Variations in multiple genes associate with CVID and IgAD

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Steven Offer

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Reuben S. Harris, PhD
Thesis Advisor

September 2009

ACKNOWLEDGEMENTS

I would like to acknowledge my current advisor, Dr. Reuben Harris, and past advisors Dr. Timothy Behrens and Dr. Patrick Gaffney, for their support while at the University of Minnesota. I would also like to thank my collaborators, Dr. Lennart Hammarström and Dr. Qiang Pan-Hammarström of the Karolinska Institute, for access to the resources of the Karolinska Institute and for guidance throughout my entire graduate career.

Next, I would like to thank my thesis committee – Brian VanNess, Deanna Koepp, Erik Peterson, and Michael Farrar. Their advice and support has been pivotal to my continued graduate education.

I would additionally like to thank Robert Graham and Ricardo Ferreira for initiating the studies that comprise Chapter III and Appendix III of this thesis. I learned much of what I know of the principles of population genetics studies from Robert Graham, and for that I am grateful. I would also like to acknowledge William Brown for his work with APEX1, which serves as the basis for Appendix II of this thesis, and Donna MacDuff, for giving me the opportunity to co-author the book chapter "Antibody Gene Diversification by AID-catalyzed DNA Editing."

I would like to thank all of the present and past members of the Harris, Behrens, and Gaffney labs with whom I have had the opportunity to work with. I would like to especially thank William Brown for all of his assistance and generally maintaining the functionality of the Harris lab. I thank Zach Demorest, Rebecca LaRue, Guylaine Haché, April Schumacher, and Mark Stenglein for technical assistance, constructive advice, and for creating an enjoyable work environment. I would also like to thank Brian Ruis and Yongbao Wang from the Hendrickson lab for technical assistance.

I would finally like to acknowledge the University of Minnesota Graduate School for awarding me a Doctoral Dissertation Fellowship, and the University of Minnesota Medical School for the Curtis L. Carlson Medical Research and Education Program Award.

DEDICATION

This dissertation is dedicated to my parents,

Philip and Gladys,

who have supported me completely in every endeavor.

ABSTRACT

A diverse antibody repertoire is paramount to an effective adaptive immune response in higher eukaryotes. This diversity is generated through the orchestrated introduction of somatic DNA mutations, rearrangements, and deletions in specific regions of the B cell genome. Class switch recombination (CSR) is the process by which a B cell alters its antibody-encoding DNA to produce a new class (isotype) of antibodies with a different effector function. This is accomplished by replacing the μ constant region (encoding for the IgM antibody isotype) with a downstream α, γ, or ε constant region (encoding for IgA, IgG, or IgE, respectively). The study of inherited antibody deficiency syndromes has provided valuable insight into the signaling cascades that stimulate CSR and the DNA metabolism machinery that coordinates CSR at the molecular level. The goal of this thesis was to improve our understanding of the molecular events of CSR through the study of two antibody deficiency syndromes, IgA deficiency (IgAD) and common variable immunodeficiency (CVID).

The tumor necrosis factor-like B cell surface receptor, TACI, is important for B cell activation independent of stimulation by T cells. Through a large-scale genetic association study of non-synonymous *TACI* polymorphisms, I provided evidence that the heterozygous *TACI* sequence variants p.C104R, p.A181E, and c.204insA constitute risk factors for the development of CVID, but likely have only a minor role, if any, in the development of IgAD.

Next, prompted by the knowledge that a subset of well characterized antibody deficiencies (termed hyper-IgM syndromes) were due to defects in the DNA metabolism genes *AID* and *UNG2*, I hypothesized that a subset a IgAD and CVID cases could be due to defects in additional DNA repair genes. In an association study of 27 DNA metabolism genes, I noted significant association with IgAD and/or CVID for markers in *AID* and genes of the mismatch repair pathways (*MSH2* and *MLH1*), the MRN complex (MRE11-RAD50-NBS1), and the homologous recombination pathway. Subsequent resequencing of associated pathways yielded a number of novel non-synonymous alleles in *MSH2*, *MLH1*, *RAD50*, and *NBS1* that were IgAD/CVID-specific.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. GENERAL INTRODUCTION

Multiple facets of the immune system coordinate in response to infection

The human immune system is constantly challenged with the detection and eradication of a virtually limitless assortment of potentially infectious pathogens, including bacteria, viruses, and other microorganisms. To provide immediate protection against infection, higher mammals such as humans have evolved a number of defenses against pathogens, both innate and adaptive. The first line of defense is the protective physical barrier formed by the epithelial cells of our skin, gastrointestinal tract, respiratory system, and urogenital tract. To provide additional protection against colonization, some specialized cells in the epithelium secrete anti-microbial compounds. If a pathogen is able to breach the epithelial barrier and infect the host, it is usually rapidly recognized and internalized by phagocytic cells of the innate immune system. Specialized phagocytic cells, known as dendritic cells, are capable of initiating a subsequent adaptive immune response [reviewed by (Fearon and Locksley 1996)].

Antibody-producing B cells, along with T cells, comprise the adaptive immune response. Mature B and T cells both display antigen-specific surface receptors (the B cell receptor, BCR, and T cell receptor, TCR, respectively), however B cells are also able to secrete a form of their receptor as antibodies. Thus, B cells provide the antibody (classically known as the humoral) immune response, and T cells confer the cell-mediated immune response. The antibodies produced by B cells are capable of recognizing infected cells and pathogens directly, whereas T cell receptors mainly recognize antigens that are displayed on surfaces of other cells. The T cell mediated immune response is outside of the scope of this thesis, and as such will only be discussed as it pertains to B cell activation.

Somatic DNA alterations generate a diverse, high affinity antibody repertoire during B cell development and activation

Antibodies are composed of four subunits, two heavy (H) chains and two light (L) chains, all linked by disulfide bridges. Antibodies contain two functionally important domains, the variable (V) and constant (C) regions (**Figure I-1**). The V region is responsible for antigen recognition, while the C region determines the isotype of the antibody. Antibodies of different isotypes are localized to different tissues and have different effector mechanisms by which pathogens are cleared.

The primary antibody repertoire is generate by V(D)J recombination during B cell development

B cells are derived from pluripotent hematopoietic stems cells in the bone marrow. During B cell development, the immunoglobulin (Ig) DNA loci that encode the antibody genes must undergo sequential rearrangements to produce functional and effective antibody proteins (**Figure I-2**) [reviewed by (Tonegawa 1983)]. The assembly of the heavy and light immunoglobulin chains is an ordered process, and successful completion of each step is required for the B cells to progress through development and rearrange subsequent segments.

First, the Ig heavy chain is assembled by recombination between distant variable (V), diversity (D), and joining (J) gene segments on human chromosome 14 to create a single, continuous protein-coding exon during the pro-B cell stage (Early et al. 1980; Maki et al. 1980). This is accomplished by first joining a D and J segment during the early pro-B cell stage to generate a late-pro-B cell in which a V segment is abutted with the assembled DJ segments. Effective VDJ rearrangement to produce a functional heavy chain protein is a hallmark of the pre-B cell stage. During the large pre-B cell stage, the H chain is expressed as the pre-B cell receptor with the help of a surrogate light chain. At this time further rearrangements of the H

chain locus are halted, and the pre-B cells divide several times giving rise to the small pre-B cells, which undergo L chain rearrangement. The light chain can either be encoded from chromosome 22 (the lambda light chain) or chromosome 2 (the kappa light chain). Light chain genes do not encode for a D region, so only a single recombination event juxtaposing a V and J region is needed.

Once assembled, the heavy and light chains are expressed at the cell surface as a complete IgM molecule (so named because at this time the constant region of the heavy chain is being encoded by the μ constant region gene segment) during the immature B cell stage. The immature B cells leave the bone marrow and travel to peripheral lympoid tissues where they undergo selection for self-tolerance to ensure they are not reactive against the cells and tissues of the host. Surviving peripheral B cells further differentiate into naïve mature B cells (so called as they have not yet been exposed to antigen) that express IgD, an alternative splice form of the heavy chain transcript, in addition to IgM on their cell surfaces. These naïve B cells migrate into the peripheral blood and lymphoid tissue ready to bind and respond to antigen.

Molecular mechanism of V(D)J recombination

Over three million different antibody proteins can be created in humans by combining different V(D)J-assembled H and L chains (Cook and Tomlinson 1995). For the heavy chain locus, most humans encode for 40 V, 25 D, and 6 J regions, and as such, 10,530 different H chain combinations are possible. The kappa chain contains 40 different V regions and 5 J regions; the lambda chain contains 30 V regions and 4 J regions. In all, 320 different L chains (200 kappa and 120 lambda chains) can be produced by B cells. End processing that occurs during recombination can also add additional diversity during V(D)J assembly.

The lymphoid specific RAG1/RAG2 complex initiates V(D)J recombination by introducing site-specific DNA breaks at repeated DNA sites called recombination signal

sequences (RSSs) that lie adjacent to Ig gene segments (**Figure I-2A**) (Agrawal and Schatz 1997; Oettinger et al. 1990). The exposed DNA 3'-hyodroxl attacks a phosphodiester bond on the anti-parallel DNA strand to generate a closed hairpin on the coding end (the end adjacent to the Ig gene segment). These hairpins are subsequently cleaved by the endonuclease Artemis to form double-strand DNA breaks (Moshous et al. 2001). Artemis is activated through phosphorylation by the DNA-dependent protein kinase catalytic subunit (DNA-PKcs) (Blunt et al. 1995). The free DNA ends can be further processed by exonucleases that remove nucleotides and polymerases that add nucleotides. Finally, the free DNA ends are brought together by the NHEJ factors KU70/KU80 and DNA-PKcs where they can be ligated by a complex of X-ray cross complementation factor 4 (XRCC4), DNA ligase IV (LIG4), and XRCC4-like factor (XLF) [reviewed by (Lieber 2008)]. As a byproduct, the intervening DNA containing the RSSs is lost from the genome as a DNA circle.

Antigen recognition by the B cell receptor initiates further diversification

Upon binding antigen in peripheral lymphoid tissues (lymph nodes, spleen and mucosal-associated lymphoid tissues), the B cell receptor (BCR) initiates a cascade of events that initiate further cellular specialization. Bound BCRs internalize antigen where it is processed and presented on the cell surface by MHC (major histocompatibility) class II molecules. Surface-presented antigen can be recognized by helper T cells, also present in the lymph tissue (Garside et al. 1998; Pape et al. 2003). T cells express their own form of antigen-specific receptor, the TCR, which is encoded by genes that undergo somatic rearrangements similar to those of the BCR. The T cells also expresses surface CD40 ligand that in turn activates CD40 receptors on the B cell, stimulating B cell proliferation, and in some cells, differentiation into antibody-secreting plasma cells and memory B cells (Lane et al. 1992; Noelle et al. 1992; Valle et al. 1989). Cytokines secreted by T cells further amplify this response.

Some of the proliferating B cells migrate to the primary lymphoid follicle, located in the outer sections of a lymph node, where they continue to divide forming a germinal center [reviewed by (Batista and Harwood 2009)]. Germinal centers are made up mostly of dividing B cells, but also contain antigen-specific T cells that provide additional survival signals to the B cells. Germinal centers also contain specialized follicular dendritic cells that secrete chemokines to attract additional B cells to the follicle. It is in the germinal center where B cells undergo affinity maturation, somatic hypermutation (SHM) and class switch recombination (CSR). The high affinity B cells that are created in the germinal center will further differentiate into antibody-secreting plasma cells, which will provide protection against the current infection, or memory B cells, which can be rapidly activated and converted to plasma cells in the case of subsequent infections by the same, or closely related, pathogen.

B cells may also directly respond to antigen outside of the germinal center, independent of T cell assistance. T cell independent stimulation generates a more rapid antibody-mediated response. However, the antibodies produced generally have limited affinity for antigen, as the B cells have not undergone repeated rounds of affinity maturation. Furthermore, in the absence of T cell co-stimulation, CSR to certain isotypes is comparatively inefficient. Therefore antibodies generated by T cell independent activation exhibit less specificity due to limited variability and decreased functionality.

Highly specific antibodies are generated by affinity maturation

Upon recognizing antigen in the germinal center, non-templated mutations (mostly point mutations) are introduced into V regions by SHM (**Figure I-2B**). If a mutation creates a BCR with a higher affinity for antigen the expressing cell is able to outcompete other cells for antigen, and in turn, receive higher amounts of positive stimuli from the T cells. In this process of "affinity maturation," cells with high-affinity BCRs proliferate and those with low affinity

receptors do not receive sufficient positive stimuli and undergo apoptosis (Sarvas and Makela 1970; Wagner and Neuberger 1996). Repeated rounds of SHM and clonal selection drive the evolution of B cells that produce high-affinity antibodies.

Some species, such as birds, rabbits, cows, pigs, and horses, gain little variable region diversity during V(D)J recombination as they encode only a single functional V segment and J segment (Arakawa and Buerstedde 2004; Maizels 2005; Reynaud et al. 1985; Reynaud et al. 1987; Reynaud et al. 1994; Thompson and Neiman 1987). Instead theses animals rely upon a pseudogene-templated recombination process called immunoglobulin gene conversion (IGC) to diversify variable region DNA, in addition to SHM (**Figure I-2C**). Highly similar pseudo variable region genes flank the true variable region locus in these species and serve as templates for the specialized homologous recombination-like IGC reaction (sometimes called homeologous recombination). Much of the investigation into the mechanism of IGC has been conducted in the chicken B cell lymphoma line DT40, which constitutively undergoes IGC (Arakawa and Buerstedde 2004; Buerstedde and Takeda 1991; Sale 2004; Sonoda et al. 2001; Takeda et al. 1992). In the currently accepted model of IGC, a DNA double-strand break in the expressed V region is thought to trigger strand invasion into one of the upstream homologous sequences with the net result being the partial replacement with variable region pseudogene sequence.

Class switch recombination generates antibodies with specialized functions

The fourth mechanism of immunoglobulin gene diversification, class switch recombination (CSR), alters the antibody constant region through DNA rearrangements that join the heavy chain variable region with a different downstream constant region to produce antibodies of different isotypes with various effector functions (**Figure I-2D**). There are nine constant regions encoded in the human heavy chain locus (Cμ, Cδ, Cγ3, Cγ1, Cα1, Cγ2, Cγ4,

Cε, and Cα2; 5'->3'). Each of the different isotypes (IgM, IgD, IgG3, IgG1, IgA1, IgG2, IgG4, IgE, and IgA2) has a different effector function to promote clearing of a pathogen. IgA is secreted across epithelial tissues, usually into mucous-containing microenvironments, to neutralize pathogens by preventing them from binding to and entering tissues. Antigen-bound IgE can activate mast cells to initiate an inflammatory response, and IgG stimulates engulfment of bound pathogens by phagocytes. Both IgM and IgG are able to activate the complement cascade upon antigen binding. Lastly, the functional of IgD, which is expressed by alternative splicing of IgM transcripts, has no known function other than to mark the surface of naïve mature B cells.

AID is a DNA cytidine deaminase that directly triggers SHM, IGC, and CSR

A single protein, activation induced deaminase (AID), is responsible for initiating the three distinct processes of SHM, IGC, and CSR at the DNA level [reviwed by (Maizels 2005)]. Consistent with being a requirement for CSR, mice deficient for AID exhibited undetectable levels of switched antibody isotypes compared to heterozygous littermates and isolated B cells failed to undergo CSR when stimulated *in vitro* (Muramatsu et al. 2000). Serum levels of IgM were normal or even elevated, signifying that AID is not required for early B cell development or V(D)J recombination. Additional characterization of $AID^{-/-}$ mice revealed enlarged germinal centers that contained many activated B cells. Variable region DNA sequencing revealed that SHM was also defective. AID was later shown to be essential for a third antibody diversification mechanism IGC using the DT40 chicken B cell lymphoma line (Arakawa et al. 2002; Harris et al. 2002).

AID is a member of a large family of deaminase proteins, the APOBECs, that share the capacity to deaminate cytidine to uridine (Conticello 2008; Harris and Liddament 2004; Jarmuz et al. 2002). When expressed in *E. coli*, AID triggered an increase in C/G to T/A mutations, a

pattern that was enhanced in cells defective for uracil excision repair (Petersen-Mahrt et al. 2002). Biochemical experiments reaffirmed the genetic studies by showing that AID can specifically deaminate cytidine residues on single-stranded DNA, but not double-stranded DNA, RNA, or RNA/DNA hybrids (Bransteitter et al. 2003; Chaudhuri et al. 2003; Dickerson et al. 2003).

The non-canonical uracils introduced by AID undergo further processing by ubiquitous DNA repair pathways converting them to DNA lesions that are substrates for antibody gene diversification. In the widely accepted DNA deamination model, AID deaminates cytosine bases to uracils, which in turn can then be recognized and removed by the base-excision repair (BER) pathway (Petersen-Mahrt et al. 2002). The mismatch-repair (MMR) DNA repair pathway provides a complementary means of uracil metabolism during antibody diversification [(Rada et al. 2004) and reviewed by (Longerich et al. 2006)].

Uracil metabolism by the base-excision repair pathway

In BER-mediated uracil excision, uracil nucleotide glycosylase 2 (UNG2) excises the uracil from the DNA backbone to generate an abasic site. The UNG locus in humans and mice encodes two alternative splice products – UNG1, the ubiquitously expressed mitochondrial form, and UNG2, the nuclear form that shows highest expression in the thymus, testes, placenta, and other rapidly proliferating cells (Nilsen et al. 1997; Otterlei et al. 1998).

Somatic hypermutation at G/C base pairs usually occurs as a mixture of transition (G/C to A/T or A/T to G/C) and transversion (G/C to T/A or C/G; or A/T to C/G or T/A) mutations. Inhibiting the function of UNG2, either by expressing the bacteriophage UNG2 inhibitor protein Ugi or by creating UNG2-deficient cells, caused SHM at G/C bases to become skewed toward transition mutations (Di Noia and Neuberger 2002; Rada et al. 2004; Rada et al. 2002; Saribasak et al. 2006). In one study, inhibition of UNG2 increased the number of transition

mutations at C and G bases to 86% compared to 38% in controls (Di Noia and Neuberger 2002). This mutational bias is consistent with the interpretation that following BER-mediated uracil excision, error prone DNA synthesis can introduce any base at the location of the abasic site. However when the uracil is not removed, DNA synthesis over the uracil produces a transition mutation.

UNG2-deficiencies were also shown to reduce IGC in the DT40 cell line and CSR in mice. Inhibition of UNG2 by Ugi dramatically inhibited IGC in DT40 cells (Di Noia and Neuberger 2004), a result that was later confirmed by the creation of an UNG2-deficient DT40 cell line (Saribasak et al. 2006). The study of UNG2-deficient mice provided additional support for BER-mediated uracil excision during Ig diversification (Rada et al. 2004; Rada et al. 2002). Serum form UNG2-deficient mice contained lower (but measurable) titers of IgG1, IgG3 and IgA compared to controls. When stimulated *in vitro*, B cells from Ung2-deficient mice were impaired in their ability to switch to either IgG3 or IgG1.

Further processing of BER-mediated abasic sites creates DNA breaks

During normal BER, the UNG2-generated abasic site would be recognized by the endonuclease APEX, which is capable of cleaving the DNA phophodiester backbone [reviewed by (Krokan et al. 1997)]. As it is the major endonuclease in vertebrates, APEX1 was originally proposed to act downstream of UNG2 in the DNA deamination model of Ig diversification (Petersen-Mahrt et al. 2002). Both APEX1 and APEX2 (the overall role of APEX2 is largely undefined) have been reported to participate in the DNA cleavage step of CSR (Guikema et al. 2007), however a recent report shows that APEX1 depletion has no effect on CSR efficiency (Sabouri et al. 2009). APEX2-deficient mice exhibit an overall decrease in somatic hypermutation, but no change in transition/transversion mutation rate at G/C base pairs (Sabouri et al. 2009). Given the striking differences between these results and those obtained from

UNG2-deficient mouse stains, one must assume that APEX2 functions in a pathway independent from UNG2 and does not act as an endonuclease. Rather, APEX2 may be functioning as 3'-5' exonuclease that progressively removes DNA bases surrounding AID-induced uracils during SHM.

An alternative complex has also been proposed to promote DNA cleavage downstream of UNG2 in the AID-dependent deamination model, the MRE11-RAD50-NBS1 (MRN) complex (Larson et al. 2005; Yabuki et al. 2005). A number of reports have established that the MRN complex participates in CSR, (Kracker et al. 2005; Lahdesmaki et al. 2004; Pan et al. 2002; Pan-Hammarström et al. 2003; Petersen et al. 2001; Reina-San-Martin et al. 2005), however the exact step at which it functions is still unclear. The MRN complex has key functions in both major pathways of DNA break repair in eukaryotic cells, homologous recombination and non-homologous end joining [reviewed by (Kobayashi et al. 2004)]. The MRN complex has a number of functions during DNA repair, including the processing of DNA breaks, tethering of broken DNA ends, and initiation of cell cycle arrest via ATM recruitment (de Jager et al. 2001; Morales et al. 2005). The ability to cleave single-stranded DNA at abasic sites has also been demonstrated for a complex of MRE11 and RAD50 (Larson et al. 2005).

