweig, M. C. (2014). Clonal selection in the germinal centre by regulated proliferation and hypermutation. *Nature, 509*(7502), 637-640. doi:10.1038/nature13300

Glaser, R., Harder, J., Lange, H., Bartels, J., Christophers, E., & Schroder, J. M. (2005). Antimicrobial psoriasin (S100A7) protects human skin from Escherichia coli infection. *Nat Immunol, 6*(1), 57-64. doi:10.1038/ni1142

Goldstone, D. C., Ennis-Adeniran, V., Hedden, J. J., Groom, H. C., Rice, G. I., Christodoulou, E., . . . Webb, M. (2011). HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature, 480*(7377), 379-382. doi:10.1038/nature10623

Goncalves, A., Karayel, E., Rice, G. I., Bennett, K. L., Crow, Y. J., Superti-Furga, G., & Burckstummer, T. (2012). SAMHD1 is a nucleic-acid binding protein that is mislocalized due to aicardi-goutieres syndrome-associated mutations. *Hum Mutat, 33*(7), 1116-1122. doi:10.1002/humu.22087

Goujon, C., Riviere, L., Jarrosson-Wuilleme, L., Bernaud, J., Rigal, D., Darlix, J. L., & Cimarelli, A. (2007). SIVSM/HIV-2 Vpx proteins promote retroviral escape from a proteasome-dependent restriction pathway present in human dendritic cells. *Retrovirology, 4*, 2. doi:10.1186/1742-4690-4-2

Gramberg, T., Kahle, T., Bloch, N., Wittmann, S., Mullers, E., Daddacha, W., . . . Landau, N. R. (2013). Restriction of diverse retroviruses by SAMHD1. *Retrovirology, 10*, 26. doi:10.1186/1742-4690-10-26

Guyader, M., Emerman, M., Montagnier, L., & Peden, K. (1989). VPX mutants of HIV-2 are infectious in established cell lines but display a severe defect in peripheral blood lymphocytes. *EMBO J, 8*(4), 1169-1175.

Hamel, K. M., Liarski, V. M., & Clark, M. R. (2012). Germinal center B-cells. *Autoimmunity, 45*(5), 333-347. doi:10.3109/08916934.2012.665524

Han, L., Masani, S., & Yu, K. (2011). Overlapping activation-induced cytidine deaminase hotspot motifs in Ig class-switch recombination. *Proc Natl Acad Sci U S A, 108*(28), 11584-11589. doi:10.1073/pnas.1018726108

Hansen, E. C., Seamon, K. J., Cravens, S. L., & Stivers, J. T. (2014). GTP activator and dNTP substrates of HIV-1 restriction factor SAMHD1 generate a long-lived activated state. *Proc Natl Acad Sci U S A, 111*(18), E1843-1851. doi:10.1073/pnas.1401706111

Harder, J., & Schroder, J. M. (2002). RNase 7, a novel innate immune defense antimicrobial protein of healthy human skin. *J Biol Chem, 277*(48), 46779-46784. doi:10.1074/jbc.M207587200

Healy, J. I., Dolmetsch, R. E., Timmerman, L. A., Cyster, J. G., Thomas, M. L., Crabtree, G. R., . . . Goodnow, C. C. (1997). Different nuclear signals are activated by the B cell receptor during positive versus negative signaling. *Immunity, 6*(4), 419-428.

Heesters, B. A., Myers, R. C., & Carroll, M. C. (2014). Follicular dendritic cells: dynamic antigen libraries. *Nat Rev Immunol, 14*(7), 495-504. doi:10.1038/nri3689

Hoege, C., Pfander, B., Moldovan, G. L., Pyrowolakis, G., & Jentsch, S. (2002). RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature, 419*(6903), 135-141. doi:10.1038/nature00991

Hollenbaugh, J. A., Gee, P., Baker, J., Daly, M. B., Amie, S. M., Tate, J., . . . Kim, B. (2013). Host factor SAMHD1 restricts DNA viruses in non-dividing myeloid cells. *PLoS Pathog, 9*(6), e1003481. doi:10.1371/journal.ppat.1003481

Hombauer, H., Srivatsan, A., Putnam, C. D., & Kolodner, R. D. (2011). Mismatch repair, but not heteroduplex rejection, is temporally coupled to DNA replication. *Science, 334*(6063), 1713-1716. doi:10.1126/science.1210770

Honjo, T. (1983). Immunoglobulin genes. *Annu Rev Immunol, 1*, 499-528. doi:10.1146/annurev.iy.01.040183.002435

Hrecka, K., Hao, C., Gierszewska, M., Swanson, S. K., Kesik-Brodacka, M., Srivastava, S., . . . Skowronski, J. (2011). Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature, 474*(7353), 658-661. doi:10.1038/nature10195

Iwasaki, A., & Medzhitov, R. (2015). Control of adaptive immunity by the innate immune system. *Nat Immunol, 16*(4), 343-353. doi:10.1038/ni.3123

Jacobs, Fukita, Y., van der Horst, G. T., de Boer, J., Weeda, G., Essers, J., . . . Rajewsky, K. (1998). Hypermutation of immunoglobulin genes in memory B cells of DNA repair-deficient mice. *J Exp Med, 187*(11), 1735-1743.

Jacobs, & Schar, P. (2012). DNA glycosylases: in DNA repair and beyond. *Chromosoma, 121*(1), 1-20. doi:10.1007/s00412-011-0347-4

Janeway, C., Travers, P., & Walport, M. (2001). Immunobiology: The immune system in health and disease.