Overexpression of NBS1 led to accelerated SHM in the human B lymphoma line, Ramos, and increased IGC in DT40 cells (Yabuki et al. 2005). Furthermore, MRE11 was shown to specifically associate with the Ig gene locus in cells undergoing SHM (Larson et al. 2005). A mixture of purified MRE11 and RAD50 was shown to be capable of cleaving single-strand DNA such as that generated at switch regions during CSR (Larson et al. 2005). Later studies have shown that transgenic expression of a truncated, inactive form of NBS1 in an NBS1-deficient background results in reduction in the rate of ICG (Nakahara et al. 2009). Interestingly, the authors were able to restore IGC to wildtype levels by ectopically expressing a 3' to 5' single-strand exonuclease from *E. coli*, SbcB. This result suggests that instead of

introducing double-strand DNA breaks, the MRN complex may be promoting IGC by converting AID-induced lesions into single-strand gaps. This result is similar to that observed for APEX2 deficiency in SHM, and hints that double-strand DNA breaks may not be a requisite intermediate of Ig locus differentiation.

Taken together these reports indicate that BER-mediated uracil excision is an important intermediate for effective SHM, IGC and CSR. However, depletion of UNG2 did not totally obliterate the introduction of V region mutations during SHM. Furthermore, detectable levels of CSR and IGC persisted in UNG2-deficient cells. Therefore while the BER pathway may be the predominant means of uracil metabolism during antibody gene diversification, complementary mechanisms must process a significant subset of uracils into Ig diversification intermediates.

Uracil metabolism by the mismatch repair pathway

Mismatched DNA bases can be directly recognized and excised by mismatch repair pathway machinery. MutSα, a heterodimeric complex of MutS homolog 2 (MSH2) and MSH6, is capable of recognizing the single base U:G mismatches introduced by AID (Wilson et al. 2005). A second mismatch repair recognition complex, MutSβ (a complex of MSH2 and MSH3), mainly recognizes short mismatched patches of DNA and is likely not involved in Ig diversification. Following mismatch recognition, the MutLα complex, composed of post-meiotic segregation 2 (PMS2) and Mut L homolog 1 (MLH1), excises the mismatched base, and DNA polymerases fill the resulting gap.

Prior to the discovery of AID, a number of mouse studies showed that mismatch repair proteins were involved in SHM (Cascalho et al. 1998; Frey et al. 1998; Kim et al. 1999; Kong and Maizels 1999; Phung et al. 1999; Phung et al. 1998; Rada et al. 1998; Wiesendanger et al. 2000). A two-stage model for SHM was postulated and later amended following the discovery of AID (Petersen-Mahrt et al. 2002; Rada et al. 1998). In the first stage, mutations are

introduced at G/C bases following uracil excision by the BER pathway, as discussed above. The second phase is triggered by mutations introduced during the first phase and creates further mutations at G/C and A/T bases by error-prone DNA synthesis. Mutagenesis during the second phase of SHM requires the mismatch repair pathway. In support of this model, mice deficient for both UNG2 and MSH2 (or MSH6) only accumulate C to T and G to A mutations in the variable region, consistent with the incorporation of A opposite U during DNA replication across the unrepaired uracil (Rada et al. 2004; Shen et al. 2006; Wiesendanger et al. 2000; Xue et al. 2006).

Mismatch repair does not appear to be involved in IGC, as disruption of MSH3 or MSH6 does not alter the rate of IGC in DT40 (Arakawa and Buerstedde 2004; Arakawa et al. 2002). However as UNG2-deficiency has been shown to nearly ablate IGC (Saribasak et al. 2006), the effects of mismatch repair-deficiency may be undetectable using single-gene disruption. The construction of UNG/MSH2 double-deficient DT40 cell lines should more clearly define the role of mismatch repair in IGC (if any).

Like SHM, fully functional CSR requires both the BER and mismatch repair pathways. Mice deficient for MSH2, MSH6, MLH1, or PMS2 showed reduced levels of CSR (Ehrenstein and Neuberger 1999; Frey et al. 1998; Rada et al. 1998; Schrader et al. 1999; Vora et al. 1999). CSR was completely ablated in MSH2/UNG2 and MSH6/UNG2 double-deficient mice, strongly suggesting that all CSR-specific DNA lesions are created by either BER or mismatch repair-mediated metabolism of AID-induced uracils (Rada et al. 2004; Shen et al. 2006).

In addition to the gross defect in CSR, switch joints cloned from MMR gene knockout mice exhibited varying levels of microhomology (stretches of sequence similarity between the two donor switch regions). Compared to wildtype littermates, MSH2-deficient mice showed a decreased reliance on microhomology, whereas MLH1 or PMS2-deficient mice showed an increase (Ehrenstein and Neuberger 1999; Schrader et al. 2002). Switch joint microhomology

was comparable to wildtype littermates for MSH6-deificient mice. These data suggest that mismatch recognition by MutSα and processing by MutLα may be separable processes. In additional to DNA cleavage, the mismatch repair pathway may play a role in breakpoint targeting, processing of DNA ends prior to ligation, and/or direction of broken DNA ends to a specific ligation pathway.

The study of primary immunodeficiency due to defective CSR has shed much light on the mechanism of CSR and upstream signaling events

Valuable insights into the molecular mechanisms of Ig diversification have been garnered through the study of inherited human immune deficiency syndromes. The best-defined primary immunodeficiencies are the hyper-IgM syndromes, commonly characterized by reduced serum IgA, IgG, and IgE levels, and often elevated serum IgM levels [reviewed by (Durandy et al. 2006; Erdos et al. 2005)]. The study of AID mutations that underlie hyper-IgM2 has provided much valuable insight into how this protein functions and is regulated during CSR [reviewed by (Durandy et al. 2007)]. Having a broader clinical criteria for diagnosis, IgA deficiency (IgAD), is the most common immunodeficiency, however affected individuals often go undiagnosed due to the relatively subtle immune defect [reviewed by (Castigli and Geha 2006; Salzer and Grimbacher 2006)].

Hyper-IgM syndromes due to mutations in signaling pathways

Three hyper-IgM syndromes are caused by mutations affecting the CD40-CD40 ligand co-stimulatory signaling pathway. The first hyper-IgM syndrome to be characterized, hyper-IgM1, exhibits X-linked inheritance and was mapped to the CD40 ligand (CD40LG) coding region of the X chromosome near to the HPRT locus (Aruffo et al. 1993; Mensink et al. 1987; Padayachee et al. 1992; Padayachee et al. 1993; Pilia et al. 1994). Mutations in CD40 were later

shown to cause the autosomal recessive syndrome hyper-IgM3 (Ferrari et al. 2001; Kutukculer et al. 2003).

Signals received through the CD40 receptor are propagated by the NFκB pathway. Mutations in the NFκB protein IKKγ, also known as NEMO, cause the X-linked hyper-IgM with ectodermal dysplasia (hyper-IgMED) (Jain et al. 2001; Martinez-Pomar et al. 2005; Orstavik et al. 2006; Zonana et al. 2000). In addition to depressed serum IgG and IgA levels, patients exhibit general developmental defects in ectoderm-derived structures leading to abnormal or absent teeth, hair and nails. These disease manifestations are likely due to the ubiquitous nature of IKKγ and NFκB signaling during early development. Signaling through IKKγ leads to nuclear trafficking and activation of NFκB, which can activate transcription of many genes, some of which, like AID, are important for both SHM and CSR. Separation-of-function alleles have been identified as two patients with mutations in IKKγ have been reported with immune deficiency, but no ectodermal dysplasia (Orange et al. 2004; Puel et al. 2006). Interestingly, the X-inactivation status of peripheral blood cells from one hyper-IgMED patient became skewed over time to leave the X chromosome containing the healthy copy of IKKγ active, effectively "curing" the immunodeficiency (Martinez-Pomar et al. 2005).

Hyper-IgM syndromes due to mutations in AID *and* UNG2

Soon after the discovery and characterization of AID in the mouse, it was noted that a subset of autosomal recessive human immunodeficiencies (later called hyper-IgM2) exhibited similar characteristics to the knockout mouse (Revy et al. 2000). These including markedly diminished serum IgA, IgG and IgE levels with normal to elevated IgM, impaired somatic hypermutation, and in many cases lymph node hyperplasia due to enlarged germinal centers. In a genome-wide linkage study conducted in 12 consanguineous families containing a total of 18 hyper-IgM2 patients, a strong linkage peak was observed for microsatellite markers on

chromosome 12p13, the same region to which the gene encoding AID had been mapped (Muto et al. 2000; Revy et al. 2000). Resequencing of the *AID* coding region identified 10 independent non-synonymous alleles in patients. In all, 39 autosomal recessive mutations in AID have been shown to underlie hyper-IgM (Imai et al. 2005; Minegishi et al. 2000; Quartier et al. 2004; Revy et al. 2000; Zhu et al. 2003). Most often, these mutations present in a homozygous state, however compound heterozygous mutations have been detected in a subset of affected individuals.

An autosomal-dominant form of hyper-IgM2 was later linked to a C-terminal deletion of AID caused by a mutation that introduces a premature stop codon at position 190, truncating the last 9 amino acids (Kasahara et al. 2003; Ta et al. 2003). This mutant protein (AID-R190X) was unable to elicit CSR when expressed in AID-deficient mouse B cells (Ta et al. 2003). Interestingly, AID-R190X was, however, able to mutate an artificial SHM substrate using an *in vitro* fibroblast model (Ta et al. 2003). The importance of the C terminal portion of AID for CSR but not SHM (or IGC) was confirmed in studies of the artificially generated AID-L189X mutant (Barreto et al. 2003).

Mutations in the catalytic domain of UNG2 are responsible for the autosomal recessive hyper-IgM5 syndrome (Imai et al. 2003). These patients exhibit clinic criteria indistinguishable from those with hyper-IgM2, except that low levels of SHM are still detectable at G/C residues, most likely due to replication over unprocessed uracil residues.

CVID and IgAD represent a clinically under-defined class of primary immunodeficiency

Common variable immunodeficiency (CVID), the most common primary immunodeficiency requiring medical attention, is characterized by recurrent bacterial infections often accompanied by autoimmune complications and/or lymphoproliferation and splenomegaly [reviewed by (Di Renzo et al. 2004)]. The diagnosis for CVID is based on three clinical criteria:

1) the inability to produce specific antibodies following vaccination or pathogen exposure, 2) markedly decreased serum IgA and IgG levels, and 3) exclusion of other causes of immunodeficiency (Kokron et al. 2004).

Although many cases are asymptomatic, IgA deficiency (IgAD), characterized by absence or a very low level (<7 mg/dL) of serum IgA, is the most common primary immunodeficiency in people of western European descent (Burrows and Cooper 1997; Cunningham-Rundles 2001). IgAD and CVID often coexist in separate members of the same family (Vorechovsky et al. 1995). In some cases, IgAD cases can develop into the more severe CVID, suggesting that at least a subset of CVID and IgAD share a common genetic cause (Vorechovsky et al. 1995).

The molecular causes of CVID and IgAD remain largely unknown. The first identified genetic defect in CVID was a large genomic deletion in the T cell receptor gene "inducible costimulatory receptor" (ICOS) (Grimbacher et al. 2003). ICOS is a costimulatory molecule expressed on T cells that induces production of IL-10 and stimulates secretion of IL-4, IL-5, and IL-6. These cytokines are important for cooperation between B cells and T cells, stimulating CSR, antibody production/secretion, and differentiation into memory and plasma B cells. Like those cells bearing mutations in CD40 or CD40LG, ICOS-deficient B cells do not readily form germinal centers. However, ICOS-deficiency accounts for less than 1% of all CVID cases. Additionally, all carriers likely descended from a common founder as the same pattern of SNPs is detected flanking the deletion in all cases of ICOS-deficiency (Grimbacher et al. 2003; Salzer et al. 2004).

Recessive mutations in another B cell receptor, CD19, have been linked to CVID in a small number of families (van Zelm et al. 2006). CD19, CD21, and CD81 form the B cell co-receptor complex [reviewed by (Cambier et al. 1994; DeFranco et al. 1995)]. Co-ligation of this complex causes it to become cross-linked to the BCR. This enhances BCR signaling by

decreasing the threshold for antigen-dependent B cell activation (Carter and Fearon 1992). Upon BCR binding of antigen, BCR-associated tyrosine kinases phosphorylate the cytoplasmic tail of CD19. In turn, Src-family kinases bind to and processively phosphorylate CD19, eliciting structural changes in the receptor and the recruitment of PI 3 kinases (Fujimoto et al. 2000). These PI 3 kinases, together with those activated through direct BCR signaling, initiate a number of downstream signaling pathways to activate a series of transcription factors. As such, co-activation of the B cell co-receptor complex amplifies signaling through the BCR such that antigen recognition by only a small fraction of the total B cell receptors per cell (0.03%) is sufficient to stimulate proliferation (Carter and Fearon 1992).

Mutations in the tumor necrosis factor receptor (TNFR) superfamily were found to be associated with CVID (Castigli et al. 2007; Castigli et al. 2005; Pan-Hammarström et al. 2007; Salzer et al. 2009; Salzer et al. 2005; Warnatz et al. 2009). Signaling through TNFR superfamily receptors modulates B cell survival, maturation, and CSR (Cancro 2004; Miller et al. 2006). The TNF-ligand B cell-activating factor (BAFF) can bind to three TNF receptors expressed mainly on B cells, BAFF receptor (BAFFR), transmembrane activator and CAML (calcium modulator and cyclophilin ligand) interactor (TACI), and B cell maturation protein A (BCMA) (Gross et al. 2000; Marsters et al. 2000; Schiemann et al. 2001; Yan et al. 2001; Yan et al. 2000). Produced by macrophages and dendritic cells, BAFF provides a key survival signal for B cell maturation. A second ligand, a proliferation-inducing ligand (APRIL), can also bind to TACI and BCMA, but not BAFFR (Hahne et al. 1998; Marsters et al. 2000; Wu et al. 2000). Signaling through TACI can initiate class switch to IgG and IgA independent of T cell stimulation (Castigli et al. 2004; Castigli et al. 2005; Litinskiy et al. 2002). A number of CVID-associated mutations in TACI (mostly dominant) have been identified (Castigli et al. 2007; Castigli et al. 2005; Pan-Hammarström et al. 2007; Salzer et al. 2009; Salzer et al. 2005), and

homozygous BAFFR mutations were also recently detected in two siblings with CVID (Warnatz et al. 2009).

Chapter 1 of this dissertation provides a report of TACI polymorphisms that were found to be associated with CVID in a number of separate patient cohorts (Pan-Hammarström et al. 2007).

Additionally, a number of mutations of the MutS homolog 5 (MSH5) gene were identified in CVID and IgAD patients (Sekine et al. 2007). MSH5, paired with its obligate heterodimerization partner MSH4, forms a stabilizing ring around the cross-shaped homologous recombination intermediates known as Holliday junctions (Snowden et al. 2004). Two of the non-synonymous MSH5 mutations detected in CVID/IgAD patients were always detected together and conferred diminished ability to bind MSH4 (Sekine et al. 2007).

In all, the genetic causes of most cases of CVID and IgAD have yet to be identified. Mutations in TACI account for the highest percentage of identified CVID cases with reports that as many as 10% of CVID cases may be due to known TACI alleles (Pan-Hammarström and Hammarström 2008; Schaffer et al. 2007; Yong et al. 2008). Mutations in ICOS and CD19 likely underlie less than 1% of CVID cases, owing to their autosomal recessive inheritance. Individuals with mutations in MSH5 exhibit variable penetrance of immunodeficiency, suggesting that additional factors may also be required for symptoms to manifest. Overall, over 80% of CVID and IgAD cases are still undefined at the molecular level.

Chapter 2 of this dissertation provides a report of additional DNA repair factor defects associated with CVID and IgAD.

Hypothesis and research question

The overall goal of this body of research was to build upon our understanding of the molecular events that are necessary for successful CSR. I hypothesize that a subset of human

immune deficiency syndromes of unknown etiology are due to mutations in signaling factors that act upstream of CSR, and in DNA metabolism proteins that orchestrate CSR at the molecular level.

Research question – Chapter II:

Our collaborators had previously detected rare TACI mutations in CVID patients. We sought to determine the abundance of these alleles in the general population and to increase the total number of CVID/IgAD patients genotyped. To address this question we genotyped 2,209 healthy controls, 424 CVID patients, and 254 IgAD patients for nonsynonymous TACI mutations.

Patients with the autoimmune disease system lupus erythematosus (SLE) have an elevated risk of immune-deficiency, and immunodeficient patients are more likely to develop autoimmune symptoms than immune-competent individuals. De-regulation of BAFF-mediated signaling in mice can produce either an immune-deficient or autoimmune phenotype (Craxton et al. 2005; Gross et al. 2001; Gross et al. 2000; Lesley et al. 2004; Schneider et al. 2001; Thompson et al. 2000). The apparent link between these two seemingly disparate diseases prompted us to ask if TACI alleles were also contributing to SLE.

To address this question I genotyped 643 SLE patients for known TACI polymorphisms and screened IgAD and autoimmune patients for additional mutations in TACI.

Research question – Chapter III:

DNA metabolism pathway defects cause a subset of hyper-IgM syndromes (hyper-IgM 2 and 5). Additionally, altered CSR is noted in severe DNA damage sensitivity syndromes such as ATLD and NBS. We asked if additional, potentially subtle, defects in DNA repair pathways could underlie a subset of CVID and/or IgAD cases.

To address this question, we conducted a genetic association screen of HapMap tagged markers spread across 27 DNA repair genes in CVID and IgAD affected individuals. I then resequenced the mRNA coding regions of select pathways that showed association. Next, I determined the allele frequency of detected nonsynonymous SNPs in a large cohort of healthy controls and CVID/IgAD patients. Finally, I assessed the ability of a novel allele unique to CVID to respond to DNA damage.

Figures

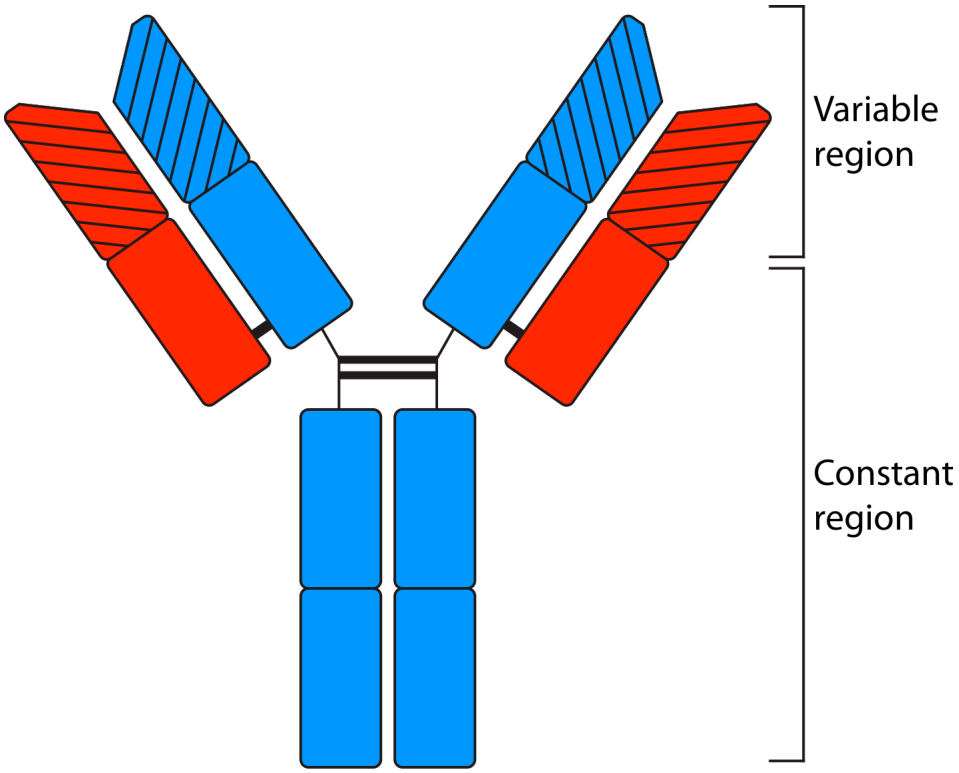


Figure I-1. Antibody schematic. Antibodies are composed of two heavy (H) chain proteins (blue) and two light (L) chain proteins (red). The amino-terminal variable (V) regions (hashed portions) recognize antigen. The isotype of the antibody is determined by the carboxy constant (C) regions of the heavy chain. Disulfide bomnds are depicted as lines.

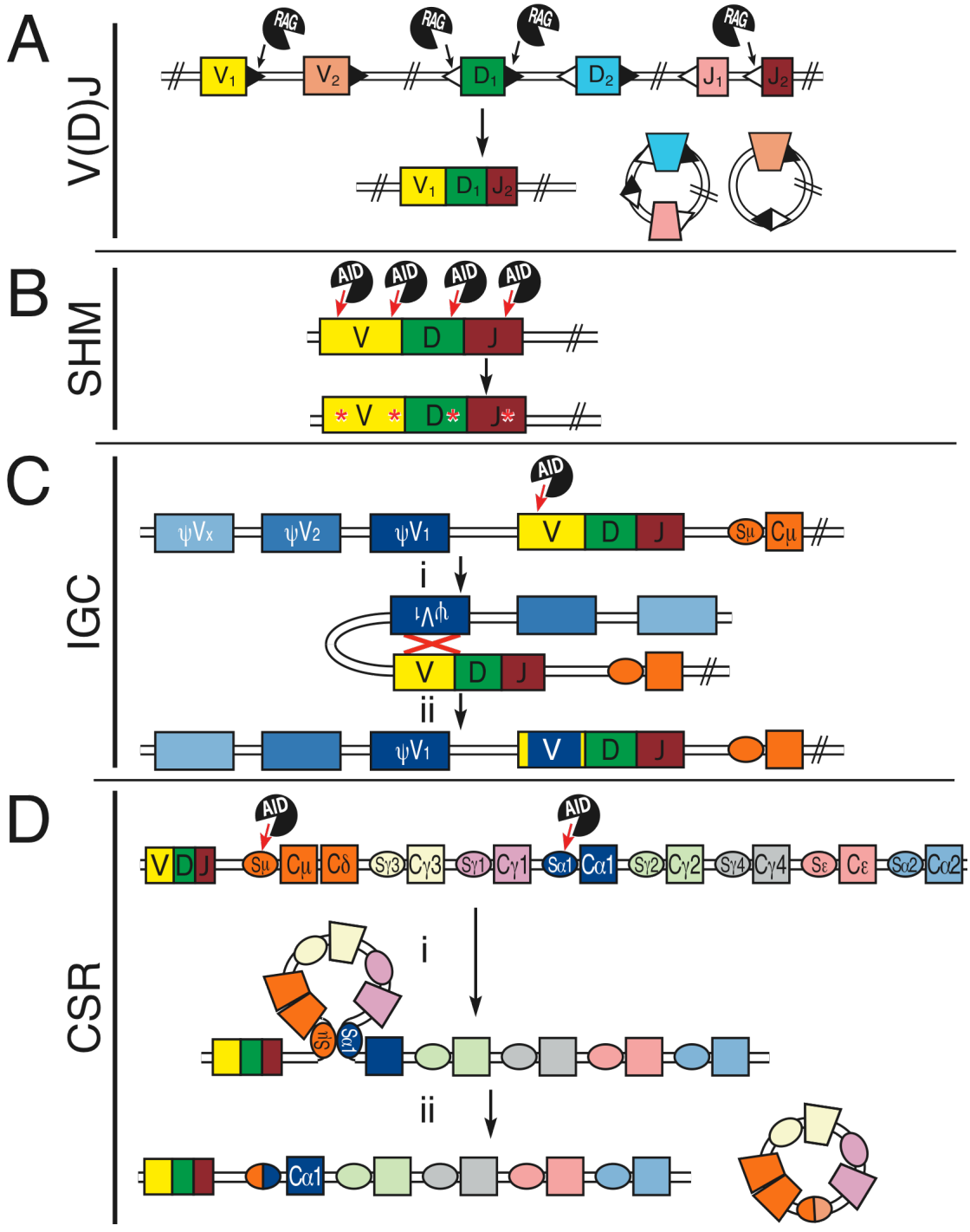

A
V(D)J
RAG
V1
V2
D1
D2
J1
J2
V1 D1 J2
B
SHM
AID
V
D
J
C
IGC
ψVx
ψV2
ψV1
Sμ
Cμ
i
ii
D
CSR
Cμ
Cδ
Sγ3
Cγ3
Sγ1
Cγ1
Sα1
Cα1
Sγ2
Cγ2
Sγ4
Cγ4
Sε
Cε
Sα2
Cα2

Figure I-2. Antibody gene diversification processes.
A. V(D)J recombination. (i, ii) RAG1 and RAG2 initiate the reaction by nicking the DNA at RSS sequences (open and filled triangles). DNA-PK binds to the DNA ends and recruits Artemis. A ligation complex of XRCC4/LIG4/XLF is recruited to complete the non-homologous end joining reaction.
B. Somatic hypermutation (SHM). Point mutations are introduced into the Ig gene variable (V) region by deamination of cytosine bases to uracils. SHM can occur independently or simultaneously with CSR during B cell development.
C. Immunoglobulin gene conversion (IGC). The chicken IgH locus is illustrated. (i) AID-dependent DNA deaminations in the V region creates regions of single-stranded DNA. (ii) A homologous-recombination-like reactions results in the templated replacement of a portion of the V gene with sequence from a ΨV donor region.
D. Class switch recombination (CSR). (i) AID-mediated deamination within the Sμ switch region and a downstream S region (e.g., Sα1) initiates CSR at the DNA level. (ii) Intervening DNA is excised, juxtaposing the V-D-J-assembled heavy chain variable region next to a new C region (e.g., Cα1). A detectable switch circle byproduct is also generated.

CHAPTER II: GENETIC ASSOCIATION OF TACI POLYMORPHISMS WITH CVID AND SYSTEMIC LUPUS ERYTHEMATOSUS

In this chapter, two reports are presented. The first was published in the journal *Nature Genetics* with the title *Reexamining the role of TACI coding variants in common variable immunodeficiency and selective IgA deficiency* (Pan-Hammarström et al. 2007). My contribution to this manuscript included the genotyping of all US CVID and control samples for all of the markers genotyped. The Swedish samples were genotyped by both the Hammarström lab and by myself using two different techniques to confirm that results obtained by the two methods were equivalent. The German samples were genotyped elsewhere using the same method employed by the Hammarström lab. Finally, I performed all statistical analyses for association including Mantel- Haenszel tests, which allowed us to combine allele frequencies from multiple, independent sample groups. Table II-1 and Supplemental Table II-1 were prepared by me.

In the second unpublished report entitled *Association of IgA deficiency and the TACI C104R mutation with human SLE*, I conducted the majority of data acquisition and analyses. Specifically, TACI resequencing was conducted jointly by Michael Weinrich and myself. I performed all additional genotyping and statistical analyses, as well as assaying serum BAFF levels. IgA levels were determined by the Hammarström lab and I analyzed the results. All figures, with the exception of Figure II-5 (prepared by Sarah Hymowitz), were prepared by me.

Reexamining the role of TACI coding variants in common variable immunodeficiency and selective IgA deficiency

Qiang Pan-Hammarström [1,9], Ulrich Salzer [2,9], Likun Du [1], Janne Björkander [3], Charlotte Cunningham-Rundles [4], David L Nelson [5], Chiara Bacchelli [6], H Bobby Gaspar [6], Steven Offer [7], Timothy W Behrens [7], Bodo Grimbacher [8], and Lennart Hammarström [1]

[1]Division of Clinical Immunology, Karolinska University Hospital Huddinge, SE-14186 Stockholm, Sweden.

[2]Division of Rheumatology and Clinical Immunology, Medical Center, University Hospital, Hugstetterstrasse 55, D- 79106 Freiburg, Germany.

[3]Immunodeficiency Unit, Department of Respiratory Medicine and Allergy, Sahlgrenska University Hospital, SE-41345 Gothenburg, Sweden.

[4]Division of Clinical Immunology, Mount Sinai Medical Center, New York, New York 10029, USA.

[5]Center for Cancer Research, National Cancer Institute, US National Institutes of Health, Bethesda, Maryland 20892, USA.

[6]Molecular Immunology Unit, Institute of Child Health, London WC1N 1EH, UK.

[7]Center for Immunology, University of Minnesota Medical School, Minneapolis, Minnesota 55455, USA.

[8]Department of Immunology, Royal Free Hospital, University College London, London 3NW 2QG, UK.

[9]These authors contributed equally to this work

e-mail: lennart.hammarstrom@ki.se or qiang.pan-hammarstrom@ki.se

Note: Supplementary information is available on the Nature Genetics website.

To the Editor:

Tumor necrosis factor (TNF)-like receptors are members of a superfamily of proteins involved in regulating maturation and survival of lymphocytes. One of these receptors, TACI (transmembrane activator and CAML interactor; encoded by *TNFRSF13B*), binds two ligands, BAFF and APRIL. Deletion of *Tnfrsf13b* in mice results in an impaired response to thymus-independent antigens (von Bulow et al. 2001) and virtually abolishes APRIL-induced switching to IgA, IgE and IgG1 (Castigli et al. 2005). Conversely, lack of APRIL, owing to a targeted inactivation of *Tnfsf13* in mice, results in an impaired ability to switch to IgA production (Castigli et al. 2004).

Recently, two studies have reported that sequence variants in *TNFRSF13B* are associated with primary immunodeficiency diseases in humans (Castigli et al. 2005; Salzer et al. 2005). Specifically, common variable immunodeficiency (CVID) was found to be associated with homozygosity for several coding variants (S144X, C104R and A181E) (Salzer et al. 2005); in addition, heterozygous coding variants (C104R, A181E, S194X, R202H and ins204A) were identified in several individuals with CVID (Castigli et al. 2005; Salzer et al. 2005). Furthermore, Castigli et al. (Castigli et al. 2005) showed a strict correlation between the presence of heterozygous coding variants (C104R, A181E and R202H) and CVID and selective IgA deficiency (IgAD) in members of their multicase families, suggesting a causal relationship. In our study (Salzer et al. 2005), one of the siblings of a CVID proband, who is heterozygous for the A181E variant, suffered from IgAD. However, her mother, who is also heterozygous for the A181E variant, had normal immunoglobulin levels, suggesting incomplete penetrance (Salzer et al. 2005).

Here we report that a significant proportion of the normal population carries heterozygous coding variants in *TNFRSF13B* (**Table II-1**), necessitating a re-evaluation of the potential role of these variants in CVID/IgAD. Therefore, we analyzed *TNFRSF13B* in 115 Swedish, 154

German and 155 US individuals with CVID (**Supplementary Table II-2**). We found a highly significant increase in the frequency of the C104R and A181E variants (**Table II-1**) and a significant increase in the frequency of ins204A (**Supplementary Table II-2**), suggesting that these variants, even in a heterozygous form, constitute risk factors for the development of CVID. However, other variants, including R72H, R122W, R202H, V220A and P251L, occurred at similar frequencies in affected individuals and controls (**Supplementary Table II-2**), suggesting that these latter variants do not contribute to the risk of CVID.

Subsequently, we screened 254 Swedish adults with sporadic IgAD or probands in IgAD multicase families for genetic alterations in *TNFRSF13B*. In 55 randomly selected cases, we sequenced the entire gene but did not observe any additional mutations besides those present in controls and individuals with CVID (**Supplementary Table II-2**). Using a SNP-based assay (Hannelius et al. 2005; Salzer et al. 2005) (for primer sequences, see **Supplementary Table II-3** online), we did not find an overrepresentation of C104R, A181E or ins204A in our IgAD case series (**Table II-1** and **Supplementary Table II-2**), nor did we identify any individuals who were homozygous or compound heterozygous for any of the coding variants. However, we found two individuals with IgAD who were heterozygous for the R202H variant, whereas we did not find any R202H heterozygotes among the Swedish controls (**Table II-1**). Subsequently, we investigated the family members of the ten probands with IgAD who carried the C104R, A181E or R202H variants, five of whom belonged to multicase families. However, IgA deficiency did not cosegregate with the TACI coding variants (see families 1, 5, 7 and 10 in **Supplementary Table II-4** and **Supplementary Fig. II-1**), and the transmission disequilibrium test (calculated using TRANSMIT) showed no linkage or association.

Thus, our sequence analysis of *TNFRSF13B* in a large number of cases and controls provides supporting evidence that heterozygous C104R, A181E and ins204A sequence variants in *TNFRSF13B* constitute risk factors for the development of CVID. However, our data suggest

that these variants have only minor roles, if any, in the development of selective IgAD. We are currently examining the functional effects of the heterozygous TACI variants to better understand the mechanism by which these variants contribute to the development of CVID.

ACKNOWLEDGMENTS

We are indebted to C. Lindgren (core facility at the Kliniskt forskningscentrum, Karolinska Institute); J. Birmelin (University of Freiburg); S. Buckridge and A. Thrasher (Institute of Child Health) for the mutation analyses; R. Engqvist (Division of Clinical Immunology, Karolinska University Hospital Huddinge) for collection of family samples and A. Schaffer (National Center for Biotechnology Information, US National Institutes of Health) for statistical advice. This work was supported by ALF funding from the Stockholm County Council, the Swedish Research Council, the Deutsche Forschungsgemeinschaft (DFG) grant SFB620/C2, USIDnet grant NO1-A1-30070, the Primary Immunodeficiency Association, the Medical Research Council UK and European Union grant SP23-CT-2005-006411.

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

TABLES

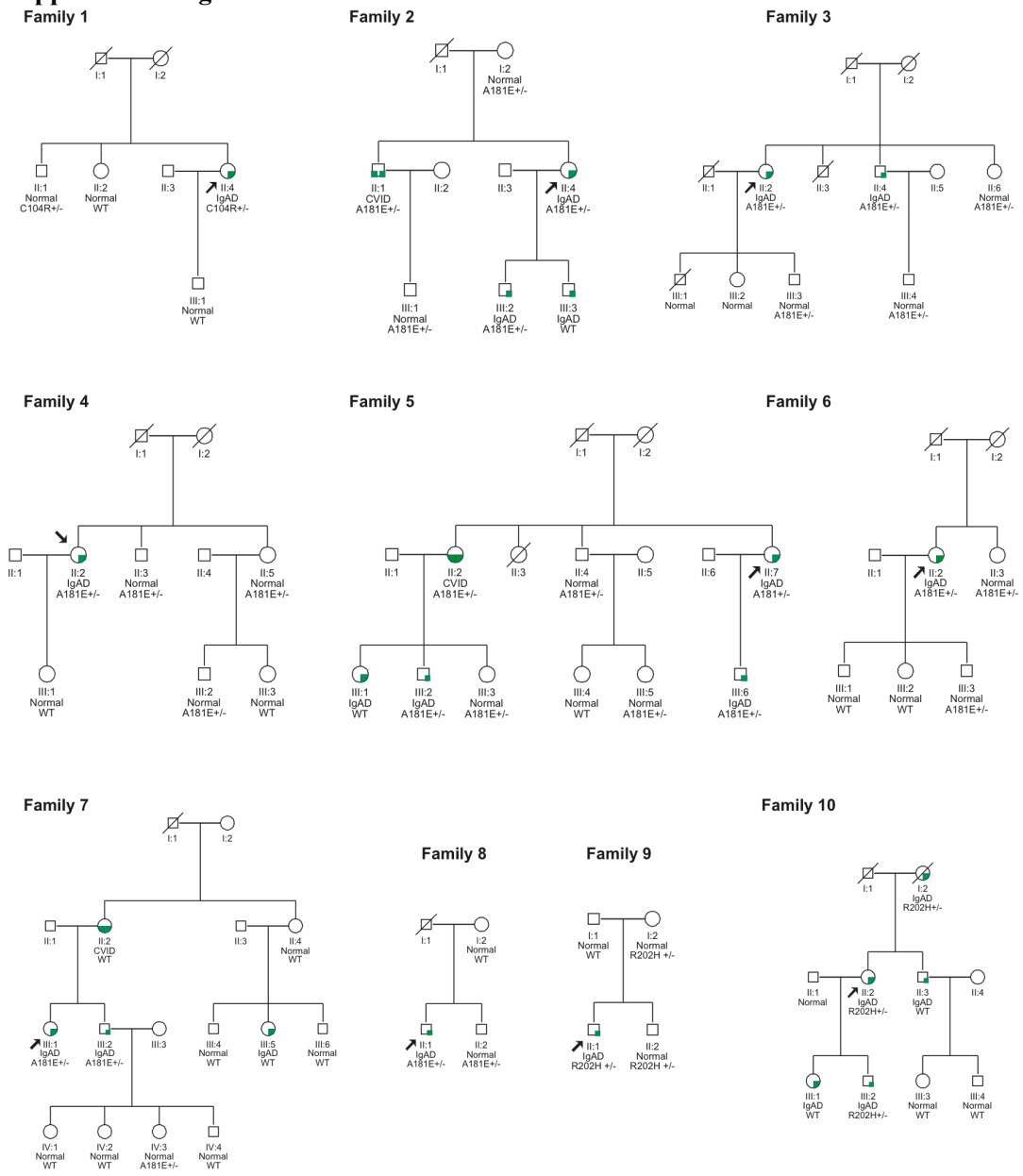
Table II-1. Frequency of selected TACI coding variants in individuals with CVID and IgAD[a]

	C104R		A181E		R202H
CVID (all)[b]	16/846	14/844d	22/844	16/844[d]	1/852
Controls (all)[c]	17/4,208	17/4,208	20/3,924[e]	20/3,924[e]	6/4,624
P value	8.2×10^{-5}	3.8×10^{-4}	3.9×10^{-8}	2.4×10^{-5}	0.69
OR	4.16	3.69	5.60	4.08	0.57
	(1.98–8.74)	(1.73–7.89)	(2.99–10.51)	(2.07–8.04)	(0.06–5.40)
IgAD (Swedish)	1/478		7/464		2/480
Controls (Swedish)	8/2,038		12/1,730		0/2,082
P value	0.55		0.09		3.0×10^{-3}
OR	0.53		2.19		–
	(0.07–4.12)		(0.88–5.47)		–

a) (Number of variant alleles)/(number of tested alleles).
b) The 'CVID (all)' group includes 115 Swedish, 154 German and 155 US individuals with CVID.
c) The 'Controls (all)' group includes 1,080 healthy Swedish blood donors and consecutive samples from the Swedish national neonatal screening program for phenylketonuria (PKU)6, 342 healthy German blood donors and 787 healthy US donors from the New York cancer project collection (Mitchell et al. 2004).
d) Homozygous sequence variants were excluded from the calculation.
e) One 'control' heterozygous for the A181E allele was found to be hypogammaglobulinemic (serum level of IgG = 3.7 g/l).
The institutional review boards at the Karolinska Institute approved this study, and informed consent was obtained from all patients.

SUPPLEMENTAL DATA

Supplemental Figure II-1.



Supplementary Fig. 1 Mutations in *TNFRSF13B* in families with IgAD.

IgAD and CVID in families where the proband was shown to carry a heterozygous TACI mutation (C104R, family 1, A181E, families 2-8 and R202H, families 9-10).

The proband is marked with an arrow. Individuals with IgAD are denoted by a filled quadrant and patients with CVID with a half filled symbol.