Jansen, J. G., Langerak, P., Tsaalbi-Shtylik, A., van den Berk, P., Jacobs, H., & de Wind, N. (2006). Strand-biased defect in C/G transversions in hypermutating

immunoglobulin genes in Rev1-deficient mice. *J Exp Med, 203*(2), 319-323. doi:10.1084/jem.20052227

Ji, X., Wu, Y., Yan, J., Mehrens, J., Yang, H., DeLucia, M., . . . Xiong, Y. (2013). Mechanism of allosteric activation of SAMHD1 by dGTP. *Nat Struct Mol Biol, 20*(11), 1304-1309. doi:10.1038/nsmb.2692

Jiricny, J. (2013). Postreplicative mismatch repair. *Cold Spring Harb Perspect Biol, 5*(4), a012633. doi:10.1101/cshperspect.a012633

Jolles, P., & Jolles, J. (1984). What's new in lysozyme research? Always a model system, today as yesterday. *Mol Cell Biochem, 63*(2), 165-189.

Jolly, C. J., Wagner, S. D., Rada, C., Klix, N., Milstein, C., & Neuberger, M. S. (1996). The targeting of somatic hypermutation. *Semin Immunol, 8*(3), 159-168. doi:10.1006/smim.1996.0020

Jung, D., & Alt, F. W. (2004). Unraveling V(D)J recombination; insights into gene regulation. *Cell, 116*(2), 299-311.

Jung, D., Giallourakis, C., Mostoslavsky, R., & Alt, F. W. (2006). Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu Rev Immunol, 24*, 541-570. doi:10.1146/annurev.immunol.23.021704.115830

Kannouche, P. L., Wing, J., & Lehmann, A. R. (2004). Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage. *Mol Cell, 14*(4), 491-500.

Kano, C., Hanaoka, F., & Wang, J. Y. (2012). Analysis of mice deficient in both REV1 catalytic activity and POLH reveals an unexpected role for POLH in the generation of C to G and G to C transversions during Ig gene hypermutation. *Int Immunol, 24*(3), 169-174. doi:10.1093/intimm/dxr109

Kilchert, C., Wittmann, S., & Vasiljeva, L. (2016). The regulation and functions of the nuclear RNA exosome complex. *Nat Rev Mol Cell Biol, 17*(4), 227-239. doi:10.1038/nrm.2015.15

Kim, E. T., White, T. E., Brandariz-Nunez, A., Diaz-Griffero, F., & Weitzman, M. D. (2013). SAMHD1 restricts herpes simplex virus 1 in macrophages by limiting DNA replication. *J Virol, 87*(23), 12949-12956. doi:10.1128/JVI.02291-13

Koharudin, L. M., Wu, Y., DeLucia, M., Mehrens, J., Gronenborn, A. M., & Ahn, J. (2014). Structural basis of allosteric activation of sterile alpha motif and histidine-aspartate domain-containing protein 1 (SAMHD1) by nucleoside triphosphates. *J Biol Chem, 289*(47), 32617-32627. doi:10.1074/jbc.M114.591958

Kohnken, R., Kodigepalli, K. M., & Wu, L. (2015). Regulation of deoxynucleotide metabolism in cancer: novel mechanisms and therapeutic implications. *Mol Cancer, 14*, 176. doi:10.1186/s12943-015-0446-6

Kosciuczuk, E. M., Lisowski, P., Jarczak, J., Strzalkowska, N., Jozwik, A., Horbanczuk, J., . . . Bagnicka, E. (2012). Cathelicidins: family of antimicrobial peptides. A review. *Mol Biol Rep, 39*(12), 10957-10970. doi:10.1007/s11033-012-1997-x

Krautler, N. J., Suan, D., Butt, D., Bourne, K., Hermes, J. R., Chan, T. D., . . . Brink, R. (2017). Differentiation of germinal center B cells into plasma cells is initiated by high-affinity antigen and completed by Tfh cells. *J Exp Med, 214*(5), 1259-1267. doi:10.1084/jem.20161533

Kretschmer, S., Wolf, C., Konig, N., Staroske, W., Guck, J., Hausler, M., . . . Lee-Kirsch, M. A. (2015). SAMHD1 prevents autoimmunity by maintaining genome stability. *Ann Rheum Dis, 74*(3), e17. doi:10.1136/annrheumdis-2013-204845

Krijger, P. H., Langerak, P., van den Berk, P. C., & Jacobs, H. (2009). Dependence of nucleotide substitutions on Ung2, Msh2, and PCNA-Ub during somatic hypermutation. *J Exp Med, 206*(12), 2603-2611. doi:10.1084/jem.20091707

Krijger, P. H., Tsaalbi-Shtylik, A., Wit, N., van den Berk, P. C., de Wind, N., & Jacobs, H. (2013). Rev1 is essential in generating G to C transversions downstream of the Ung2 pathway but not the Msh2+Ung2 hybrid pathway. *Eur J Immunol, 43*(10), 2765-2770. doi:10.1002/eji.201243191

Krokan, H. E., & Bjoras, M. (2013). Base excision repair. *Cold Spring Harb Perspect Biol, 5*(4), a012583. doi:10.1101/cshperspect.a012583

Kumar, A., Priya, A., Ahmed, T., Grundstrom, C., Negi, N., & Grundstrom, T. (2018). Correction: Regulation of the DNA Repair Complex during Somatic Hypermutation and Class-Switch Recombination. *J Immunol, 201*(11), 3472-3473. doi:10.4049/jimmunol.1801371

Kunos, C. A., Ferris, G., Pyatka, N., Pink, J., & Radivoyevitch, T. (2011). Deoxynucleoside salvage facilitates DNA repair during ribonucleotide reductase blockade in human cervical cancers. *Radiat Res, 176*(4), 425-433.