Supplementary Table II-2. Mutations/sequence variants in *TNFRSF13B* in CVID and IgAD patients [a]

	Origin	Ins204A	R72H	C104R	R122W	A181E	R202H	V220A	P251L
CVID	Sweden	0/230	1/230	1/230	1/228	10/226	0/230	8/228	25/230
Controls	Sweden	0/716	0/1070	8/2038	4/1784	12/1730	0/2082	27/1070	186/1824
P value		-	**0.031**	0.92	0.54	**5.7×10^{-7}**	-	0.40	0.75
OR		-	-	1.11 (0.14-8.89)	1.96 (0.23-16.91)	6.63 (3.16-13.91)	-	1.40 (0.63-3.12)	1.07 (0.69-1.67)
CVID	Germany	1/308	1/308	9/308	0/308	5/308	1/308	3/308	29/308
Controls	Germany	0/636	4/636	4/660	1/636	4/684 [b]	1/636	2/98	22/98
P value		0.15	0.55	**3.5×10^{-3}**	0.49	0.11	0.60	0.40	0.00070
OR		-	0.51 (0.06-4.45)	4.94 (1.69-14.44)	-	2.81 (0.79-9.95)	2.07 (0.14-31.26)	0.47 (0.08-2.75)	0.36 (0.20-0.65)
CVID	USA	1/314	2/306	6/308	NA	7/310	0/314	6/310	29/310
Controls	USA	0/1512	7/1504	5/1510	NA	4/1510	5/1528	33/1514	150/1464
P value		**0.028**	0.67	**8.5×10^{-4}**		**3.7×10^{-5}**	0.31	0.79	0.64
OR		-	1.41 (0.29-6.75)	5.98 (2.09-17.11)		8.70 (3.11-24.32)	-	0.89 (0.37-2.13)	0.90 (0.60-1.37)
CVID all		2/852	4/844	16/846	1/536	22/844	1/852	17/846	83/848
Controls all		0/2864	11/3210	17/4208	5/2420	20/3924	6/4246	62/2682	358/3386
P value [c]		**9.5×10^{-3}** [e]	0.34	**8.2×10^{-5}**	0.45	**3.9×10^{-8}**	0.69	0.45	0.92
Pooled OR [d]		-	1.27 (0.41-3.97)	4.16 (1.98-8.74)	1.14 (0.14-9.35)	5.60 (2.99-10.51)	0.57 (0.06-5.40)	1.04 (0.59-1.83)	0.82 (0.62-1.08)
IgAD	Sweden	0/474	1/476	1/478	1/480	7/464	2/480	17/480	56/480
Controls	Sweden	0/716	0/1070	8/2038	4/1784	12/1730	0/2082	27/1070	186/1824
P value		-	0.13	0.55	0.95	0.09	**3.0×10^{-3}**	0.26	0.35
OR		-	-	0.53 (0.07-4.12)	0.93 (0.10-8.33)	2.19 (0.88-5.47)	-	1.42 (0.77-2.62)	1.16 (0.85-1.60)
IgAD	Sweden	0/474	1/476	1/478	1/480	7/464	2/480	17/480	56/480
Controls all		0/2864	11/3210	17/4208	5/2420	20/3924	6/4246	62/2682	358/3386
P value [c]		-	0.64 [e]	0.72	0.53	0.05 [e]	**1.8×10^{-3}**	0.13	0.18
Pooled OR [d]		-	0.61 [e] (0.08-4.66)	0.53 (0.07-4.26)	0.93 (0.10-8.33)	2.19 [e] (0.86-5.60)	5.91 (1.94-18.07)	1.42 (0.77-2.63)	1.16 (0.85-1.60)

a) Number of mutated alleles/number of tested alleles
b) One control carrying the A181E allele was found to be hypogammaglobulinema (serum level of IgG = 3.7 g/L)
c) Normal distribution of pooled Mantel-Haenszel OR.
d) Mantel-Haenszel test of pooled odds ratios and 95% confidence intervals.
e) P-value calculated on pooled case-control data using one-tailed chi-squared test.

Supplementary Table II-3. Primer sequences for detection of *TNFRSF13B* mutation/sequence variants using a SNP based mutation assay.

Mutations/sequence variants	Primer 1	Primer 2	Extension primer
ins204A	ACGTTGGATGAGTTCCTGCTAGTCTGTGTC	ACGTTGGATGAGAACTTGCCTTGCTCCTTG	TCTCTCTCCCAGGGTCA
R202H	ACGTTGGATGAGTTTCATGCACTCACCCTG	ACGTTGGATGCTGCTTCCTCAAGAAGAGGG	ACTTGGCCGGACTTTGA
C104R	ACGTTGGATGTCTGTGGACAGCACCCTAAG	ACGTTGGATGTCTGGTGGAAGGTTCACTGG	AAGCAATGTGCATACTTC
A181E	ACGTTGGATGATCAGGTGGCCCTGGTCTAC	ACGTTGGATGCCCTCTTCTTGAGGAAGCAG	CTCTGCTGCTTCCTGGTGG
R72H	ACGTTGGATGGGTCATAGAACTTGCCTTGC	ACGTTGGATGAGTTCCTGCTAGTCTGTGTC	CTTGCCTTGCTCCTTG
R122W	ACGTTGGATGAACCTTCCACCAGAGCTCAG	ACGTTGGATGTCCAATCCTTGGTACCTTCC	AGAGCTCAGGAGACAG
P251L	ACGTTGGATGCCCACCTTCCAGCACAAGT	ACGTTGGATGAGCTTCTGCTTCCCTGAGTG	AGCACAAGTGGGGTCG
V220A	ACGTTGGATGTGCAGATCACGCGATGGAAG	ACGTTGGATGAAGCAGAAGCTGCAGGTCTC	GAAGCCGGCAGCCCTG

Supplementary Table II-4. Lack of correlation between heterozygous *TNFRSF13B* mutations/sequence variants and IgA deficiency

Family	Mutation: Phenotype:	TACI IgAD	TACI normal [a]	wt IgAD	wt normal [a]
1	C104R	1	1	0	2
2	A181E	3 [b]	2	1	0
3	A181E	2	3	0	0
4	A181E	1	3	0	2
5	A181E	4 [c]	3	1	1
6	A181E	1	2	0	2
7	A181E	2	1	2 [d]	6
8	A181E	1	1	0	1
9	R202H	1	2	0	1
10	R202H	3	0	2	2
Sum		19 (17) [e] 18		6 (5) [e]	17

a) Normal serum levels for age of IgM, IgG and IgA
b) The brother of the proband has CVID
c) The sister of the proband has CVID
d) The mother of the proband has CVID
e) Excluding patients with CVID

Association of IgA deficiency and the TACI C104R mutation with human SLE

Steven Offer[1], Mike Weinrich[1], Qiang Pan-Hammarström[2], Sarah Hymowitz[3],

Lennart Hammarström[2], and Timothy W. Behrens[1,3†]

[1] Center for Immunology, University of Minnesota Medical School, Minneapolis, Minnesota 55455, USA.

[2] Division of Clinical Immunology, Department of Laboratory Medicine, Karolinska University Hospital, Huddinge, SE-14186 Stockholm, Sweden.

[3] Genentech, 1 DNA Way, South San Francisco, CA 94080.

† To whom correspondence should be addressed: behrens.timothy@gene.com

INTRODUCTION

Signaling through tumor necrosis factor (TNF)-receptor family members modulates B cell survival, maturation, and isotype switching (Cancro 2004; Miller et al. 2006). The TNF-like ligand BAFF (B cell activating factor, *TNFSF13B*) is known to bind to three TNF receptors on B cells - transmembrane activator and CAML interactor (TACI, encoded by *TNFRSF13B*), B cell-activating factor receptor (BAFFR, encoded by *TNFRSF13C*), and B cell maturation antigen (BCMA, encoded by *TNFRSF17*) (Gross et al. 2000; Marsters et al. 2000; Schiemann et al. 2001; Yan et al. 2001; Yan et al. 2000). A second TNF-like ligand, APRIL (encoded by *TNFSF13*) is able to bind both TACI and BCMA, but not BAFF-R (Hahne et al. 1998; Marsters et al. 2000). APRIL is the primary ligand for TACI and BCMA, with binding affinities for these receptors >100-fold higher than those of BAFF (Cao et al. 2005; Day et al. 2005; Hymowitz et al. 2005). APRIL has also been shown to bind cell-surface heparin proteoglycans, an interaction that may promote ligand oligomerization (Ingold et al. 2005).

BAFF signaling plays an important role in promoting B cell survival and homeostasis (Gorelik et al. 2003). Mice deficient for either BAFF (Gross et al. 2001; Schiemann et al. 2001) or BAFFR (Thompson et al. 2001; Yan et al. 2001) show a reduced number of circulating mature B cells. Mice lacking BCMA (Schiemann et al. 2001; Xu and Lam 2001), TACI (von Bulow et al. 2001; Yan et al. 2001), or both BCMA and TACI (Mackay et al. 2003) have normal levels of circulating mature B cells. BAFF transgenic mice show a large expansion of the mature B cell pool and high serum levels of all isotypes of immunoglobulin (Ig). In addition, these animals produce elevated titers of autoantibodies to anti-double strand DNA antibodies, a hallmark of the autoimmune disorder systemic lupus erythematosus (SLE) (Gross et al. 2000; Khare et al. 2000).

Under normal conditions, self-reactive B cells are tolerized by one of several mechanisms: central deletion in the bone marrow, receptor editing at the pre-B cell stage, or by

the induction of anergy. Elevated BAFF levels were found to allow self-reactive anergic B cells to survive chronic B cell antigen receptor (BCR)-initiated signaling that would normally induce apoptosis (Craxton et al. 2005; Lesley et al. 2004). Restricting the available pool of BAFF in mice by the introduction of decoy receptors Ig-TACI or Ig-BCMA led to a loss of circulating mature B cells, and a delay in the onset of autoimmunity (Gross et al. 2001; Gross et al. 2000; Schneider et al. 2001; Thompson et al. 2000).

Ig isotype class switch in response to T cell-dependent and independent antigen presentation is mediated in part by signaling through TNF-receptors. Synergistic signals produced in B cells by cytokine stimulation and activation of TNF-receptors induces to Ig constant region germline transcription, induction of activation induced cytosine deaminase (AID), and initiation of class switch recombination (CSR) through NFκB pathway signaling (Cancro 2004; Miller et al. 2006). Co-stimulation of the TNF-receptor CD40 on the B cell surface by CD40LG (expressed on the surface of T cells) is required for proliferation and CSR during T cell dependent B cell activation (Lane et al. 1992; Noelle et al. 1992; Valle et al. 1989). Mutations in CD40 or CD40LG confer the primary immunodeficiencies hyper-IgM syndrome type 1 (hyper-IgM1) (Aruffo et al. 1993; Mensink et al. 1987; Padayachee et al. 1992; Padayachee et al. 1993; Pilia et al. 1994) and hyper-IgM3, respectively (Ferrari et al. 2001; Kutukculer et al. 2003). Mutations in the NFκB pathway protein IKKγ, also known as NEMO, cause an X-linked form of hyper-IgM with ectodermal dysplasia (hyper-IgMED) (Jain et al. 2001; Martinez-Pomar et al. 2005; Orstavik et al. 2006; Zonana et al. 2000).

Signaling through TACI, and to a lesser extent BAFF-R, can stimulate T-independent class switch to IgG and IgA (Castigli et al. 2004; Castigli et al. 2005; Litinskiy et al. 2002). Mice deficient for either TACI or APRIL exhibit low serum IgA levels and decreased antibody responses to T-independent antigens (Castigli et al. 2004; Yan et al. 2001). Mice deficient for BAFF or BAFF-R exhibit decreased serum IgA levels and impaired T-dependent response

(Schiemann et al. 2001). BAFF-R$^{-/-}$ mice are, however, able to mount an effective T-independent B cell response while BAFF$^{-/-}$ mice are not (Shulga-Morskaya et al. 2004). No role has been elucidated for APRIL-BCMA signaling in B cell activation or function, however this interaction is thought to promote plasma cell survival in the bone marrow (O'Connor et al. 2004).

Two recent reports have identified polymorphisms in the gene encoding TACI that are associated with the primary immunodefiencies IgA deficiency (IgAD) and common variable immunodeficiency (CVID) in humans. In one report, B cells isolated from patients carrying homozygous TACI mutations were unable to class switch to IgG production following stimulation with APRIL or BAFF ligand (Salzer et al. 2005). In a second report, B cells from patients carrying heterozygous TACI mutations were unable to class switch to IgA or IgG following APRIL stimulation, however BAFF stimulation did result in the production of IgG (Castigli et al. 2005). Patients carrying mutations in TACI had an increased prevalence of pathogenic autoantibodies and autoimmunity (Salzer et al. 2005).

We sought to characterize the effect of TACI mutations on the human autoimmune disorder systemic lupus erythematosus (SLE). To determine the contribution of TACI alleles toward disease, we performed targeted resequencing and genetic association studies of the TACI locus in SLE, CVID, and IgAD patients. Here we identify a new TACI polymorphism, I87N, that is present in SLE and IgAD cases, but not control DNA samples. Furthermore, we provide evidence associating the TACI-C104R mutation with elevated BAFF levels in SLE patients. Finally, we show that SLE patients in general have an increased prevalence of IgAD compared to population controls. These findings are consistent with a model whereby perturbations of the machinery participating in CSR induction could additionally contribute to systemic autoimmunity.

RESULTS

Identification of the TACI I87N missense polymorphism

Exons 1-5 of TNFRSF13B were sequenced in 69 unrelated U.S. CVID cases collected at the NCI/NIH (n=59) and the University of Minnesota (n=10). TNFRSF13B was also sequenced in four IgAD patients from two families collected by the Multiple Autoimmune Disease Genetics Consortium (MADGC), which has successfully identified a large number of families with two or more individuals with different autoimmune disorders (Criswell et al. 2005).

Three patients were found to be heterozygous for a previously unreported missense mutation that results in the replacement of isoleucine with asparagine at amino acid position 87 (I87N) of TACI. One carrier was a CVID patient while the other two were a father-son pair of IgAD patients (3231-211 and 3231-300, **Figure II-2**). Further genotyping of this family identified patient 3231-130 as a heterozygous carrier of the I87N SNP. The results of the sequencing and genotyping in the remaining CVID subjects is reported elsewhere (Pan-Hammarström et al. 2007).

Patient 3231-211 was male, and had a history of idiopathic thrombocytopenic purpura (ITP), frequent infections, and IgAD. His son, subject 3231-300, also had IgAD, frequent serious infections, and was diagnosed with juvenile rheumatoid arthritis at age 8. The grandfather, subject 3231-130, was diagnosed with Grave's hyperthyroidism and had a positive high-titer ANA, but exhibited no Ig deficiencies. The apparent linkage of the I87N SNP with autoimmune phenotypes in this family prompted us to screen SLE patients for TACI polymorphisms.

Genetic association of SLE with known TACI polymorphisms

A total of 487 SLE cases of collected at the University of Minnesota, 156 SLE cases from John Hopkins University School of Medicine and 764 control patients (all of European-Caucasian ancestry, (Mitchell et al. 2004)) were genotyped for the I87N SNP using a specific Taqman SNP genotyping assay purchased from Applied Biosystems. The I87N mutation was not detected in the control population, however one SLE patient was found to be heterozygous. Thus, I87N is a rare TACI polymorphism in SLE.

To evaluate other TACI SNPs in SLE, the previously reported CVID/IgAD associated SNPs were also genotyped in the SLE patient population. Both C104R and P251L were significantly associated with SLE (**Table II-5**). Enrichment of C104R was observed in two independent SLE cohorts and significance strengthened with the addition of other previously reported control samples (Pan-Hammarström et al. 2007). Both R202H and V220A trended toward being protective alleles. In the US samples, R72H was detected uniquely in SLE patients and not controls, however a number of healthy German controls have previously been reported to carry this mutation (Pan-Hammarström et al. 2007). A181E occurred at similar levels in SLE patients and controls, and the 204insA frameshift mutation was not detected in any cohorts tested.

With the exception of P251L, no patients or controls were found to be homozygous for a given SNP. However, four SLE patients were compound heterozygous for TACI mutations. Three patients carried two mutations that were singly associated with disease (two V220A/P251L and one R72H/P251L patient). The fourth patient was compound heterozygous for the mutations A181E and P251L. A single control also carried two TACI mutations, A181E and V220A.

Association of the C104R TACI polymorphism with severe SLE

We next asked whether the identified TACI polymorphisms were associated with the number of diagnosed American College of Rheumatology criteria for SLE (Smith and Shmerling 1999) (**Table II-6**). The prevalence of individual criteria by TACI genotype can be found in **Supplemental Table II-8**. On average, SLE patients used in this study were diagnosed as having 6 criteria. Patients carrying the TACI-C104R mutation exhibited an average of 7.5 criteria ($p=0.015$). Ten of the eleven individual disease criteria also showed elevated prevalence in C104R heterozygous patients as compared to the rest of the patient population (**Supplemental Table II-8**). Pedigrees in which TACI-C104R was detected are presented in **Figure II-3.**

Increased BAFF levels in TACI C104R+ SLE patients

We measured BAFF by ELISA (R&D Systems) in the sera of 13 SLE patients carrying the C104R mutation and one SLE patient carrying the I87N mutation and compared these levels to those found in 14 SLE patients matched for age, gender and SLE criteria and who carried no mutations in TACI. Sera from 12 healthy control donors were also assayed. Overall, BAFF levels were higher in SLE cases than in healthy controls (**Figure II-4**). One SLE patient with TACI-C104R was excluded from statistical analysis as an outlier since the BAFF content was greater than twice that of the next highest sample. In remaining SLE patients carrying the TACI-C104R mutation, the median BAFF level was approximately 150 pg/ml higher than that of matched SLE cases.

IgA levels in SLE patients

Next we investigated the potential link between SLE and IgA deficiency. IgA levels were measured in 593 SLE patients randomly selected from the MN and John Hopkins SLE

collections. The overall prevalence of IgAD (<0.07 g/L serum IgA) in the general Caucasian population is estimated at 1/600 (Pan-Hammarström and Hammarström 2008). In the SLE cohort, 22 of 593 patients (1 in 27, 3.7%) exhibited IgA deficiency, 22-times a high as in the general population IgA (**Table II-6**).

Modeling of TACI polymorphisms

To gain additional insight into potential functional consequences, the location of TACI mutations were modeled onto the three-dimensional protein structure. Three mutations, R72H, I87N, and C104R, were localized to CRD2 (residues 58-109) of the extra-cellular domain of TACI (**Figure II-5**). R72H exchanged a positively charged residue with a potentially charged polar residue located in a poorly conserved region of CRD2. Protein alignment yields that this residue differs in the TACI CRD1 domain and in BCMA and BAFFR. This residue does not contact APRIL and is not predicted to contact BAFF, so there is likely no functional consequence of this mutation.

The TACI I87N mutation is localized at a residue conserved with both human BAFFR and BCMA as either an isoleucine or valine. This residue is largely buried during binding to APRIL. Shotgun alanine scanning mutagenesis has identified this residue as functionally important for binding to both BAFF and APRIL (Hymowitz et al. 2005), making it likely that the I87N allele would decrease TACI binding affinity for both BAFF and APRIL.

The C104R mutation ablates one of the TACI CRD2 disulfide domains. This change may also destabilize the D2 module of TACI, which contains residues that are functionally and structurally important resulting in a protein with decreased affinity for both BAFF and APRIL.

The A181E mutation replaces a hydrophobic residue with a polar one in the expected transmembrane helix, possibly leading to destabilization. The R202H, V220A, and P251L

mutations all affect the intracellular domain. None of them would alter affinity for BAFF or APRIL binding.

DISCUSSION

There have been many published case reports connecting immune deficiency to autoimmunity, however this link is not well characterized. An estimated 30% of patients diagnosed with CVID also develop autoimmune symptoms including thrombocytopenia and autoimmune hemolytic anemia. Similar autoimmune associations have been noted to various extents in other immune deficiencies, notably selective IgAD, X-linked hyper-IgM syndrome, and combined B cell and antibody deficiencies such as Wiskott Aldrich syndrome. IgA patients are also more likely to develop organ-specific autoimmune conditions including IDDM, myasthenia gravis, Crohn's disease, SLE, JRA, and adult rheumatoid arthritis.

It is unclear if the increased incidence of immune deficiency noted in autoimmune patients is indeed due to a common genetic component. Antibody deficiency could be a part of autoimmune disease progression and/or caused by immunosuppressive therapies used to treat autoimmunity. For instance drugs commonly used to treat SLE including cyclophosphamide, corticosterioids and azathioprine have all been shown to produce transient hypogammaglobulinemia in a subset of recipients.

In this report, we have identified a rare highly penetrant autoimmunity allele in the gene encoding TACI (I87N). Additionally we have demonstrated association of both the C104R and P251L alleles with SLE. Patients carrying the C104R allele exhibited more severe SLE and elevated serum BAFF levels. Lastly we have established that IgAD occurs more frequently in SLE patients than controls.

Signaling through TACI by BAFF or APRIL has a negative regulatory effect on B cells survival. Patients carrying polymorphisms in TACI resemble the phenotype of the TACI

knockout mouse. The TACI knockout mouse exhibits an expanded pool of peripheral B cells and develops autoimmunity as it ages. The I87N and C104R mutations likely interfere with the ability of TACI to bind BAFF and APRIL and thus interfere with the ability of TACI to modulate the B cell survival.

De-regulation of BAFF may also play a role in the development of autoimmune symptoms in patients carrying mutant TACI alleles. Mutations affecting BAFF binding sites could reduce the number of potential binding targets and may elevate serum BAFF levels, as was observed for the C104R mutation. High levels of serum BAFF have been observed in SLE patients previously. Our data corroborated this, as BAFF levels were also high in SLE patients that carried no polymorphisms in the gene encoding TACI. There is an obvious similarity in the disease manifestations and BAFF levels between SLE patients and BAFF transgenic mice.

FIGURES

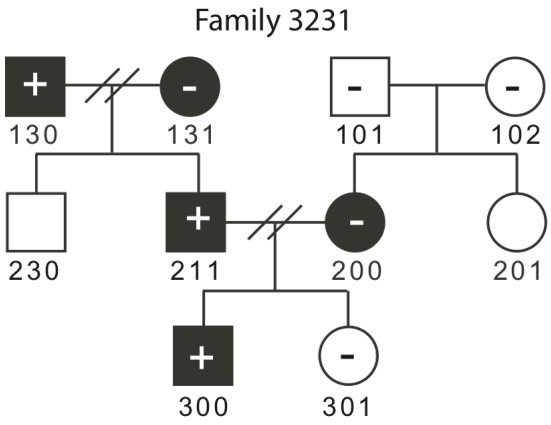


Figure II-2. Transmission of TACI I87N SNP in MADGC family 3231. Squares – males, circles – females, affected individuals – filled symbols. Individuals heterozygous for I87N are denoted with a "+." Individuals with a "-" do not carry the I87N mutation. 3231-230 and 3231-201 were not available for genotyping. Individual identification numbers are located below each symbol.

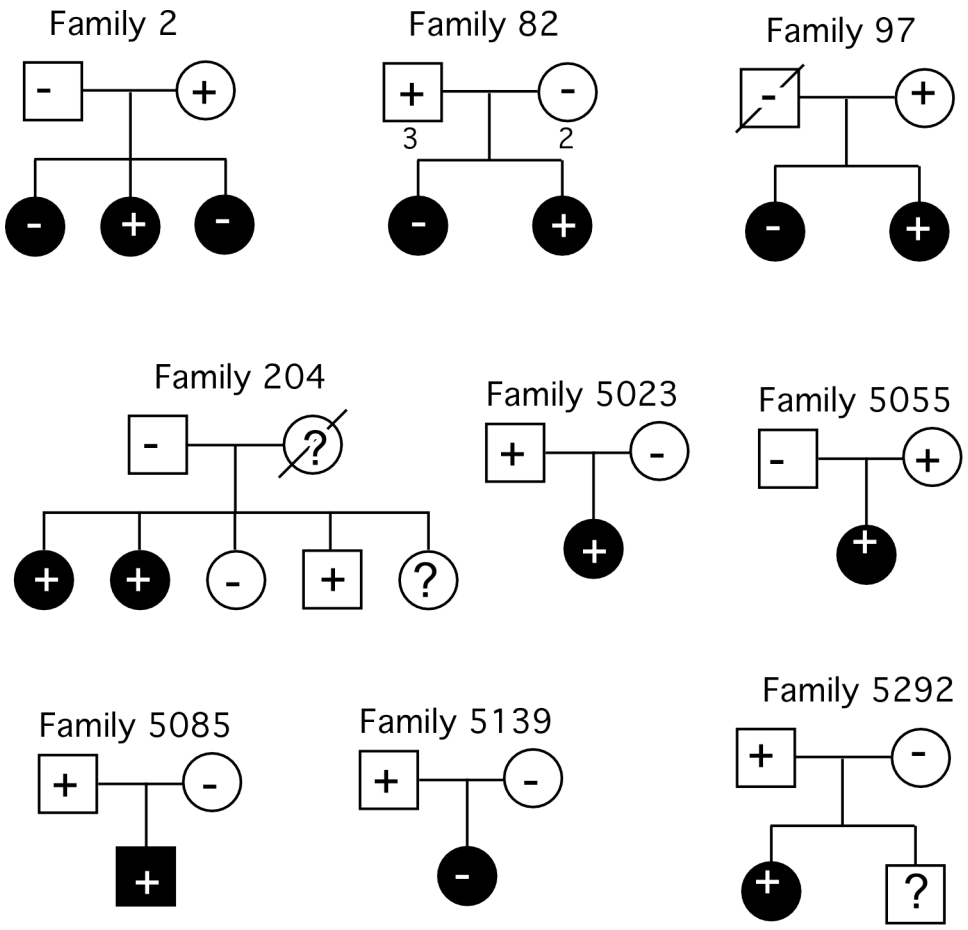


Figure II-3. Transmission of TACI C104R SNP in SLE families. Filled symbols represent patients diagnosed with SLE. Individuals heterozygous for C104R are denoted with a "+." Individuals with a "-" do not carry the C104R variation. A "?" denotes that an individual was unavailable to be genotyped. All other annotations are as in figure 2.

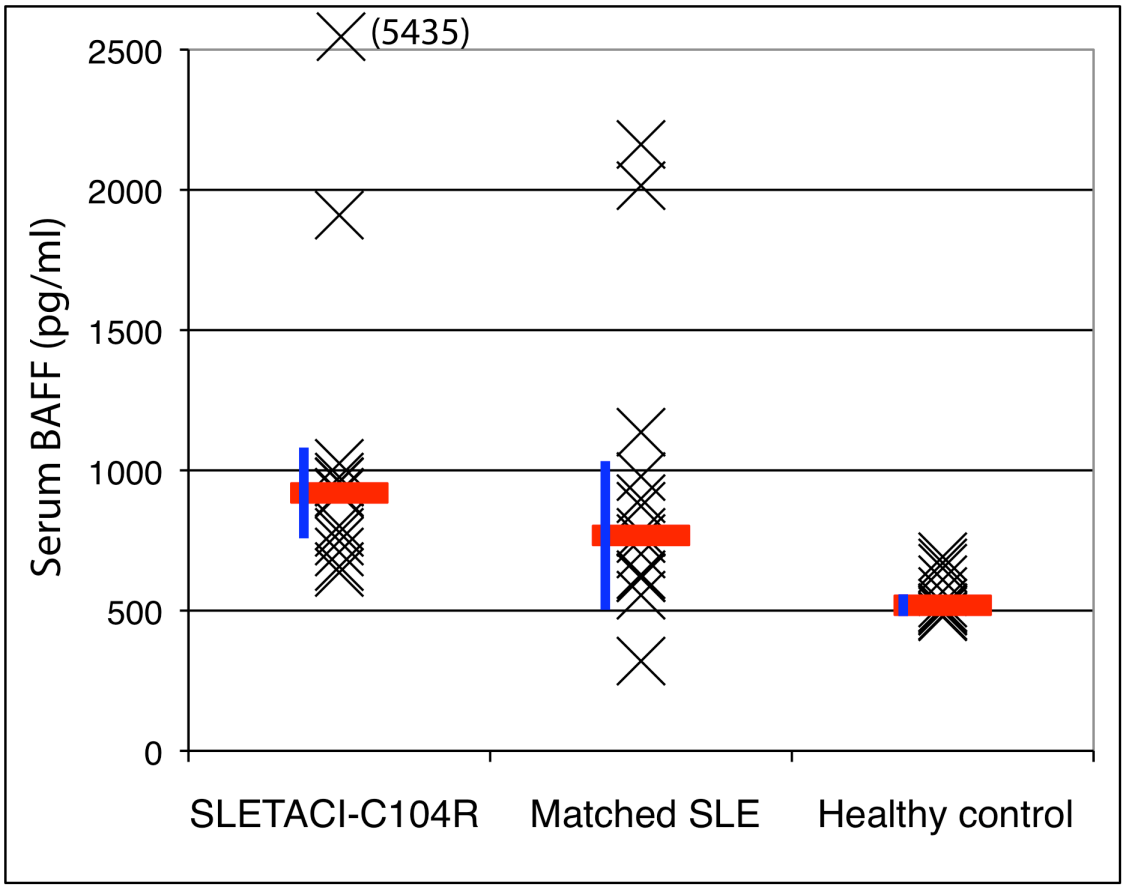


Figure II-4. BAFF levels in C104R subjects. Serum BAFF levels were determined in serum from SLE patients carrying the TACI-C104R mutation, matched SLE patients (based on age, gender, and disease state), and healthy control donors using the Quantikine Human BAFF/BLyS/TNFSF13B Immunoassay (R&D Systems). Median (red bar) and standard deviation (blue bar) are also reported. One outlier sample in the SLE group carrying TACI-C104R was an outlier with BAFF levels of 5435 pg/µl and was excluded from statistical analysis.

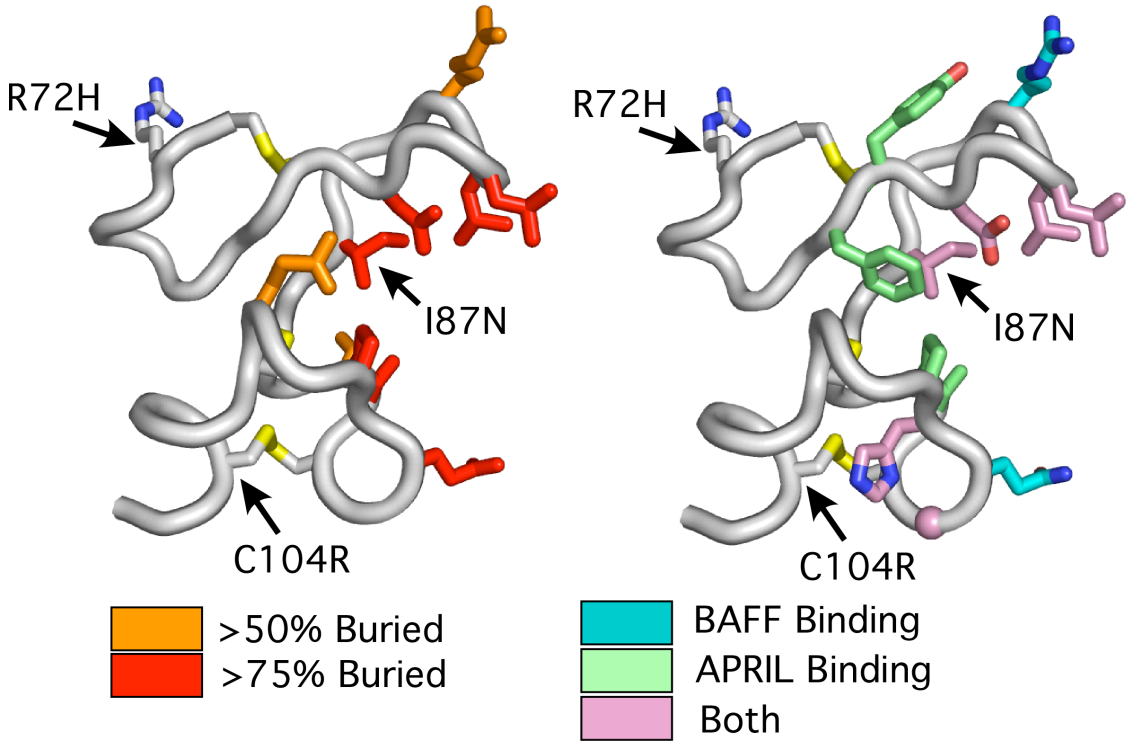


Figure II-5. Three-dimension structure of the TACI CD2 domain (residues 58-109) with location of mutated residues marked. Left panel – amino acid side chains colored based upon surface exposure. Right panel – residues predicted to bind the ligands BAFF and/or APRIL. Figure prepared by Sarah Hymowitz.

TABLES

Table II-5. Mutations/sequence variants in *TNFRSF13B* in SLE patients [a]

	204insA	R72H	I87N	C104R	A181E	R202H	V220A	P251L
US Controls [b]	0/1512	0/1504	0/1504	5/1510	4/1510	5/1528	33/1514	150/1464
JHP SLE	0/276	3/310	0/308	3/296	1/286	0/312	9/312	38/312
P value	-	0.28	-	0.11	0.80	0.31	0.45	0.31
OR	-	2.09 (0.55-7.89)	-	3.08 (0.79-12.06)	1.32 (0.15-11.78)	-	1.33 (0.63-2.81)	1.21 (0.83-1.77)
MN SLE	0/974	6/958	1/976	10/966	4/976	0/976	9/962	135/954
P value	-	0.59	0.21	**0.028**	0.53	0.074	**0.019**	**0.0036**
OR	-	1.35 (0.45-4.01)	-	3.15 (1.13-8.74)	1.55 (0.39-6.14)	-	0.42 (0.21-0.87)	1.44 (1.13-1.85)
All US SLE	1/1250	9/1268	1/1284	13/1262	5/1262	0/1288	18/1274	173/1266
P value	-	0.40	0.28	**0.023**	0.55	**0.040**	0.13	**0.0058**
Pooled OR	-	1.53 (0.57-4.09)	-	3.13 (1.17-8.36)	1.50 (0.40-5.54)	-	0.64 (0.36-1.14)	1.39 (1.10-1.75)
Swedish Controls [c,d]	0/716	0/1070	0/396	8/2038	12/1730	0/2082	27/1070	186/1824
German Controls [d]	0/636	4/636	-	4/660	4/684	1/636	9/98	22/98
All Controls	0/2864	11/3210	0/1900	17/4208	20/3924	6/4246	69/2682	358/3386
P value	-	0.10	0.22	**0.0083**	0.61	0.18	**0.020**	**0.0032**
Pooled OR	-	2.08 (0.88-4.93)	-	2.57 (1.28-5.16)	0.78 (0.29-2.07)	-	0.54 (0.32-0.91)	1.34 (1.10-1.62)

a) Number of polymorphic alleles/number of tested alleles.
b) Controls (n = 787) were obtained from the New York Cancer project collection (Criswell et al. 2005).
c) Healthy Swedish blood donors and consecutive samples from the Swedish national neonatal screening program for phenylketonurea (PKU) (Mitchell et al. 2004).
d) Data from (Pan-Hammarström et al. 2007).

Table II-6. Number of disease criteria per patient by genotype.

Allele	Genotype	Total Criteria [a]	Average [b]	p-value [c]
204insA	0 0	3790 / 624	6.1	
	0 1	0 / 0	-	-
	1 1	0 / 0	-	-
R72H	3 3	3798 / 624	6.1	
	1 3	52 / 9	5.8	0.57
	1 1	0 / 0	-	-
I87N	4 4	3892 / 640	6.1	
	1 4	4 / 1	4.0	-
	1 1	0 / 0	-	-
C104R	4 4	3736 / 617	6.1	
	2 4	97 / 13	7.5	**0.0015**
	2 2	0 / 0		-
A181E	2 2	3798 / 625	6.1	
	1 2	30 / 5	6.0	0.91
	1 1	0 / 0	-	-
R202H	3 3	3912 / 643	6.1	
	1 3	0 / 0	-	-
	1 1	0 / 0	-	-
V220A	4 4	3770 / 618	6.1	
	2 4	96 / 18	5.3	**0.046**
	2 2	0 / 0		-
P251L	2 2	2863 / 471	6.1	
	2 4	903 / 149	6.1	0.81
	4 4	70 / 12	5.8	0.61

a) Total criteria over number of patients with a given genotype
b) Average number of criteria per patient with a given genotype
c) Calculated using two-tailed student's t-test

Table II-7. Prevalence of IgAD in SLE.

	Normal IgA	IgAD	% IgAD
MN-A	461	14	2.9%
JHP	110	8	6.8%
All SLE	571	22	3.7%

Supplemental Table II-8. Proportion of total SLE patients exhibiting a given disease criteria.

Allele	Genotype	MALAR [a]	DISCOID [b]	PHOTO [c]	MULCER [d]	ARTHRITS [e]	SERO [f]	RENAL [g]	NEURO [h]	HEME [i]	IMM [j]	ANA [k]
204insA	0 0	378 / 624	58 / 624	402 / 624	324 / 624	508 / 624	336 / 624	250 / 624	121 / 624	375 / 624	436 / 624	602 / 624
	0 1	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
	1 1	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
R72H	3 3	376 / 624	61 / 624	403 / 624	320 / 624	509 / 624	334 / 624	254 / 624	118 / 624	379 / 624	442 / 624	602 / 624
	1 3	9 / 9	0 / 9	7 / 9	5 / 9	6 / 9	2 / 9	2 / 9	3 / 9	6 / 9	3 / 9	9 / 9
	1 1	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
I87N	4 4	391 / 640	62 / 640	412 / 640	331 / 640	522 / 640	342 / 640	255 / 640	122 / 640	389 / 640	448 / 640	618 / 640
	1 4	0 / 1	0 / 1	1 / 1	0 / 1	0 / 1	0 / 1	1 / 1	0 / 1	0 / 1	1 / 1	1 / 1
	1 1	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
C104R	4 4	373 / 617	58 / 617	396 / 617	314 / 617	501 / 617	330 / 617	247 / 617	115 / 617	372 / 617	434 / 617	596 / 617
	2 4	11 / 13	2 / 13	10 / 13	10 / 13	12 / 13	9 / 13	6 / 13	5 / 13	10 / 13	9 / 13	13 / 13
	2 2	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
A181E	2 2	382 / 625	58 / 625	406 / 625	324 / 625	510 / 625	333 / 625	251 / 625	118 / 625	377 / 625	435 / 625	604 / 625
	1 2	2 / 5	1 / 5	2 / 5	3 / 5	4 / 5	4 / 5	0 / 5	2 / 5	2 / 5	5 / 5	5 / 5
	1 1	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
R202H	3 3	392 / 643	62 / 643	414 / 643	333 / 643	524 / 643	344 / 643	258 / 643	122 / 643	391 / 643	451 / 643	621 / 643
	1 3	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
	1 1	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
V220A	4 4	378 / 618	60 / 618	397 / 618	320 / 618	502 / 618	331 / 618	251 / 618	120 / 618	379 / 618	436 / 618	596 / 618
	2 4	9 / 18	2 / 18	13 / 18	9 / 18	16 / 18	7 / 18	5 / 18	0 / 18	7 / 18	10 / 18	18 / 18
	2 2	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
P251L	2 2	279 / 471	44 / 471	306 / 471	253 / 471	377 / 471	248 / 471	181 / 471	97 / 471	292 / 471	333 / 471	453 / 471
	2 4	96 / 149	16 / 149	92 / 149	68 / 149	125 / 149	81 / 149	64 / 149	23 / 149	87 / 149	105 / 149	146 / 149
	4 4	7 / 12	0 / 12	7 / 12	5 / 12	11 / 12	7 / 12	7 / 12	1 / 12	7 / 12	7 / 12	11 / 12

a) Malar rash
b) Discoid rash
c) Photosensitivity
d) Oral ulcers
e) Arthritis
f) Serositis
g) Renal disorder
h) Neurological disorder
i) Hematologic disorder
j) Immunological disorder
k) Antinuclear antibody positive

CHAPTER III: ASSOCIATION OF MULTIPLE DNA REPAIR PATHWAYS WITH IGAD AND CVID

In this chapter, one report is presented, *Unique DNA repair gene variations associate with the primary antibody deficiencies IgAD and CVID*. At the time of thesis publication, this manuscript was submitted for publication in Human Mutation, pending review.

The majority of the work presented in this manuscript is my own. Rob Graham selected SNPs for the initial genotyping screen and was involved with the resequencing performed at the Broad Institute. With the exception of initial SNP selection, I designed the studies presented in this chapter, analyzed the data, and prepared the manuscript and figures.

Unique DNA repair gene variations associate with the primary antibody deficiencies IgAD and CVID

Steven M. Offer,[1] Qiang Pan-Hammarström,[2] Lennart Hammarström,[2] and Reuben S. Harris[1*]

[1]*Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, Minneapolis, Minnesota 55455, USA;*

[2]*Division of Clinical Immunology, Department of Laboratory Medicine, Karolinska Institutet at Karolinska University Hospital Huddinge, SE-141 86 Stockholm, Sweden*

[*]Correspondence to: Reuben S. Harris; Department of Biochemistry, Molecular Biology, and Biophysics; University of Minnesota, Minneapolis, Minnesota 55455, USA.

Tel: +1 612 624 0457, Fax: +1 612 625 2163.

Email: rsh@umn.edu

KEY WORDS: antibody; class switch recombination; CVID; IgAD; immunodeficiency genes

Introduction

Upon antigen stimulation, the constant region exons of the expressed antibody heavy chain gene can be replaced with downstream exons that encode an alternative antibody isotype (IgM -> IgG, IgE or IgA). At the molecular level, this antibody class switch recombination (CSR) process occurs between C/G-rich switch region DNA sequences located upstream of each set of constant region exons (**Figure III-1**). In recent years, a consensus model has emerged whereby CSR is initiated by cytidine to uridine deamination events within switch region DNA catalyzed by activation-induced deaminase (AID) (Bransteitter et al. 2003; Chaudhuri et al. 2003; Muramatsu et al. 2000; Petersen-Mahrt et al. 2002; Rada et al. 2004). These DNA uridines are subsequently recognized and removed by the base excision repair enzyme uracil DNA glycosylase 2 (UNG2) or the mismatch repair proteins MutSα and MutLα (heterodimers of MSH2/MSH6 and MLH1/PMS2, respectively) (Ehrenstein and Neuberger 1999; Ehrenstein et al. 2001; Li et al. 2004; Martin et al. 2003; Martomo et al. 2004; Petersen-Mahrt et al. 2002; Rada et al. 2002; Schrader et al. 1999; Schrader et al. 2005). Additional base excision repair, mismatch repair, and recombination repair factors are then proposed to help convert these DNA repair intermediates to double-strand breaks and ultimately to CSR products (Chaudhuri and Alt 2004; Honjo et al. 2004; Longerich et al. 2006; Maizels 2005; Neuberger 2008; Stavnezer et al. 2008).