Laffleur, B., Basu, U., & Lim, J. (2017). RNA Exosome and Non-coding RNA-Coupled Mechanisms in AID-Mediated Genomic Alterations. *J Mol Biol, 429*(21), 3230-3241. doi:10.1016/j.jmb.2016.12.021

Laguette, N., Sobhian, B., Casartelli, N., Ringeard, M., Chable-Bessia, C., Segeral, E., . . . Benkirane, M. (2011). SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature, 474*(7353), 654-657. doi:10.1038/nature10117

Lahouassa, H., Daddacha, W., Hofmann, H., Ayinde, D., Logue, E. C., Dragin, L., . . . Margottin-Goguet, F. (2012). SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat Immunol, 13*(3), 223-228. doi:10.1038/ni.2236

Langerak, P., Nygren, A. O., Schouten, J. P., & Jacobs, H. (2005). Rapid and quantitative detection of homologous and non-homologous recombination events using three oligonucleotide MLPA. *Nucleic Acids Res, 33*(22), e188. doi:10.1093/nar/gni187

Le, Q., & Maizels, N. (2015). Cell Cycle Regulates Nuclear Stability of AID and Determines the Cellular Response to AID. *PLoS Genet, 11*(9). doi:ARTN e1005411
10.1371/journal.pgen.1005411

LeBien, T. W., & Tedder, T. F. (2008). B lymphocytes: how they develop and function. *Blood, 112*(5), 1570-1580. doi:10.1182/blood-2008-02-078071

Leshinsky-Silver, E., Malinger, G., Ben-Sira, L., Kidron, D., Cohen, S., Inbar, S., . . . Lerman-Sagie, T. (2011). A large homozygous deletion in the SAMHD1 gene causes atypical Aicardi-Goutieres syndrome associated with mtDNA deletions. *Eur J Hum Genet, 19*(3), 287-292. doi:10.1038/ejhg.2010.213

Li, Zhang, W., & Cao, X. (2000). Identification of human homologue of mouse IFN-gamma induced protein from human dendritic cells. *Immunol Lett, 74*(3), 221-224.

Li, Zhao, C., Iglesias-Ussel, M. D., Polonskaya, Z., Zhuang, M., Yang, G., . . . Scharff, M. D. (2006). The mismatch repair protein Msh6 influences the in vivo AID targeting to the Ig locus. *Immunity, 24*(4), 393-403. doi:10.1016/j.immuni.2006.02.011

Li, Z., Woo, C., Iglesias-Ussel, M., Ronai, D., & Scharff, M. (2004). The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes Dev, 18*(1), 1-11. doi:10.1101/gad.1161904

Lim, Y. W., Sanz, L. A., Xu, X., Hartono, S. R., & Chedin, F. (2015). Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi-Goutieres syndrome. *Elife, 4*. doi:10.7554/eLife.08007

Liu, M., Duke, J. L., Richter, D. J., Vinuesa, C. G., Goodnow, C. C., Kleinstein, S. H., & Schatz, D. G. (2008). Two levels of protection for the B cell genome during somatic hypermutation. *Nature, 451*(7180), 841-845. doi:10.1038/nature06547

Loureiro, L. R., Carrascal, M. A., Barbas, A., Ramalho, J. S., Novo, C., Delannoy, P., & Videira, P. A. (2015). Challenges in Antibody Development against Tn and Sialyl-Tn Antigens. *Biomolecules, 5*(3), 1783-1809. doi:10.3390/biom5031783

Malu, S., Malshetty, V., Francis, D., & Cortes, P. (2012). Role of non-homologous end joining in V(D)J recombination. *Immunol Res, 54*(1-3), 233-246. doi:10.1007/s12026-012-8329-z

Mann, G. J., Musgrove, E. A., Fox, R. M., & Thelander, L. (1988). Ribonucleotide reductase M1 subunit in cellular proliferation, quiescence, and differentiation. *Cancer Res, 48*(18), 5151-5156.

Masuda, K., Ouchida, R., Hikida, M., Kurosaki, T., Yokoi, M., Masutani, C., . . . J, O. W. (2007). DNA polymerases eta and theta function in the same genetic pathway to generate mutations at A/T during somatic hypermutation of Ig genes. *J Biol Chem, 282*(24), 17387-17394. doi:10.1074/jbc.M611849200

Masuda, K., Ouchida, R., Hikida, M., Nakayama, M., Ohara, O., Kurosaki, T., & J, O. W. (2006). Absence of DNA polymerase theta results in decreased somatic hypermutation frequency and altered mutation patterns in Ig genes. *DNA Repair (Amst), 5*(11), 1384-1391. doi:10.1016/j.dnarep.2006.06.006

Masuda, K., Ouchida, R., Yokoi, M., Hanaoka, F., Azuma, T., & Wang, J. Y. (2008). DNA polymerase eta is a limiting factor for A:T mutations in Ig genes and contributes to antibody affinity maturation. *Eur J Immunol, 38*(10), 2796-2805. doi:10.1002/eji.200838502

Maul, R. W., & Gearhart, P. J. (2010a). AID and somatic hypermutation. *Adv Immunol, 105*, 159-191. doi:10.1016/S0065-2776(10)05006-6

Maul, R. W., & Gearhart, P. J. (2010b). Controlling somatic hypermutation in immunoglobulin variable and switch regions. *Immunol Res, 47*(1-3), 113-122. doi:10.1007/s12026-009-8142-5

Maul, R. W., & Gearhart, P. J. (2014). Refining the Neuberger model: Uracil processing by activated B cells. *Eur J Immunol, 44*(7), 1913-1916. doi:10.1002/eji.201444813

Maul, R. W., Saribasak, H., Martomo, S. A., McClure, R. L., Yang, W., Vaisman, A., . . . Gearhart, P. J. (2011). Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions. *Nat Immunol, 12*(1), 70-76. doi:10.1038/ni.1970

Mayadas, T. N., Cullere, X., & Lowell, C. A. (2014). The multifaceted functions of neutrophils. *Annu Rev Pathol, 9*, 181-218. doi:10.1146/annurev-pathol-020712-164023

Mayorov, V. I., Rogozin, I. B., Adkison, L. R., & Gearhart, P. J. (2005). DNA polymerase eta contributes to strand bias of mutations of A versus T in immunoglobulin genes. *J Immunol, 174*(12), 7781-7786.

McCormick, T. S., & Weinberg, A. (2010). Epithelial cell-derived antimicrobial peptides are multifunctional agents that bridge innate and adaptive immunity. *Periodontol 2000, 54*(1), 195-206. doi:10.1111/j.1600-0757.2010.00373.x

Methot, S. P., & Di Noia, J. M. (2017). Molecular Mechanisms of Somatic Hypermutation and Class Switch Recombination. *Adv Immunol, 133*, 37-87. doi:10.1016/bs.ai.2016.11.002

Miethke, M., & Skerra, A. (2010). Neutrophil gelatinase-associated lipocalin expresses antimicrobial activity by interfering with L-norepinephrine-mediated bacterial iron acquisition. *Antimicrob Agents Chemother, 54*(4), 1580-1589. doi:10.1128/AAC.01158-09

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., & Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell, 102*(5), 553-563.

Muramatsu, M., Sankaranand, V. S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N. O., & Honjo, T. (1999). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J Biol Chem, 274*(26), 18470-18476.

Nemazee, D. (2006). Receptor editing in lymphocyte development and central tolerance. *Nat Rev Immunol, 6*(10), 728-740. doi:10.1038/nri1939

Neuberger, M. S. (2008). Antibody diversification by somatic mutation: from Burnet onwards. *Immunol Cell Biol, 86*(2), 124-132. doi:10.1038/sj.icb.7100160

Neuberger, M. S., Di Noia, J. M., Beale, R. C., Williams, G. T., Yang, Z., & Rada, C. (2005). Somatic hypermutation at A.T pairs: polymerase error versus dUTP incorporation. *Nat Rev Immunol, 5*(2), 171-178. doi:10.1038/nri1553

Neuberger, M. S., & Milstein, C. (1995). Somatic hypermutation. *Curr Opin Immunol, 7*(2), 248-254.

Newton, K., & Dixit, V. M. (2012). Signaling in innate immunity and inflammation. *Cold Spring Harb Perspect Biol, 4*(3). doi:10.1101/cshperspect.a006049

Niewold, T. B., Clark, D. N., Salloum, R., & Poole, B. D. (2010). Interferon alpha in systemic lupus erythematosus. *J Biomed Biotechnol, 2010*, 948364. doi:10.1155/2010/948364

Niida, H., Shimada, M., Murakami, H., & Nakanishi, M. (2010). Mechanisms of dNTP supply that play an essential role in maintaining genome integrity in eukaryotic cells. *Cancer Sci, 101*(12), 2505-2509. doi:10.1111/j.1349-7006.2010.01719.x

Nordlund, P., & Reichard, P. (2006). Ribonucleotide reductases. *Annu Rev Biochem, 75*, 681-706. doi:10.1146/annurev.biochem.75.103004.142443

Orcesi, S., La Piana, R., & Fazzi, E. (2009). Aicardi-Goutieres syndrome. *Br Med Bull, 89*, 183-201. doi:10.1093/bmb/ldn049

Orthwein, A., & Di Noia, J. M. (2012). Activation induced deaminase: how much and where? *Semin Immunol, 24*(4), 246-254. doi:10.1016/j.smim.2012.05.001

Parijs, V., Refaeli, Y., Lord, J. D., Nelson, B. H., Abbas, A. K., & Baltimore, D. (1999). Uncoupling IL-2 signals that regulate T cell proliferation, survival, and Fas-mediated activation-induced cell death. *Immunity, 11*(3), 281-288.

Parkin, J., & Cohen, B. (2001). An overview of the immune system. *Lancet, 357*(9270), 1777-1789. doi:10.1016/S0140-6736(00)04904-7

Paull, T. T. (2018). 20 Years of Mre11 Biology: No End in Sight. *Mol Cell, 71*(3), 419-427. doi:10.1016/j.molcel.2018.06.033

Pauls, E., Badia, R., Torres-Torronteras, J., Ruiz, A., Permanyer, M., Riveira-Munoz, E., . . . Este, J. A. (2014a). Palbociclib, a selective inhibitor of cyclin-dependent kinase4/6, blocks HIV-1 reverse transcription through the control of sterile alpha motif and HD domain-containing protein-1 (SAMHD1) activity. *AIDS, 28*(15), 2213-2222. doi:10.1097/QAD.0000000000000399

Pauls, E., Ruiz, A., Badia, R., Permanyer, M., Gubern, A., Riveira-Munoz, E., . . . Este, J. A. (2014b). Cell cycle control and HIV-1 susceptibility are linked by CDK6-dependent CDK2 phosphorylation of SAMHD1 in myeloid and lymphoid cells. *J Immunol, 193*(4), 1988-1997. doi:10.4049/jimmunol.1400873