A subset of primary antibody deficiency syndromes has been explained by defects in these DNA repair pathways. For instance, hyper-IgM syndrome type 2 (HIGM2; MIM# 605258) is caused by mutations in *AID* and this disease is characterized by high levels of IgM at the expense of the other antibody isotypes ((Revy et al. 2000); reviewed in (Durandy et al. 2006)). The less severe HIGM5 (MIM# 608106) is caused by mutations in *UNG2* (Imai et al. 2003). Additionally, defects in *PMS2* have been associated with decreased antibody production (Peron et al. 2008). Varying degrees of antibody deficiency have also been noted in

chromosomal instability syndromes such as ataxia-telangiectasia (A-T, *ATM* mutations; MIM# 607585), Nijmegen breakage syndrome (NBS, *NBS1* mutations; MIM# 602667), and ataxia-telangiectasia-like disorder (ATLD, *MRE11* mutations; MIM# 600814) (Carney et al. 1998; Concannon and Gatti 1997; Lahdesmaki et al. 2004; Pan et al. 2002; Savitsky et al. 1995; Stewart et al. 1999; Varon et al. 1998).

Prior studies have shown that missense mutations in *MSH5*, which impair binding of MSH5 with its obligate heterodimerization partner MSH4, associate with immunoglobulin A deficiency (IgAD; MIM# 137100) and common variable immunodeficiency (CVID; MIM# 240500) (Sekine et al. 2007). IgAD and CVID are thought to have common genetic components in at least a subset of patients, as these antibody syndromes often occur in different individuals of the same family (Vorechovsky et al. 1995). Mutations in the B cell surface receptor genes *TACI* (Castigli et al. 2005; Pan-Hammarström et al. 2007; Salzer et al. 2005) and *CD19* (Kanegane et al. 2007; van Zelm et al. 2006), and the T cell receptor gene *ICOS* (Grimbacher et al. 2003) are also responsible for a subset of CVID cases. Nevertheless, despite this considerable progress, the genetic causes of >90% of IgAD and CVID cases have yet to be identified (reviewed in (Pan-Hammarström and Hammarström 2008; Schaffer et al. 2007; Yong et al. 2008)).

Defects in DNA metabolism proteins are responsible for a fraction of primary antibody deficiency syndromes, leading us to hypothesize that variations in DNA repair genes could also underlie additional IgAD and CVID cases. To test this hypothesis, we performed a candidate pathway genetic association study of 27 DNA metabolism genes in IgAD/CVID. Pathways that contained the strongest associated SNPs and multi-marker haplotypes were sequenced in a panel of 96 IgAD/CVID patients. All non-synonymous alleles that were discovered were genotyped in a large case-control association study. Novel coding and non-coding variations were detected in several genes, including seven variants found only in IgAD – MSH2-A727S, MLH1-S247A,

MLH1-G638R, RAD50-P165H, RAD50-R1077Q, NBS1-P401R, and NBS1-D527Y – and two specific to CVID – MLH1-Q409P, and RAD50-Q372X. A few patients were found to carry multiple non-synonymous alleles. One IgAD patient carried MSH2-A727S, MLH1-G638R, and NBS1-P401R. Another IgAD patient was found to be compound heterozygous for RAD50-P165H and a previously reported RAD50-R327H variation.

Materials and Methods

DNA Repair Gene Mendelian Inheritance in Man (MIM) Accession Numbers

AID (605257); *APEX1* (107748); *CD19* (107265); *ERCC1* (126380); *ICOS* (604558); *MLH1* (120436); *MRE11* (600814); *MSH2* (609309); *MSH4* (602105); *MSH5* (603382); *MSH6* (600678); *MUS81* (606591); *NBS1* (600814); *PMS2* (600259); *RAD50* (604040); *RAD52* (600392); *RAD54B* (604289); *TACI* (604907); *UNG2* (191525).

Human DNA Samples

The initial association screen was performed in DNA from 140 IgAD patients, 48 CVID patients, and 92 healthy controls. To characterize all coding variation for the selected candidate genes in the CVID population, genomic resequencing was performed in 58 CVID DNAs, 48 from the original population and 10 that were subsequently recruited. 38 IgAD patients, each carrying at least one allele/haplotype associating with disease, were also included. 334 IgAD patients, 108 CVID patients, and 991 healthy controls were used for the subsequent large genetic association study of detected non-synonymous SNPs. Where adequate DNA was available, samples that had been used for earlier studies were utilized. All DNA samples used in this study were collected from Swedish individuals and isolated at the Karolinska Institutet (Stockholm, Sweden). Informed consent was obtained from all subjects, and these studies were approved by the Karolinska Institutet human subjects research institutional review board.

SNP Selection and Genotyping

For the initial association screen, tag SNPs were selected to cover the genomic locus of each gene plus an additional 10 kb upstream and 5 kb downstream using Tagger (de Bakker et al. 2005). As phase II hapmap data was not yet available at the time, phase I data was used. Additional non-tag SNPs were included to fill out the assay. Genotypes were determined using the Illumina GoldenGate platform at the Broad Institute. Of the 384 markers genotyped, 81 failed to pass quality control tests and were not used in subsequent studies (**Supplementary Table III-4**). Two markers rs2020908 and rs689754 deviated significantly from Hardy-Weinberg equilibrium in the control DNA set ($p<0.001$), and 79 markers had genotype call rates that were below the 90% confidence threshold. Therefore, a total of 303 SNPs passed quality control criteria and were used in the subsequent studies.

For the large genetic association study, genotyping was performed by the BioMedical Genomics Center at the University of Minnesota. Two Sequenom iPLEX gold genotyping assays were designed using Spectrodesigner, part of the MassARRAY software package, to include non-synonymous coding SNPs detected by resequencing. Assays were filled out by including synonymous coding SNPs, UTR SNPs, and intronic SNPs located within 10 bases of an exon. Primers were synthesized by IDT and sequences are available upon request.

DNA Repair Gene Resequencing

Primer design, sequencing, and polymorphism detection were performed by the Broad Institute Center for Genotyping and Analysis. Briefly, primers were designed for each target gene to sequence the known exons, the 5' UTR, and the 3'UTR, with at least 100 nucleotides of flanking sequence on each side. Genetic variants were identified by a combination of the PolyPhred and PolyDhan programs. Primer sequences for each assay are available upon request.

To increase the power to detect informative alleles, the sample set was enriched for DNAs that carried disease-associated single markers and haplotypes as determined in the initial genotyping screen. In all, 3,086,784 nucleotides were examined, with average 1x coverage of 92.0% and 2x coverage of 88.4% (**Supplementary Table III-5**). All novel non-synonymous SNPs were confirmed by fluorescence-based Sanger di-deoxy sequencing at either the University of Minnesota Sequencing and Analysis Facility or the Macrogen Company (South Korea).

Data Processing and Statistical Analyses

For both genotyping studies, allele calls were made using the appropriate built-in automatic allele calling specific to each platform. Call clustering was inspected manually and adjusted when obvious calling errors were made by the software. For genetic association studies, we required that each SNP have a missing genotype rate <10% and, in controls, have a Hardy-Weinberg p-value >0.001. Samples were excluded if genotyping failed for >50% of markers. Haploview was used to determine haplotype structure from control genotypes and to perform association testing for both single markers and haplotypes (Barrett et al. 2005).

To correct for multiple testing bias, association results were subjected to 100,000 permutation tests by pathway using Haploview. For this analysis, genes were grouped based on functional pathways. Each group that contained at least one marker or haplotype with an uncorrected p-value <0.01 was permuted. These pathways were MutSα (MSH2 and MSH6), MutSβ (MSH2 and MSH3), MutLα (MLH1 and PMS2), MRN (MRE11, RAD50, NBS1), MutSγ (MSH4 and MSH5), and RAD51-mediated homologous recombination (RAD51, RAD52, RAD54B, DMC1, and BLM). AID and MUS81 was permuted on an individual gene basis. Permutation testing was conducted for SNPs and haplotypes independently.

Results

IgAD/CVID Association Screen of 27 DNA Repair Genes

To assess the role of DNA repair gene variations in IgAD and CVID, we performed a genetic association screen of 303 SNPs in candidate DNA metabolism pathways. Significant single marker association with IgAD and/or CVID ($p<0.01$) was noted for SNPs in the mismatch repair complexes, MutSα, MutSβ, and MutSγ, the MRN complex, the extended RAD52 epistasis group, and AID. A summary of single marker associations with p-values <0.01 is presented in **Table III-1**. Full association data can be found in **Supplementary Table III-6**.

To correct for multiple testing bias, SNPs were grouped by DNA repair complex or pathway and subjected to permutation testing. After applying this statistical method, 13 SNPs remained significantly associated with IgAD/CVID (**Table III-1**). Two MSH2 SNPs, rs3771276 and rs6729015, were significantly associated with CVID (corrected p=0.045 and 0.0498, respectively). The *RAD50* marker rs2237060 exhibited strong association with CVID (corrected p=0.029) and borderline association in the IgAD/CVID combined cohort (corrected p=0.057). The *RAD52* marker rs10849605 was associated with CVID (corrected p=0.040), and the *RAD54B* SNP rs3019279 showed association in the IgAD/CVID combined cohort (corrected p=0.035). One marker in *AID*, rs2580874, showed significant association with all three groups, with the combined IgAD/CVID cohort most strongly associating (corrected p=0.0055). An additional *AID* SNP, rs714629, was associated with CVID and with the combined cohort (corrected p=0.037 and 0.044, respectively). Finally, markers in the MHC class III region gene *MSH5* were strongly associated with IGAD, in agreement with our prior studies (Sekine et al. 2007).

Next, we constructed multi-marker haplotypes and tested for association with IgAD and/or CVID. The markers comprising each haplotype block are listed in **Supplementary Table III-7**. Association data for all haplotypes with a frequency greater than 0.1% are reported

in **Supplementary Table III-8**, and those associated significantly with IgAD/CVID (uncorrected p-value <0.01) are summarized in **Table III-2**. After pathway-based permutation testing, haplotypes of *MLH1*, *MSH5*, and *MUS81* were significantly associated with IgAD and the combined IgAD/CVID group. Association with CVID was noted for haplotypes of *MSH2* and *RAD50*.

Identification of Novel IgAD/CVID Alleles

Since notable associations were observed for genes encoding the DNA repair proteins that convert AID-catalyzed uracils into DNA breaks, we resequenced select DNA repair genes of IgAD and CVID patient samples. Specifically, we resequenced the MutSα genes *MSH2* and *MSH6*, the MutLα gene *MLH1*, and the MRN complex genes *MRE11*, *RAD50*, and *NBS1*. We also resequenced the genes of other CSR implicated proteins, including those for APEX1 (endonuclease), ERCC1 (endonuclease), and RAD52 (recombinase). We also resequenced *AID*.

DNA repair gene resequencing yielded 242 genetic variants. 93 variations did not have records in dbSNP or SwissProt and were therefore considered novel (**Supplementary Table III-9**). We found 24 alleles that encode amino acid changes, 13 of which were previously unreported. Of the remaining 80 novel SNPs, 5 were synonymous coding alleles, 31 were located in mRNA untranslated regions (UTRs), 40 were intronic, and 4 were in flanking nontranscribed regions.

We next determined the allele frequencies of identified non-synonymous alleles and selected UTR, synonymous, and intronic SNPs in a large cohort of healthy controls, IgAD patients, and CVID patients (**Table III-3**, **Supplementary Table III-10**). In the mismatch repair pathway, two novel MSH2 variations were discovered in IgAD samples: MSH2-T292S and MSH2-A727S. The latter variation was only in a single IgAD patient. Four novel amino acid mutations were detected in MLH1. MLH1-S247A and MLH1-G638R were specific to

IgAD, and MLH1-Q409P were specific to CVID (**Table III-3**). MLH1-H727L and the previously reported mutations MLH1-K618A, MLH1-V716M, and MLH1-I219V were detected at similar levels in IgAD/CVID patients and controls.

The most striking variations were observed in *RAD50*. In all, 16 novel SNPs were found: 3 encoded amino acid changes, 2 were located in the 5' UTR, 1 was synonymous, and 10 were intronic (**Supplementary Table III-9**). Most notably, in a single CVID sample, a heterozygous transition mutation (1114C>T) was found that introduces a premature stop codon at amino acid 372 (RAD50-Q372X, **Table III-3**). Additionally, an IgAD patient was found with the RAD50-R1077Q mutation and another carried both RAD50-P165H and RAD50-R327H. The latter variation was also detected in our control samples, but never linked with RAD50-P165H. Pedigree analysis of the RAD50-P165H/R327H patient revealed that the two alleles were present in a compound heterozygous state.

Novel variations were also detected in the other two MRN complex genes, *MRE11* and *NBS1*. In *MRE11*, 18 novel SNPs were detected (**Supplementary Table III-9**). Of these, two resulted in amino acid changes (MRE11-A492D and MRE11-E494K), eight were in the 3' UTR, and eight were intronic. Both MRE11-A492D and MRE11-E494K were detected at similar frequencies in controls as in IgAD and CVID cases (**Table III-3**). Two 3' UTR SNPs, Broad13928907 and Broad13928926, were, however, significantly enriched in the CVID population. In *NBS1*, 11 unreported SNPs were found (**Supplementary Table III-9**). Three were non-synonymous, and the remaining 8 were a mixture of synonymous, 5' UTR, 3' UTR, and intronic. Two novel alleles, NBS1-P401R and NBS1-D527Y, were specific to IgAD (**Table III-3**). NBS1-V210F and NBS1-L421S were detected in both IgAD/CVID and control samples, as was the previously reported common polymorphism, rs1805794 (NBS1-E185Q).

A number of previously unreported SNPs were also found in other genes. In *AID,* two novel 3' UTR SNPs and three novel synonymous SNPs were detected (**Supplementary Table**

III-9). Two novel SNPS were found in *APEX1*, one synonymous and one 5'UTR SNP that showed enrichment in the CVID group. In *ERCC1*, one 3' flanking SNP and five novel intronic SNPs were detected. 18 new SNPs were discovered in *RAD52*, 14 in the 3' UTR, and four in introns. None of these four genes had variations encoding amino acid substitutions.

Some patients were found to carry mutant alleles in more than one DNA repair gene. An IgAD patient heterozygous for the previously reported APEX1-Q51H allele also carried three novel mutations, MLH1-G638R, MSH2-A727S, and NBS1-P401R. One IgAD patient carried both RAD50-R1077Q and MLH1-K618A. Additionally, as described earlier, an IgAD patient was found to be biallelic for RAD50-P165H and RAD50-R327H. These observations suggest that multiple subtle defects in DNA repair genes may combine to explain some patients' immune deficiencies.

Discussion

We have shown that the primary immunodeficiency syndromes CVID and IgAD associate with SNPs and multi-marker haplotypes in various DNA repair pathways. To our knowledge this is the first report to establish genetic association between immunodeficiency and markers in many of these genes, notably *MSH2, MLH1,* and *RAD50*. We also identified new non-synonymous alleles of *MSH2, MLH1, RAD50*, and *NBS1,* and UTR SNPs in *RAD50* and *MRE11* that were unique to IgAD and/or CVID. Overall, these data suggest that variations in the proteins of the mismatch and recombination repair pathways may underlie some IgAD and CVID cases.

The mismatch repair proteins directly recognize and excise single base DNA mismatches and short patches of mismatched DNA (reviewed in (Kunz et al. 2009)). Genetically, the involvement of mismatch repair in CSR has been well established in mice, with the most prominent defects observed in animals that also lack the uracil excision repair protein

UNG2 (Rada et al. 2004). Humans with heterozygous inactivating mutations in the mismatch repair proteins MSH2, MSH6, MLH1, or PMS2 are at increased risk for developing hereditary non-polyposis colorectal cancer (HNPCC) (Lynch and de la Chapelle 2003; Watson and Lynch 1994). IgA deficiency has been noted in a handful of HNPCC patients, one with a homozygous MSH2 mutation (Whiteside et al. 2002) and three with biallelic MSH6 mutations (Ostergaard et al. 2005; Scott et al. 2007).

Our results are consistent with a central role for mismatch repair in CSR. Most of the IgAD-CVID-specific alleles that we detected in the mismatch repair pathway are located in functionally important domains. MSH2-A727S maps to a conserved ABC domain (amino acids 718-731) that is predicted to be involved in ATP hydrolysis (Fuellen et al. 2005) (**Figure III-2**). Two of the three IgAD/CVID specific mutations identified in *MLH1* map to domains of known importance. MLH1-S247A is located in a region that has both nucleotide hydrolysis and signal transduction functions (Ban et al. 1999). Mutation of this residue to proline (S247P) has been reported in 14 HNPCC2 patients and is associated with decreased MLH1 protein expression and impaired mismatch repair activity using an *in-vitro* assay (Raevaara et al. 2005). Similar functional impacts would be expected for S247A. MLH1-G638R introduces a positively charged amino acid into a neutral area of the region known to be important for interaction with PMS2, MLH1, and PMS1 (Guerrette et al. 1999; Kondo et al. 2001). Interestingly, two additional rare alleles that we detected in both IgAD/CVID cases and control samples were previously identified and characterized in HNPCC studies. MLH1-K618A was suggested to bind less efficiently to PMS2 (Liu et al. 1999), however follow-up studies with this allele found PMS2 binding and overall mismatch repair function to be normal (Raevaara et al. 2005). The MLH1-V716M variant has been suggested to be a protective allele, as it was found enriched in controls compared to HNPCC cases (Barnetson et al. 2008). Taken together, these results

indicate that cancer predisposition alleles that are DNA damage sensitizing may also perturb the ability of the DNA repair pathways to elicit CSR.

Our results are the first to implicate MRN defects in IgAD/CVID. The premature stop codon encoded by the RAD50-Q372X allele likely leads to nonsense-mediated decay of the mRNA, potentially leading to haploinsufficiency (**Figure III-3**). Other identified mutations may slightly perturb the structure and/or function of the MRN complex, and could act as weak hypomorphic alleles. The RAD50-P165H mutation replaces a neutral, non-polar proline with a positively charged, polar histidine at a position that is evolutionarily conserved through vertebrates. This residue is just outside of the N-terminal NTPase-containing globular domain and at the edge of the region that is thought to be important for interaction with MRE11. At the edge of the N-terminal MRE11-interating domain, RAD50-R1077Q replaces a conserved, positively charged arginine with a neutral glutamine. RAD50-R327H, which by itself did not associate with IgAD/CVID, exchanges one neutral, polar residue for another in the N-terminal coiled-coil domain. Neither of the additional IgAD/CVID-specific alleles detected in the MRN complex, NBS1-P401R and NBS1-D527Y, disrupt known or predicted functionally important protein domains. Future studies on the detected MRN alleles will provide greater insight into how they functionally impact the recombinational repair and CSR.

The association noted for two markers in *AID* suggests that a subclass of IgAD/CVID syndromes may be due to variations at this locus. These markers may be linked to undiscovered alleles that impact the expression of functional AID. This could be due to variations that create splice variants, UTR mutations that impact mRNA stability and protein expression, or mutations that impact transcription of *AID*, including alterations to transcription factor binding sites or sites important for epigenetic modulation of expression. In these studies, we identified two novel SNPs in the *AID* UTR and three novel synonymous mutations (**Supplemental Table III-**

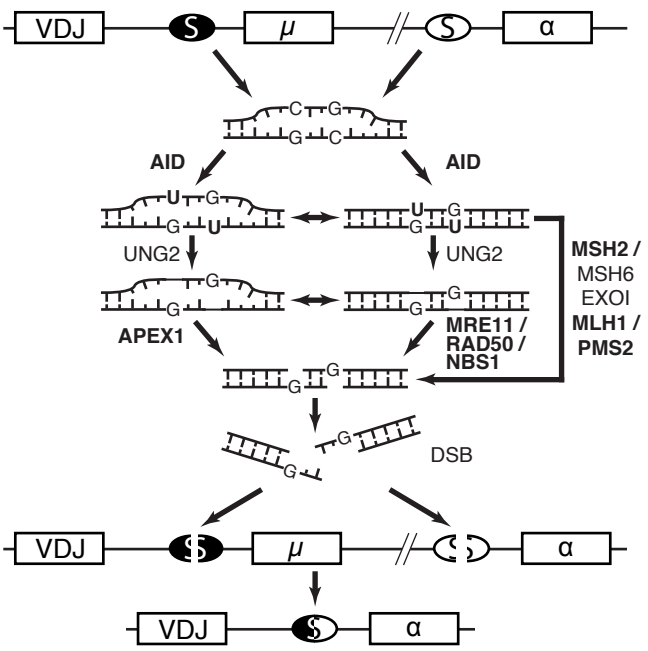
9). The further studies of these alleles may yield that some IgAD/CVID patients represent a new subclass of Hyper-IgM type 2 patients.