Pauls, E., Ruiz, A., Riveira-Munoz, E., Permanyer, M., Badia, R., Clotet, B., . . . Este, J. A. (2014c). p21 regulates the HIV-1 restriction factor SAMHD1. *Proc Natl Acad Sci U S A, 111*(14), E1322-1324. doi:10.1073/pnas.1322059111

Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., . . . Nussenzweig, M. C. (2010). Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell, 143*(1), 122-133. doi:10.1016/j.cell.2010.09.017

Pavri, R., & Nussenzweig, M. C. (2011). AID targeting in antibody diversity. *Adv Immunol, 110*, 1-26. doi:10.1016/B978-0-12-387663-8.00005-3

Pena-Diaz, J., Bregenhorn, S., Ghodgaonkar, M., Follonier, C., Artola-Boran, M., Castor, D., . . . Jiricny, J. (2012). Noncanonical mismatch repair as a source of genomic instability in human cells. *Mol Cell, 47*(5), 669-680. doi:10.1016/j.molcel.2012.07.006

Pena-Diaz, J., & Jiricny, J. (2012). Mammalian mismatch repair: error-free or error-prone? *Trends Biochem Sci, 37*(5), 206-214. doi:10.1016/j.tibs.2012.03.001

Petermann, E., Orta, M. L., Issaeva, N., Schultz, N., & Helleday, T. (2010). Hydroxyurea-stalled replication forks become progressively inactivated and require two different RAD51-mediated pathways for restart and repair. *Mol Cell, 37*(4), 492-502. doi:10.1016/j.molcel.2010.01.021

Petersen-Mahrt, S. K., Harris, R. S., & Neuberger, M. S. (2002). AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. *Nature, 418*(6893), 99-103. doi:10.1038/nature00862

Phan, T. G., Amesbury, M., Gardam, S., Crosbie, J., Hasbold, J., Hodgkin, P. D., . . . Brink, R. (2003). B cell receptor-independent stimuli trigger immunoglobulin (Ig) class switch recombination and production of IgG autoantibodies by anergic self-reactive B cells. *J Exp Med, 197*(7), 845-860. doi:10.1084/jem.20022144

Phung, Q. H., Winter, D. B., Cranston, A., Tarone, R. E., Bohr, V. A., Fishel, R., & Gearhart, P. J. (1998). Increased hypermutation at G and C nucleotides in immunoglobulin variable genes from mice deficient in the MSH2 mismatch repair protein. *J Exp Med, 187*(11), 1745-1751.

Pieper, K., Grimbacher, B., & Eibel, H. (2013). B-cell biology and development. *J Allergy Clin Immunol, 131*(4), 959-971. doi:10.1016/j.jaci.2013.01.046

Pitman, R. S., & Blumberg, R. S. (2000). First line of defense: the role of the intestinal epithelium as an active component of the mucosal immune system. *J Gastroenterol, 35*(11), 805-814.

Pontarin, G., Ferraro, P., Bee, L., Reichard, P., & Bianchi, V. (2012). Mammalian ribonucleotide reductase subunit p53R2 is required for mitochondrial DNA replication and DNA repair in quiescent cells. *Proc Natl Acad Sci U S A, 109*(33), 13302-13307. doi:10.1073/pnas.1211289109

Powell, R. D., Holland, P. J., Hollis, T., & Perrino, F. W. (2011). Aicardi-Goutieres syndrome gene and HIV-1 restriction factor SAMHD1 is a dGTP-regulated deoxynucleotide triphosphohydrolase. *J Biol Chem, 286*(51), 43596-43600. doi:10.1074/jbc.C111.317628

Rada, C., Di Noia, J. M., & Neuberger, M. S. (2004). Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-

focused phase of somatic mutation. *Mol Cell, 16*(2), 163-171. doi:10.1016/j.molcel.2004.10.011

Rada, C., Ehrenstein, M. R., Neuberger, M. S., & Milstein, C. (1998). Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity, 9*(1), 135-141.

Rada, C., Williams, G. T., Nilsen, H., Barnes, D. E., Lindahl, T., & Neuberger, M. S. (2002). Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr Biol, 12*(20), 1748-1755.

Rajagopal, D., Maul, R. W., Ghosh, A., Chakraborty, T., Khamlichi, A. A., Sen, R., & Gearhart, P. J. (2009). Immunoglobulin switch mu sequence causes RNA polymerase II accumulation and reduces dA hypermutation. *J Exp Med, 206*(6), 1237-1244. doi:10.1084/jem.20082514

Ramesh, V., Bernardi, B., Stafa, A., Garone, C., Franzoni, E., Abinun, M., . . . Crow, Y. J. (2010). Intracerebral large artery disease in Aicardi-Goutieres syndrome implicates SAMHD1 in vascular homeostasis. *Dev Med Child Neurol, 52*(8), 725-732. doi:10.1111/j.1469-8749.2010.03727.x

Ramiro, A. R., Stavropoulos, P., Jankovic, M., & Nussenzweig, M. C. (2003). Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nat Immunol, 4*(5), 452-456. doi:10.1038/ni920

Rampazzo, C., Miazzi, C., Franzolin, E., Pontarin, G., Ferraro, P., Frangini, M., . . . Bianchi, V. (2010). Regulation by degradation, a cellular defense against deoxyribonucleotide pool imbalances. *Mutat Res, 703*(1), 2-10. doi:10.1016/j.mrgentox.2010.06.002

Rampazzo, C., Tozzi, M. G., Dumontet, C., & Jordheim, L. P. (2015). The druggability of intracellular nucleotide-degrading enzymes. *Cancer Chemother Pharmacol*. doi:10.1007/s00280-015-2921-6

Ravot, E., Comolli, G., Lori, F., & Lisziewicz, J. (2002). High efficiency lentiviral gene delivery in non-dividing cells by deoxynucleoside treatment. *J Gene Med, 4*(2), 161-169.

Rehwinkel, J., Maelfait, J., Bridgeman, A., Rigby, R., Hayward, B., Liberatore, R. A., . . . Reis e Sousa, C. (2013). SAMHD1-dependent retroviral control and escape in mice. *EMBO J, 32*(18), 2454-2462. doi:10.1038/emboj.2013.163

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., . . . Durandy, A. (2000). Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell, 102*(5), 565-575.

Rice, G. I., Bond, J., Asipu, A., Brunette, R. L., Manfield, I. W., Carr, I. M., . . . Crow, Y. J. (2009). Mutations involved in Aicardi-Goutieres syndrome implicate SAMHD1 as regulator of the innate immune response. *Nat Genet, 41*(7), 829-832. doi:10.1038/ng.373

Roa, S., Avdievich, E., Peled, J. U., Maccarthy, T., Werling, U., Kuang, F. L., . . . Scharff, M. D. (2008). Ubiquitylated PCNA plays a role in somatic hypermutation and class-switch recombination and is required for meiotic progression. *Proc Natl Acad Sci U S A, 105*(42), 16248-16253. doi:10.1073/pnas.0808182105

Rogozin, I. B., & Diaz, M. (2004). Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. *J Immunol, 172*(6), 3382-3384.

Roy, S., de Melo, A. J., Xu, Y., Tadi, S. K., Negrel, A., Hendrickson, E., . . . Meek, K. (2015). XRCC4/XLF Interaction Is Variably Required for DNA Repair and Is Not Required for Ligase IV Stimulation. *Mol Cell Biol, 35*(17), 3017-3028. doi:10.1128/MCB.01503-14

Ryoo, J., Choi, J., Oh, C., Kim, S., Seo, M., Kim, S. Y., . . . Ahn, K. (2014). The ribonuclease activity of SAMHD1 is required for HIV-1 restriction. *Nature Medicine, 20*(8), 936-941. doi:10.1038/nm.3626

Sakaue-Sawano, A., Kurokawa, H., Morimura, T., Hanyu, A., Hama, H., Osawa, H., . . . Miyawaki, A. (2008). Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell, 132*(3), 487-498. doi:10.1016/j.cell.2007.12.033

Sattler, U., Frit, P., Salles, B., & Calsou, P. (2003). Long-patch DNA repair synthesis during base excision repair in mammalian cells. *EMBO Rep, 4*(4), 363-367. doi:10.1038/sj.embor.embor796

Schrader, C. E., Guikema, J. E., Linehan, E. K., Selsing, E., & Stavnezer, J. (2007). Activation-induced cytidine deaminase-dependent DNA breaks in class switch recombination occur during G1 phase of the cell cycle and depend upon mismatch repair. *J Immunol, 179*(9), 6064-6071.

Schroeder, H. W., Jr., & Cavacini, L. (2010). Structure and function of immunoglobulins. *J Allergy Clin Immunol, 125*(2 Suppl 2), S41-52. doi:10.1016/j.jaci.2009.09.046

Seamon, K. J., Hansen, E. C., Kadina, A. P., Kashemirov, B. A., McKenna, C. E., Bumpus, N. N., & Stivers, J. T. (2014). Small molecule inhibition of SAMHD1 dNTPase by tetramer destabilization. *J Am Chem Soc, 136*(28), 9822-9825. doi:10.1021/ja5035717

Seamon, K. J., Sun, Z., Shlyakhtenko, L. S., Lyubchenko, Y. L., & Stivers, J. T. (2015). SAMHD1 is a single-stranded nucleic acid binding protein with no active site-

associated nuclease activity. *Nucleic Acids Res, 43*(13), 6486-6499. doi:10.1093/nar/gkv633

Sharbeen, G., Cook, A. J., Lau, K. K., Raftery, J., Yee, C. W., & Jolly, C. J. (2010). Incorporation of dUTP does not mediate mutation of A:T base pairs in Ig genes in vivo. *Nucleic Acids Res, 38*(22), 8120-8130. doi:10.1093/nar/gkq682

Sharbeen, G., Yee, C. W., Smith, A. L., & Jolly, C. J. (2012). Ectopic restriction of DNA repair reveals that UNG2 excises AID-induced uracils predominantly or exclusively during G1 phase. *J Exp Med, 209*(5), 965-974. doi:10.1084/jem.20112379

Shen, H. M., Bozek, G., Pinkert, C. A., McBride, K., Wang, L., Kenter, A., & Storb, U. (2008). Expression of AID transgene is regulated in activated B cells but not in resting B cells and kidney. *Mol Immunol, 45*(7), 1883-1892.