Overall, these studies suggest that, like many cancer predisposition syndromes such as HNPCC, IgAD and CVID may be a heterogeneous class of diseases due to defects in multiple pathways. Rather than straightforward causal alleles giving rise to disease, IgAD and CVID alleles likely represent risk factors that predispose an individual to developing primary immune deficiency. Additional studies to determine what effects these unique variations have on DNA repair and CSR are ongoing and long-term.

Acknowledgments

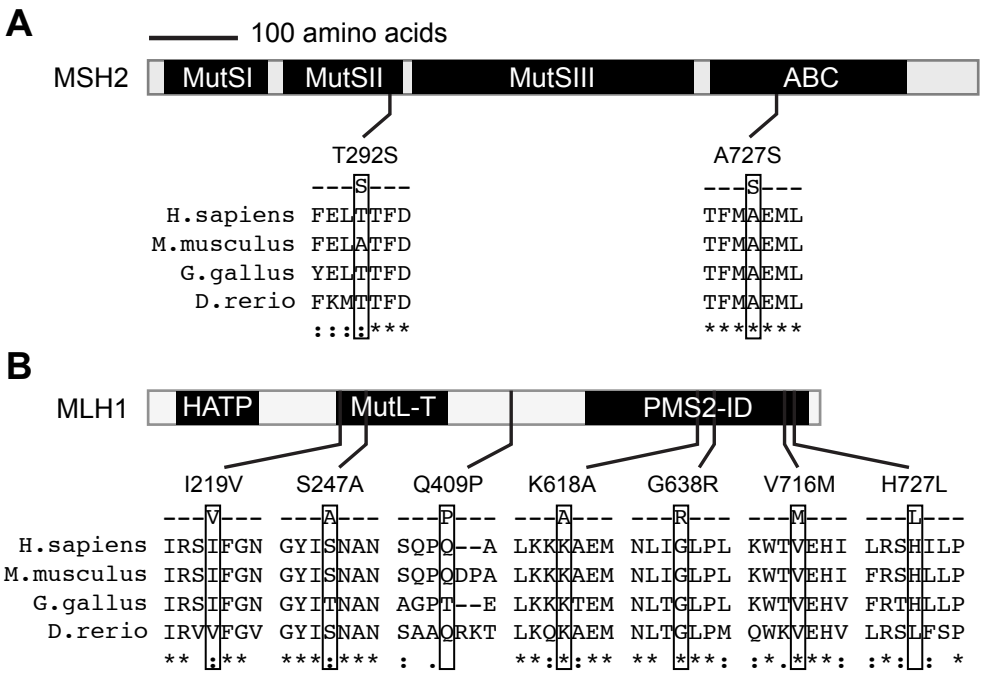
We thank T.W. Behrens, R. Ferreira, and R. Graham for initiating aspects of these studies, A. Kotnis for confirming some of the rare alleles, the Broad Institute for Illumina genotyping and DNA sequencing, the University of Minnesota Biostatistics and Informatics Shared Resource for statistical support, the University of Minnesota Supercomputing Institute for computational support, and the University of Minnesota BioMedical Genomics Center for Sequenom genotyping. This work was supported by the National Institutes of Health (U19 AI067152 to L.H., R21 AI079743 to R.H.) and by Swedish Cancerfound (to Q.P.). S. Offer was also supported in part by a University of Minnesota Graduate School Doctoral Dissertation Fellowship and by the Curtis L. Carlson University of Minnesota/Karolinska Institutet Medical Research and Education Program Award.

Figures



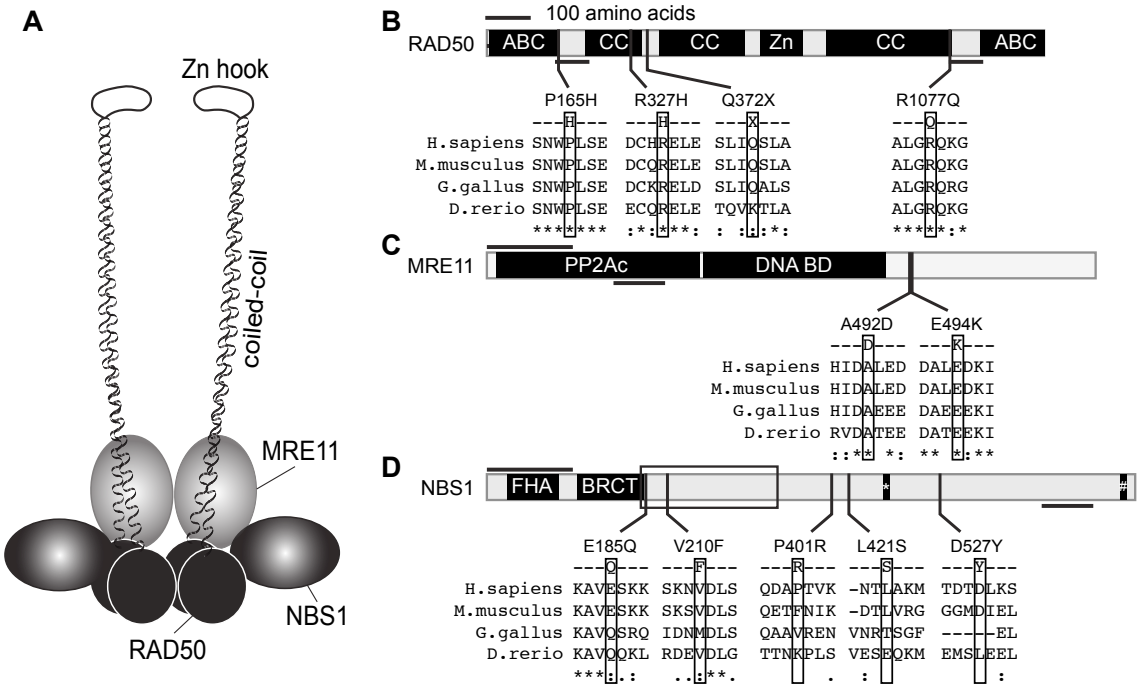
Offer et al., Figure 1.

Figure III-1. A model for DNA repair enzymes in class switch recombination. AID initiates CSR by deaminating switch (S) region DNA cytosines to uracils (the μ and α constant regions are depicted). These uracils are recognized as lesions and excised by the base excision repair enzyme UNG2 or the mismatch repair complex consisting of MutSα (MSH2/MSH6) and MutLα (PMS2/MLH1). UNG2-generated abasic sites are converted to single-strand breaks by the endonuclease activity of APEX1 or the MRN complex (MRE11-RAD50-NBS1). Mismatch repair also leads to single-strand breaks. Opposing single-strand breaks may naturally lead to double-strand breaks, which can be further processed by non-homologous end joining factors to yield a complete CSR event. A switch from IgM to IgA is depicted, and the intervening switch circle is not shown.



Offer et al., Figure 2.

Figure III-2. Location of the amino acid variations in the mismatch repair proteins MSH2 and MLH1. (**A**) MSH2 is divided into four domains. MutSI, II, and III represent a DNA binding globular region, an RNaseH-like motif, and a major structured domain, respectively, revealed by the *Thermus aquaticus* MutS crystal structure (Obmolova et al. 2000). The fourth domain is an ABC-type ATPase domain (Fuellen et al. 2005). The indicated amino acid variations lie within conserved regions. (**B**) MLH1 has three domains, a histidine kinase-like ATPase domain (HATP), a MutL transducer domain (MutL-T), which is homologous to a domain in DNA gyrase B, and a PMS2 interaction domain (PMS2-ID). The indicated amino acid variations lie within conserved regions. The illustrated protein segments were derived from ClustalW reference sequence alignments of the human, mouse, chicken and zebrafish proteins (MSH2, NP_000242, NP_032654, XP_426110, and NP_998689; MLH1, NP_000240, NP_081086, XP_418828, and NP_956953).



Offer et al., Figure 3.

Figure III-3. Location of the amino acid variations in the recombination proteins MRE11, RAD50, and NBS1. (**A**) A schematic of the MRN complex, indicating two copies of each protein and the extended coil-coil and zinc hook domains of RAD50. (**B, C, & D**) Schematics depicting the major protein domains of RAD50, MRE11, and NBS1, with the IgAD/CVID amino acid variations shown below aligned to the homologous regions of mouse, chicken, and zebrafish proteins. RAD50 has a N-terminal Walker A-type ATPase and DNA binding domain (ABC), three coiled-coil regions (CC), a zinc-hook (Zn), and a C-terminal Walker B-type ATPase and DNA binding domain (ABC). MRE11 has a protein phosphatase 2A-like catalytic domain (PP2AC) and a DNA binding domain (DNA BD). NBS1 has forkhead homology region (FHA) and a BRCA1 C-terminus-like domain (BRCT). A box in NBS1 indicates the location containing most of the mutations responsible for Nijmegen breakage syndrome. The illustrated protein segments were derived from ClustalW reference sequence alignments (RAD50, NP_005723, NP_033038, XP_414645, and XP_696859; MRE11, NP_005582, NP_061206, NP_990109, and NP_001001407; NBS1, NP_002476, NP_038780, NP_989668, and NP_001014819).

Table III-1. SNPs in multiple DNA repair genes associate with IgAD/CVID

		Controls	IgAD			CVID			Combined IgAD & CVID		
Gene	SNP	Allele frequency	Allele frequency	p-value	Corrected p-value [a]	Allele frequency	p-value	Corrected p-value [a]	Allele frequency	p-value	Corrected p-value [a]
MSH2	rs4952887	14/184	31/280	0.22	0.96	18/96	0.0054	0.062	49/376	0.06	0.49
MSH2	rs2347794	60/182	117/278	0.049	0.42	47/94	0.0059	0.070	164/372	0.012	0.17
MSH2	rs3771274	62/184	118/278	0.059	0.46	49/96	0.0049	0.053	167/374	0.013	0.18
MSH2	rs3771275	64/184	117/276	0.10	0.77	48/94	0.0088	0.098	165/370	0.027	0.30
MSH2	rs3771276	59/182	116/276	0.038	0.34	48/96	0.0042	**0.0498**	164/372	0.0085	0.13
MSH2	rs6729015	61/184	122/278	0.021	0.22	49/96	0.0036	**0.045**	171/374	0.0046	0.061
MSH2	rs3771281	56/184	116/280	0.016	0.19	43/96	0.017	0.21	159/376	0.0068	0.079
RAD50	rs2237060	59/184	117/280	0.035	0.23	48/96	0.0034	**0.029**	165/376	0.0073	0.057
RAD52	rs9634161	28/184	54/280	0.26	1.00	27/96	0.0099	0.26	81/376	0.076	0.88
RAD52	rs10849605	75/178	123/276	0.61	1.00	59/94	0.0012	**0.040**	182/370	0.12	0.97
RAD54B	rs2046666	71/184	143/280	0.0083	0.23	46/96	0.13	0.98	189/376	0.0092	0.26
RAD54B	rs3019279	63/184	134/276	0.0024	0.076	47/94	0.011	0.29	181/370	0.0010	**0.035**
RAD54B	rs2921385	71/182	145/276	0.0045	0.14	48/96	0.078	0.89	193/372	0.0044	0.13
RAD54B	rs2930968	64/184	132/280	0.0084	0.23	47/94	0.014	0.35	179/374	0.0034	0.11
AID	rs2580874	73/184	151/280	0.0026	**0.010**	52/96	0.021	**0.021**	203/376	0.0015	**0.0055**
AID	rs1561559	2/184	15/280	0.017	0.068	2/94	0.49	0.49	17/374	0.034	0.13
AID	rs714629	73/178	142/274	0.024	0.086	49/90	0.037	**0.037**	191/364	0.012	**0.044**
MSH5	rs805304	68/182	170/280	9.2E-07	**0** [b]	37/94	0.75	1.00	207/374	6.9E-05	**0.0020**
MSH5	rs3117572	142/182	251/278	2.7E-04	**0.0064**	76/96	0.83	1.00	327/374	0.0042	0.090
MSH5	rs3131379	27/182	87/278	6.4E-05	**0.0016**	14/92	0.93	1.00	101/370	0.0011	**0.024**
MSH5	rs3131378	26/182	89/280	2.1E-05	**5E-04**	18/96	0.33	1.00	107/376	2.3E-04	**0.0060**
MSH5	rs3117577	26/180	88/278	3.2E-05	**9E-04**	15/96	0.79	1.00	103/374	6.4E-04	**0.016**
MSH5	rs707938	63/180	154/276	1.4E-05	**3E-04**	35/96	0.81	1.00	189/372	4.7E-04	**0.012**

a) Corrected p-values were adjusted for multiple testing bias by 100,000 permutations of all markers within pathways as described in Materials and Methods. Data presented for MSH2 is from permutation testing of the MutSα pathway.

b) No permutation exceeded the chi-squared result observed for this SNP.

Table III-2. Haplotypes blocks in MLH1, MSH2, RAD50, and RAD54B associate with IgAD/CVID

	IgAD			CVID			Combined IgAD & CVID		
Haplotype	Hap freq.	p-value	Corrected p-value [a]	Hap freq.	p-value	Corrected p-value [a]	Hap freq.	p-value	Corrected p-value [a]
MSH2 – Block 16									
GCGACGCG	0.575	0.079	0.400	0.568	**0.0075**	**0.0250**	0.558	**0.013**	0.066
MSH3 – Block 18									
GTT	0.137	0.64	1.00	0.171	**0.0097**	0.14	0.157	0.19	1.00
MSH3 – Block 19									
GTC	0.009	**0.014**	0.20	0.014	0.14	0.95	0.007	**0.0043**	0.075
MLH1 – Block 11									
ATTATAT	0.008	**0.018**	**0.036**	0.013	0.16	0.64	0.006	**0.0062**	**0.014**
RAD50 – Block 23									
TGAGCA	0.365	0.064	0.500	0.367	**0.0062**	**0.041**	0.386	**0.011**	0.097
TGATCA	0.4	0.37	1.00	0.368	**0.0038**	**0.024**	0.373	0.061	0.63
RAD54B – Block 26									
TAGGGCT	0.508	**0.0062**	0.15	0.55	0.086	0.96	0.503	**0.0057**	0.16
MSH5 – Block 21									
TACGATT	0.139	**2.0E-04**	**0.0033**	0.211	0.94	1.00	0.15	**0.0066**	0.095
TGCAGCC	0.243	**7.5E-05**	**0.0010**	0.146	0.99	1.00	0.227	**0.0017**	**0.023**
MUS81 – Block 22									
AATCA	0.011	**0.0053**	**0.014**	0.018	0.11	0.50	0.009	**0.0014**	**0.0054**

a) p-value corrected for multiple testing bias by 100,000 permutations of all haplotype blocks with a frequency >0.1% within pathways as described in Materials and Methods

Table III-3. Genetic association of SNPs identified by reseqencing

			Controls	IgAD		CVID	
Gene	SNP	Variant	Allele frequency	Allele frequency	p-value	Allele frequency	p-value
MSH2	Broad11461385	T292S	1/1890	2/652	0.10	0/230	0.73
MSH2	Broad11462232	A727S	0/1888	**1/650** [a]	0.088	0/230	--
MLH1	rs1799977	I219V	563/1904	187/656	0.61	79/222	0.065
MLH1	Broad13929360	S247A	0/1876	2/642	**0.016**	0/230	--
MLH1	Broad13369339	Q409P	0/1902	0/656	--	**1/230**	**0.0040**
MLH1	c.1912-1913(AA>GC)	K618A	13/1882	5/646	0.83	2/230	0.76
MLH1	Broad13771867	G638R	0/1872	**1/642** [a]	0.088	0/230	--
MLH1	rs35831931	V716M	3/1904	1/658	0.98	1/230	0.36
MLH1	Broad13929366	H727L	1/1872	0/642	0.56	1/228	0.075
RAD50	Broad13929372	5' UTR	0/1842	**1/640**	0.090	0/230	--
RAD50	Broad13929374	5' UTR	0/1900	**1/658**	0.089	0/230	--
RAD50	Broad13929385	P165H	0/1842	**1/640**	0.0897	0/230	--
RAD50	rs28903091	R327H	4/1826	1/584	0.83	0/214	0.49
RAD50	Broad13929402	Q372X	0/1900	0/654	--	**1/230**	**0.0040**
RAD50	Broad13929448	R1077Q	0/1856	**1/596**	0.078	0/216	--
RAD50	Broad13929454	SYN	0/1900	**1/652**	0.088	0/230	--
MRE11	Broad13928907	3' UTR	2/1866	2/636	0.26	3/228	**0.00040**
MRE11	Broad13928926	3' UTR	0/1864	0/636	--	**1/228**	**0.0042**
MRE11	Broad13928970	A492D	2/1898	3/652	0.077	1/230	0.21
MRE11	Broad13928971	E494K	8/1876	0/642	0.098	2/230	0.36
NBS1	Broad13931251	Intron	0/1814	**1/602** [a]	0.083	0/226	--
NBS1	rs1805794	E185Q	625/1892	207/650	0.58	77/224	0.69
NBS1	Broad13929731	V210F	4/1858	2/638	0.66	0/230	0.48
NBS1	Broad13929706	P401R	0/1892	**2/646** [a]	**0.016**	0/230	--
NBS1	Broad13929705	L421S	1/1884	0/644	0.56	1/230	0.076
NBS1	Broad13931239	D527Y	0/1844	**1/598**	0.079	0/212	--
RAD52	rs7487683	G180R	74/1876	17/644	0.13	9/230	0.98
APEX1	rs2307490	5' UTR	3/1872	2/642	0.46	2/230	**0.037**
APEX1	rs1048945	Q51H	60/1840	12/626	0.085	9/212	0.45

a) A single IgAD patient carried all four alleles, MLH1-G638R, MSH2-A727S, NBS1-P401R, and the intronic SNP Broad13931251

Supplementary Table III-4. Control genotype data for SNPs attempted in genotyping screen. Legend: chr:location, chromosome locations relative to human genome build 36 coordinates; % Genotyped, percent of control individuals successfully genotyped for a given marker; PredHET, predicted heterozygocity based on allele frequencies; ObsHET, observed heterozygocity; HWpval, p-value from chi-squared test of ObsHET compared to PredHET; QC pass, 303 SNPs for which >90% of samples were successfully genotyped and HWpval >0.001 are marked with a check (√).

Gene	SNP	chr:location	% Genotyped	PredHET	ObsHET	HWpval	QC pass
AID_close	rs2119089	chr12:8643499	0	-	-	-	-
AID	rs2028373	chr12:8648748	98.9	0.47	0.48	1.00	√
AID	rs2580874	chr12:8650077	100	0.48	0.42	0.35	√
AID	rs1561559	chr12:8654272	100	0.02	0.02	1.00	√
AID_close	rs714629	chr12:8657628	96.7	0.48	0.46	0.77	√
BLM	rs7184015	chr15:89075579	100	0.38	0.41	0.53	√
BLM	rs8034371	chr15:89089359	100	0.34	0.35	1.00	√
BLM	rs6496724	chr15:89090734	98.9	0.40	0.40	1.00	√
BLM	rs7183841	chr15:89095901	96.7	0.33	0.35	0.90	√
BLM	rs6496725	chr15:89096964	98.9	0.33	0.35	0.84	√
BLM	rs3784782	chr15:89103252	98.9	0.34	0.37	0.64	√
BLM	rs2518968	chr15:89108414	95.7	0.49	0.58	0.16	√
BLM	rs3784780	chr15:89110479	100	0.32	0.35	0.55	√
BLM	rs3815003	chr15:89113827	100	0.39	0.36	0.56	√
BLM	rs8036601	chr15:89116502	97.8	0.47	0.42	0.42	√
BLM	rs4932363	chr15:89124105	100	0.09	0.10	1.00	√
BLM	rs2227935	chr15:89127103	0	-	-	-	-
BLM	rs7175811	chr15:89132550	100	0.49	0.46	0.64	√
BLM	rs2229035	chr15:89138509	98.9	0.00	0.00	1.00	√
BLM	rs7182287	chr15:89145670	91.3	0.36	0.35	0.94	√
BLM	rs1801256	chr15:89148467	97.8	0.00	0.00	1.00	√
BLM	rs2270132	chr15:89152872	98.9	0.48	0.50	1.00	√
BLM	rs2073919	chr15:89154123	98.9	0.37	0.37	1.00	√
BLM	rs7167216	chr15:89155525	100	0.12	0.13	1.00	√
BLM	rs414634	chr15:89157257	98.9	0.40	0.43	0.79	√
BLM	rs374294	chr15:89158033	0	-	-	-	-
BLM_close	rs447804	chr15:89162940	0	-	-	-	-
DMC1_close	rs1946990	chr22:37242717	100	0.47	0.57	0.10	√
DMC1	rs5750616	chr22:37248840	100	0.37	0.36	0.94	√
DMC1	rs1980455	chr22:37255665	98.9	0.47	0.57	0.06	√
DMC1	rs1292811	chr22:37257832	100	0.08	0.09	1.00	√
DMC1	rs2227914	chr22:37264552	98.9	0.00	0.00	1.00	√
DMC1	rs5757130	chr22:37267740	97.8	0.48	0.60	0.025	√
DMC1	rs4821792	chr22:37272872	100	0.08	0.09	1.00	√
DMC1	rs5757133	chr22:37277781	98.9	0.47	0.50	0.75	√
DMC1	rs5757135	chr22:37282519	98.9	0.48	0.59	0.040	√
DMC1	rs5757141	chr22:37287429	96.7	0.46	0.55	0.10	√
DMC1	rs1129426	chr22:37292674	0	-	-	-	-
DMC1_close	rs8140617	chr22:37300385	100	0.45	0.53	0.14	√
DMC1_close	rs1013339	chr22:37304626	0	-	-	-	-
DMC1_close	rs1569492	chr22:37305457	98.9	0.48	0.59	0.040	√
CD3EAP,ERCC1_close	rs1046282	chr19:50602512	98.9	0.29	0.33	0.37	√
ERCC1	rs3212977	chr19:50608822	100	0.00	0.00	1.00	√
ERCC1	rs11615	chr19:50615493	100	0.44	0.45	1.00	√
ERCC1	rs3188420	chr19:50616367	0	-	-	-	-
ERCC1_close	rs1319052	chr19:50623061	96.7	0.41	0.43	1.00	√
LOC100133057,EXO1_close	rs10802995	chr1:240068905	100	0.48	0.46	0.81	√
EXO1,LOC100133057	rs1635518	chr1:240077197	100	0.50	0.52	0.84	√
EXO1	rs4149864	chr1:240082281	98.9	0.00	0.00	1.00	√
EXO1	rs1635515	chr1:240083632	97.8	0.48	0.47	0.99	√
EXO1	rs2526700	chr1:240086622	98.9	0.50	0.46	0.56	√
EXO1	rs4149910	chr1:240090581	100	0.00	0.00	1.00	√
EXO1	rs1776133	chr1:240091721	97.8	0.48	0.46	0.72	√
EXO1	rs2526698	chr1:240095724	100	0.50	0.48	0.79	√
EXO1	rs735943	chr1:240096774	98.9	0.49	0.46	0.74	√
EXO1	rs851781	chr1:240101491	100	0.50	0.44	0.27	√
EXO1	rs4149963	chr1:240102005	98.9	0.16	0.15	1.00	√
EXO1	rs4149966	chr1:240102067	100	0.01	0.01	1.00	√

Supplementary Table III-4, continued.