Shen, H. M., 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. doi:10.1128/MCB.25.24.10815-10821.2005

Shulman, Z., Gitlin, A. D., Weinstein, J. S., Lainez, B., Esplugues, E., Flavell, R. A., . . . Nussenzweig, M. C. (2014). Dynamic signaling by T follicular helper cells during germinal center B cell selection. *Science, 345*(6200), 1058-1062. doi:10.1126/science.1257861

St Gelais, C., de Silva, S., Hach, J. C., White, T. E., Diaz-Griffero, F., Yount, J. S., & Wu, L. (2014). Identification of cellular proteins interacting with the retroviral restriction factor SAMHD1. *J Virol, 88*(10), 5834-5844. doi:10.1128/JVI.00155-14

St Gelais, C., & Wu, L. (2011). SAMHD1: a new insight into HIV-1 restriction in myeloid cells. *Retrovirology, 8*, 55. doi:10.1186/1742-4690-8-55

Stavnezer, J., & Schrader, C. E. (2006). Mismatch repair converts AID-instigated nicks to double-strand breaks for antibody class-switch recombination. *Trends Genet, 22*(1), 23-28. doi:10.1016/j.tig.2005.11.002

Stavnezer, J., & Schrader, C. E. (2014). IgH chain class switch recombination: mechanism and regulation. *J Immunol, 193*(11), 5370-5378. doi:10.4049/jimmunol.1401849

Stillman, B. (2013). Deoxynucleoside triphosphate (dNTP) synthesis and destruction regulate the replication of both cell and virus genomes. *Proc Natl Acad Sci U S A, 110*(35), 14120-14121. doi:10.1073/pnas.1312901110

Storb, U., & Stavnezer, J. (2002). Immunoglobulin genes: generating diversity with AID and UNG. *Curr Biol, 12*(21), R725-727.

Striz, I., & Trebichavsky, I. (2004). Calprotectin - a pleiotropic molecule in acute and chronic inflammation. *Physiol Res, 53*(3), 245-253.

Sze, A., Olagnier, D., Lin, R., van Grevenynghe, J., & Hiscott, J. (2013). SAMHD1 host restriction factor: a link with innate immune sensing of retrovirus infection. *J Mol Biol, 425*(24), 4981-4994. doi:10.1016/j.jmb.2013.10.022

Takeuchi, O., & Akira, S. (2010). Pattern recognition receptors and inflammation. *Cell, 140*(6), 805-820. doi:10.1016/j.cell.2010.01.022

Tanaka, H., Arakawa, H., Yamaguchi, T., Shiraishi, K., Fukuda, S., Matsui, K., . . . Nakamura, Y. (2000). A ribonucleotide reductase gene involved in a p53-dependent cell-cycle checkpoint for DNA damage. *Nature, 404*(6773), 42-49. doi:10.1038/35003506

Taniguchi, T., & Takaoka, A. (2002). The interferon-alpha/beta system in antiviral responses: a multimodal machinery of gene regulation by the IRF family of transcription factors. *Curr Opin Immunol, 14*(1), 111-116.

Tasher, D., & Dalal, I. (2012). The genetic basis of severe combined immunodeficiency and its variants. *Appl Clin Genet, 5*, 67-80. doi:10.2147/TACG.S18693

Thientosapol, E. S., Sharbeen, G., Lau, K. K. E., Bosnjak, D., Durack, T., Stevanovski, I., . . . Jolly, C. J. (2017). Proximity to AGCT sequences dictates MMR-independent versus MMR-dependent mechanisms for AID-induced mutation via UNG2. *Nucleic Acids Res, 45*(6), 3146-3157. doi:10.1093/nar/gkw1300

Tran, T. H., Nakata, M., Suzuki, K., Begum, N. A., Shinkura, R., Fagarasan, S., . . . Nagaoka, H. (2010). B cell-specific and stimulation-responsive enhancers derepress Aicda by overcoming the effects of silencers. *Nat Immunol, 11*(2), 148-154. doi:10.1038/ni.1829

Tungler, V., Staroske, W., Kind, B., Dobrick, M., Kretschmer, S., Schmidt, F., . . . Lee-Kirsch, M. A. (2013). Single-stranded nucleic acids promote SAMHD1 complex formation. *J Mol Med (Berl), 91*(6), 759-770. doi:10.1007/s00109-013-0995-3

Turvey, S. E., & Broide, D. H. (2010). Innate immunity. *J Allergy Clin Immunol, 125*(2 Suppl 2), S24-32. doi:10.1016/j.jaci.2009.07.016

Vert, J. P., Foveau, N., Lajaunie, C., & Vandenbrouck, Y. (2006). An accurate and interpretable model for siRNA efficacy prediction. *BMC Bioinformatics, 7*, 520. doi:10.1186/1471-2105-7-520

Wang, Kieffer-Kwon, K. R., Oliveira, T. Y., Mayer, C. T., Yao, K., Pai, J., . . . Robbiani, D. F. (2017). The cell cycle restricts activation-induced cytidine deaminase activity to early G1. *J Exp Med, 214*(1), 49-58. doi:10.1084/jem.20161649

Wang, Rada, C., & Neuberger, M. S. (2010). Altering the spectrum of immunoglobulin V gene somatic hypermutation by modifying the active site of AID. *J Exp Med, 207*(1), 141-153. doi:10.1084/jem.20092238

Weill, J. C., & Reynaud, C. A. (2008). DNA polymerases in adaptive immunity. *Nat Rev Immunol, 8*(4), 302-312. doi:10.1038/nri2281

White, T. E., Brandariz-Nunez, A., Valle-Casuso, J. C., Amie, S., Nguyen, L., Kim, B., . . . Diaz-Griffero, F. (2013a). Contribution of SAM and HD domains to retroviral restriction mediated by human SAMHD1. *Virology, 436*(1), 81-90. doi:10.1016/j.virol.2012.10.029

White, T. E., Brandariz-Nunez, A., Valle-Casuso, J. C., Amie, S., Nguyen, L. A., Kim, B., . . . Diaz-Griffero, F. (2013b). The retroviral restriction ability of SAMHD1, but not its deoxynucleotide triphosphohydrolase activity, is regulated by phosphorylation. *Cell Host Microbe, 13*(4), 441-451. doi:10.1016/j.chom.2013.03.005

Wiedemann, E. M., Peycheva, M., & Pavri, R. (2016). DNA Replication Origins in Immunoglobulin Switch Regions Regulate Class Switch Recombination in an R-Loop-Dependent Manner. *Cell Rep, 17*(11), 2927-2942. doi:10.1016/j.celrep.2016.11.041

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

Wilson, T. M., Vaisman, A., Martomo, S. A., Sullivan, P., Lan, L., Hanaoka, F., . . . Gearhart, P. J. (2005). MSH2-MSH6 stimulates DNA polymerase eta, suggesting a role for A:T mutations in antibody genes. *J Exp Med, 201*(4), 637-645. doi:10.1084/jem.20042066

Wittmann, S., Behrendt, R., Eissmann, K., Volkmann, B., Thomas, D., Ebert, T., . . . Gramberg, T. (2015). Phosphorylation of murine SAMHD1 regulates its antiretroviral activity. *Retrovirology, 12*, 103. doi:10.1186/s12977-015-0229-6

Woodrick, J., Gupta, S., Camacho, S., Parvathaneni, S., Choudhury, S., Cheema, A., . . . Roy, R. (2017). A new sub-pathway of long-patch base excision repair involving 5' gap formation. *Embo J, 36*(11), 1605-1622. doi:10.15252/embj.201694920

Wu, L. (2013a). Cellular and Biochemical Mechanisms of the Retroviral Restriction Factor SAMHD1. *ISRN Biochem, 2013*. doi:10.1155/2013/728392

Wu, L. (2013b). SAMHD1 knockout mice: modeling retrovirus restriction in vivo. *Retrovirology, 10*, 142. doi:10.1186/1742-4690-10-142

Xu, X., Page, J. L., Surtees, J. A., Liu, H., Lagedrost, S., Lu, Y., . . . Weiss, R. S. (2008). Broad overexpression of ribonucleotide reductase genes in mice specifically induces lung neoplasms. *Cancer Res, 68*(8), 2652-2660. doi:10.1158/0008-5472.CAN-07-5873

Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H. W., . . . Casellas, R. (2011). Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat Immunol, 12*(1), 62-69. doi:10.1038/ni.1964

Yan, J., Hao, C., DeLucia, M., Swanson, S., Florens, L., Washburn, M. P., . . . Skowronski, J. (2015). CyclinA2-Cyclin-dependent Kinase Regulates SAMHD1 Protein Phosphohydrolase Domain. *J Biol Chem, 290*(21), 13279-13292. doi:10.1074/jbc.M115.646588

Yang, K., Guo, R., & Xu, D. (2016). Non-homologous end joining: advances and frontiers. *Acta Biochim Biophys Sin (Shanghai), 48*(7), 632-640. doi:10.1093/abbs/gmw046

Yavuz, S., Yavuz, A. S., Kraemer, K. H., & Lipsky, P. E. (2002). The role of polymerase eta in somatic hypermutation determined by analysis of mutations in a patient with xeroderma pigmentosum variant. *J Immunol, 169*(7), 3825-3830.

Yeap, L. S., Hwang, J. K., Du, Z., Meyers, R. M., Meng, F. L., Jakubauskaite, A., . . . Alt, F. W. (2015). Sequence-Intrinsic Mechanisms that Target AID Mutational Outcomes on Antibody Genes. *Cell, 163*(5), 1124-1137. doi:10.1016/j.cell.2015.10.042

Yu, X. F., Yu, Q. C., Essex, M., & Lee, T. H. (1991). The vpx gene of simian immunodeficiency virus facilitates efficient viral replication in fresh lymphocytes and macrophage. *J Virol, 65*(9), 5088-5091.

Zarrin, A. A., Alt, F. W., Chaudhuri, J., Stokes, N., Kaushal, D., Du Pasquier, L., & Tian, M. (2004). An evolutionarily conserved target motif for immunoglobulin class-switch recombination. *Nat Immunol, 5*(12), 1275-1281. doi:10.1038/ni1137

Zeng, X., Negrete, G. A., Kasmer, C., Yang, W. W., & Gearhart, P. J. (2004). Absence of DNA polymerase eta reveals targeting of C mutations on the nontranscribed strand in immunoglobulin switch regions. *J Exp Med, 199*(7), 917-924. doi:10.1084/jem.20032022

Zhang, H., Duan, L. X., Dornadula, G., & Pomerantz, R. J. (1995). Increasing transduction efficiency of recombinant murine retrovirus vectors by initiation

of endogenous reverse transcription: potential utility for genetic therapies. *J Virol, 69*(6), 3929-3932.

Zhang, Y., He, Q., Hu, Z., Feng, Y., Fan, L., Tang, Z., . . . Zhang, L. (2016). Long noncoding RNA LINP1 regulates repair of DNA double-strand breaks in triple-negative breast cancer. *Nat Struct Mol Biol, 23*(6), 522-530. doi:10.1038/nsmb.3211

Zhou, L., Lopes, J. E., Chong, M. M., Ivanov, II, Min, R., Victora, G. D., . . . Littman, D. R. (2008). TGF-beta-induced Foxp3 inhibits T(H)17 cell differentiation by antagonizing RORgammat function. *Nature, 453*(7192), 236-240. doi:10.1038/nature06878