Gene	SNP	chr:location	% Genotyped	PredHET	ObsHET	HWpval	QC pass
EXO1	rs4149967	chr1:240102197	100	0.00	0.00	1.00	√
EXO1	rs2526697	chr1:240106752	100	0.49	0.50	1.00	√
EXO1	rs1047840	chr1:240108924	100	0.47	0.41	0.36	√
EXO1	rs12122770	chr1:240108987	100	0.03	0.03	1.00	√
EXO1	rs4149978	chr1:240109060	100	0.01	0.01	1.00	√
EXO1	rs1776148	chr1:240109168	98.9	0.48	0.44	0.49	√
EXO1	rs1635498	chr1:240111898	100	0.04	0.04	1.00	√
EXO1	rs3886571	chr1:240113323	0	-	-	-	-
EXO1	rs9350	chr1:240115297	100	0.23	0.24	1.00	√
EXO1	rs4150005	chr1:240117359	0	-	-	-	-
EXO1	rs4150018	chr1:240119244	100	0.46	0.44	0.72	√
EXO1_close	rs2526694	chr1:240124312	0	-	-	-	-
H2AFX,HMBS	rs1784304	chr11:118468026	100	0.36	0.26	0.016	√
H2AFX	rs7350	chr11:118470258	100	0.44	0.55	0.030	√
DPAGT1,H2AFX_close	rs643788	chr11:118472968	100	0.47	0.54	0.18	√
DPAGT1,H2AFX_close	rs649870	chr11:118476461	98.9	0.46	0.55	0.12	√
DPAGT1,H2AFX_close	rs3825059	chr11:118478323	0	-	-	-	-
H2AFX_close	rs4938628	chr11:118480088	0	-	-	-	-
LIG1	rs251692	chr19:53310523	0	-	-	-	-
LIG1	rs11666150	chr19:53310942	0	-	-	-	-
LIG1	rs11668325	chr19:53312774	100	0.01	0.01	1.00	√
LIG1	rs274883	chr19:53314357	100	0.23	0.20	0.34	√
LIG1	rs3731008	chr19:53318362	0	-	-	-	-
LIG1	rs156633	chr19:53319074	95.7	0.50	0.44	0.39	√
LIG1	rs3731003	chr19:53323070	100	0.00	0.00	1.00	√
LIG1	rs156641	chr19:53323220	97.8	0.48	0.44	0.54	√
LIG1	rs2288883	chr19:53328346	0	-	-	-	-
LIG1	rs3730980	chr19:53330834	0	-	-	-	-
LIG1	rs4987068	chr19:53332619	100	0.06	0.07	1.00	√
LIG1	rs3730966	chr19:53332740	0	-	-	-	-
LIG1	rs2304136	chr19:53334936	98.9	0.09	0.10	1.00	√
LIG1	rs3730947	chr19:53335082	0	-	-	-	-
LIG1	rs2288878	chr19:53338695	98.9	0.50	0.47	0.77	√
LIG1	rs3730933	chr19:53339009	100	0.02	0.02	1.00	√
LIG1	rs3730913	chr19:53344612	100	0.24	0.23	1.00	√
LIG1	rs3730911	chr19:53344862	98.9	0.00	0.00	1.00	√
LIG1	rs11879148	chr19:53345246	100	0.00	0.00	1.00	√
LIG1	rs274860	chr19:53349432	100	0.09	0.10	1.00	√
LIG1	rs12981963	chr19:53352097	100	0.00	0.00	1.00	√
LIG1	rs4987181	chr19:53356529	100	0.01	0.01	1.00	√
LIG1	rs3730862	chr19:53356610	0	-	-	-	-
LIG1,LOC374920	rs3730842	chr19:53364812	0	-	-	-	-
LOC374920,LIG1_close	rs274869	chr19:53369960	98.9	0.50	0.47	0.74	√
LOC374920,LIG1_close	rs274873	chr19:53371521	98.9	0.50	0.46	0.60	√
EPM2AIP1,MLH1_close	rs1055095	chr3:37002937	97.8	0.50	0.52	0.87	√
EPM2AIP1,MLH1	rs1800734	chr3:37009950	98.9	0.36	0.33	0.64	√
MLH1	rs9852378	chr3:37013775	100	0.49	0.50	1.00	√
MLH1	rs11541859	chr3:37017507	100	0.00	0.00	1.00	√
MLH1	rs4647224	chr3:37018234	97.8	0.49	0.48	0.90	√
MLH1	rs4234259	chr3:37023637	0	-	-	-	-
MLH1	rs4647250	chr3:37024102	96.7	0.49	0.44	0.43	√
MLH1	rs4647257	chr3:37028240	100	0.00	0.00	1.00	√
MLH1	rs1799977	chr3:37028572	0	-	-	-	-
LOC100131713,MLH1	rs1558528	chr3:37031994	97.8	0.49	0.48	0.90	√
LOC100131713,MLH1	rs4647277	chr3:37033513	0	-	-	-	-
MLH1	rs2286939	chr3:37037044	97.8	0.48	0.43	0.42	√
MLH1	rs2286940	chr3:37045110	97.8	0.49	0.49	1.00	√
MLH1	rs3774332	chr3:37049672	0	-	-	-	-
MLH1	rs9311150	chr3:37061692	100	0.00	0.00	1.00	√
MLH1	rs2241031	chr3:37065278	100	0.50	0.47	0.70	√
MLH1	rs1800149	chr3:37067062	100	0.00	0.00	1.00	√

Supplementary Table III-4, continued.

Gene	SNP	chr:location	% Genotyped	PredHET	ObsHET	HWpval	QC pass
MLH3	rs175047	chr14:74548792	96.7	0.50	0.46	0.59	√
MLH3	rs108621	chr14:74550390	0	-	-	-	-
MLH3	rs108622	chr14:74551661	100	0.50	0.47	0.66	√
MLH3	rs13712	chr14:74553565	0	-	-	-	-
MLH3	rs175052	chr14:74555159	100	0.00	0.00	1.00	√
MLH3	rs175057	chr14:74559385	100	0.50	0.44	0.28	√
MLH3	rs175062	chr14:74562047	0	-	-	-	-
MLH3	rs2098252	chr14:74562485	100	0.50	0.42	0.19	√
MLH3	rs175067	chr14:74565825	97.8	0.00	0.00	1.00	√
MLH3	rs735452	chr14:74567737	97.8	0.50	0.48	0.80	√
MLH3	rs149220	chr14:74573426	100	0.00	0.00	1.00	√
MLH3	rs3742780	chr14:74574761	97.8	0.50	0.52	0.85	√
MLH3	rs10136948	chr14:74580828	98.9	0.50	0.51	1.00	√
MLH3	rs175081	chr14:74583636	100	0.00	0.00	1.00	√
MLH3	rs175082	chr14:74587484	100	0.00	0.00	1.00	√
ACYP1,MLH3_close	rs10142770	chr14:74595202	98.9	0.50	0.51	1.00	√
MRE11	rs516037	chr11:93788460	0	-	-	-	-
MRE11	rs2155209	chr11:93790438	96.7	0.45	0.44	0.93	√
MRE11	rs1805362	chr11:93792974	98.9	0.00	0.00	1.00	√
MRE11	rs661957	chr11:93803832	97.8	0.44	0.46	0.99	√
MRE11	rs518276	chr11:93806905	100	0.47	0.54	0.23	√
MRE11	rs682213	chr11:93816167	97.8	0.50	0.62	0.031	√
MRE11	rs569143	chr11:93828035	98.9	0.49	0.57	0.18	√
MRE11	rs654718	chr11:93829763	100	0.45	0.49	0.59	√
MRE11	rs1805367	chr11:93832319	100	0.00	0.00	1.00	√
MRE11	rs680695	chr11:93851802	0	-	-	-	-
MRE11	rs1061945	chr11:93863708	100	0.00	0.00	1.00	√
ANKRD49,MRE11	rs497763	chr11:93865568	100	0.50	0.64	0.012	√
ANKRD49,MRE11_close	rs472344	chr11:93867917	97.8	0.50	0.61	0.059	√
MSH2	rs12612908	chr2:47481793	100	0.00	0.00	1.00	√
MSH2	rs1863332	chr2:47483402	97.8	0.20	0.20	1.00	√
MSH2	rs3815865	chr2:47484143	0	-	-	-	-
MSH2	rs1800151	chr2:47490943	0	-	-	-	-
MSH2	rs4987188	chr2:47496961	98.9	0.01	0.01	1.00	√
MSH2	rs4952887	chr2:47500472	100	0.14	0.15	1.00	√
MSH2	rs2347794	chr2:47510305	98.9	0.44	0.40	0.41	√
MSH2	rs3771274	chr2:47517349	100	0.45	0.39	0.31	√
MSH2	rs6726691	chr2:47523808	100	0.37	0.32	0.24	√
MSH2	rs1981928	chr2:47525979	98.9	0.38	0.34	0.40	√
MSH2	rs3771275	chr2:47528835	100	0.45	0.39	0.25	√
MSH2	rs3771276	chr2:47533677	98.9	0.44	0.41	0.61	√
MSH2	rs6729015	chr2:47540077	100	0.44	0.42	0.81	√
MSH2	rs3771278	chr2:47543908	0	-	-	-	-
MSH2	rs3771281	chr2:47545785	100	0.42	0.39	0.59	√
MSH2	rs3764959	chr2:47551812	96.7	0.39	0.37	0.81	√
MSH2	rs1800152	chr2:47555825	0	-	-	-	-
MSH2	rs2229061	chr2:47557207	0	-	-	-	-
MSH2	rs2042649	chr2:47563208	98.9	0.14	0.15	1.00	√
MSH2	rs1802577	chr2:47563570	100	0.00	0.00	1.00	√
DHFR,MSH3_close	rs844370	chr5:79978758	98.9	0.34	0.32	0.77	√
MSH3	rs380691	chr5:79987790	100	0.43	0.48	0.46	√
MSH3	rs1650670	chr5:79996605	96.7	0.35	0.27	0.066	√
MSH3	rs1650666	chr5:79998195	97.8	0.35	0.27	0.061	√
MSH3	rs1650663	chr5:79998953	97.8	0.36	0.27	0.034	√
MSH3	rs1677649	chr5:80001953	97.8	0.37	0.27	0.019	√
MSH3	rs836810	chr5:80007968	0	-	-	-	-
MSH3	rs6864493	chr5:80021458	100	0.33	0.36	0.76	√
MSH3	rs836817	chr5:80028174	0	-	-	-	-
MSH3	rs3776968	chr5:80042314	100	0.43	0.41	0.81	√
MSH3	rs245016	chr5:80056949	0	-	-	-	-
MSH3	rs3852191	chr5:80063586	0	-	-	-	-
MSH3	rs26282	chr5:80086737	100	0.39	0.26	0.0049	√
MSH3	rs33013	chr5:80095772	98.9	0.41	0.41	1.00	√

Supplementary Table III-4, continued.

Gene	SNP	chr:location	% Genotyped	PredHET	ObsHET	HWpval	QC pass
MSH3	rs6151838	chr5:80118621	100	0.19	0.19	1.00	√
MSH3	rs42290	chr5:80124778	100	0.24	0.24	1.00	√
MSH3	rs3797896	chr5:80129556	97.8	0.16	0.16	1.00	√
MSH3	rs245391	chr5:80139799	98.9	0.36	0.39	0.80	√
MSH3	rs32991	chr5:80150158	100	0.50	0.42	0.21	√
MSH3	rs26910	chr5:80156689	96.7	0.22	0.23	1.00	√
MSH3	rs3776978	chr5:80162567	100	0.20	0.19	0.64	√
MSH3	rs184967	chr5:80185737	100	0.27	0.23	0.31	√
MSH3	rs245341	chr5:80196149	96.7	0.36	0.38	0.85	√
MSH3	rs27887	chr5:80205233	98.9	0.38	0.40	0.92	√
MSH3_close	rs32964	chr5:80211553	100	0.38	0.40	0.84	√
RABGGTB,MSH4,close	rs1498311	chr1:76027314	100	0.41	0.39	0.88	√
MSH4	rs5745311	chr1:76035437	0	-	-	-	-
MSH4	rs1146646	chr1:76038088	96.7	0.35	0.27	0.066	√
MSH4	rs1146649	chr1:76041391	0	-	-	-	-
MSH4	rs5745325	chr1:76042048	98.9	0.42	0.37	0.42	√
MSH4	rs5745327	chr1:76042628	98.9	0.19	0.19	1.00	√
MSH4	rs5745329	chr1:76045310	97.8	0.00	0.00	1.00	√
MSH4	rs1146652	chr1:76046296	100	0.14	0.15	1.00	√
MSH4	rs5745353	chr1:76047725	0	-	-	-	-
MSH4	rs1144336	chr1:76050009	0	-	-	-	-
MSH4	rs1144342	chr1:76056952	98.9	0.19	0.17	0.48	√
MSH4	rs5745384	chr1:76061203	0	-	-	-	-
MSH4	rs5745390	chr1:76079179	98.9	0.33	0.26	0.10	√
MSH4	rs1565717	chr1:76093029	100	0.33	0.26	0.10	√
MSH4	rs2029682	chr1:76114207	0	-	-	-	-
MSH4	rs5745458	chr1:76118017	98.9	0.34	0.26	0.057	√
MSH4	rs5745459	chr1:76118411	100	0.07	0.05	0.22	√
MSH4	rs3819949	chr1:76126668	0	-	-	-	-
MSH4	rs1001160	chr1:76131179	98.9	0.40	0.39	0.80	√
MSH4	rs5745543	chr1:76137699	96.7	0.41	0.39	0.86	√
MSH4	rs5745545	chr1:76149944	100	0.41	0.38	0.61	√
MSH4	rs5745549	chr1:76151090	97.8	0.06	0.07	1.00	√
ASB17,MSH4_close	rs946163	chr1:76155637	100	0.41	0.39	0.88	√
CLIC1,DDAH2,MSH5_close	rs805304	chr6:31806067	98.9	0.47	0.48	0.97	√
CLIC1,MSH5_close	rs400547	chr6:31810689	100	0.10	0.09	0.45	√
MSH5	rs3131382	chr6:31815709	100	0.12	0.13	1.00	√
MSH5	rs409558	chr6:31816126	100	0.27	0.30	0.52	√
MSH5	rs2075789	chr6:31816307	0	-	-	-	-
MSH5	rs28381349	chr6:31817023	100	0.04	0.04	1.00	√
MSH5	rs707915	chr6:31818947	100	0.10	0.09	0.45	√
MSH5	rs2075788	chr6:31820160	100	0.27	0.30	0.52	√
MSH5	rs3117572	chr6:31825671	98.9	0.34	0.26	0.057	√
MSH5	rs2299851	chr6:31826581	100	0.19	0.17	0.55	√
MSH5	rs3131379	chr6:31829012	98.9	0.25	0.21	0.20	√
MSH5	rs3131378	chr6:31833264	98.9	0.25	0.20	0.15	√
MSH5	rs707939	chr6:31834667	97.8	0.41	0.43	0.76	√
MSH5	rs3117577	chr6:31835453	97.8	0.25	0.20	0.16	√
MSH5	rs3115672	chr6:31835876	0	-	-	-	-
C6orf26,MSH5	rs707938	chr6:31837338	97.8	0.46	0.46	1.00	√
C6orf26,MSH5	rs1802127	chr6:31837904	100	0.04	0.04	1.00	√
C6orf26,MSH5_close	rs707937	chr6:31838993	98.9	0.31	0.32	1.00	√
C6orf27,MSH5_close	rs707936	chr6:31841629	100	0.12	0.11	0.64	√
MSH6	rs330792	chr2:47862078	0	-	-	-	-
MSH6	rs3136245	chr2:47866350	100	0.33	0.34	1.00	√
MSH6	rs1878484	chr2:47870051	100	0.02	0.02	1.00	√
MSH6	rs1800932	chr2:47871585	97.8	0.25	0.22	0.52	√
MSH6	rs3211299	chr2:47871740	96.7	0.00	0.00	1.00	√
MSH6	rs1800935	chr2:47876619	0	-	-	-	-
MSH6	rs1800938	chr2:47879286	100	0.00	0.00	1.00	√
MSH6	rs2020908	chr2:47879812	100	0.04	0.00	0.00020	-

Supplementary Table III-4, continued.

Gene	SNP	chr:location	% Genotyped	PredHET	ObsHET	HWpval	QC pass
MSH6	rs728619	chr2:47880239	0	-	-	-	-
MSH6	rs3136334	chr2:47880493	100	0.00	0.00	1.00	√
MSH6	rs2020912	chr2:47881259	100	0.02	0.02	1.00	√
MSH6	rs2020911	chr2:47884342	100	0.45	0.38	0.19	√
FBXO11,MSH6	rs3136367	chr2:47887055	0	-	-	-	-
CFL1,MUS81_close	rs635375	chr11:65378571	96.7	0.39	0.39	1.00	√
CFL1,MUS81_close	rs652021	chr11:65379923	98.9	0.39	0.41	1.00	√
CFL1,MUS81_close	rs665306	chr11:65380601	94.6	0.44	0.48	0.61	√
MUS81	rs13817	chr11:65384910	0	-	-	-	-
MUS81	rs545500	chr11:65386510	0	-	-	-	-
EFEMP2,MUS81	rs630303	chr11:65388549	100	0.42	0.44	0.88	√
EFEMP2,MUS81	rs765593	chr11:65389308	94.6	0.00	0.00	1.00	√
EFEMP2,MUS81_close	rs659824	chr11:65393085	96.7	0.50	0.48	0.93	√
EFEMP2,MUS81_close	rs630394	chr11:65394560	98.9	0.43	0.39	0.41	√
NBS1	rs3026268	chr8:91035004	98.9	0.00	0.00	1.00	√
NBS1	rs769420	chr8:91051867	100	0.00	0.00	1.00	√
NBS1	rs1805794	chr8:91059655	96.7	0.47	0.45	0.86	√
RNF216,PMS2_close	rs710939	chr7:5657216	100	0.03	0.03	1.00	√
RNF216,PMS2_close	rs852520	chr7:5661793	100	0.50	0.53	0.71	√
RNF216,PMS2_close	rs852516	chr7:5664454	100	0.11	0.12	1.00	√
RNF216,PMS2_close	rs852417	chr7:5670195	100	0.46	0.46	1.00	√
RNF216,PMS2_close	rs852413	chr7:5675052	100	0.03	0.03	1.00	√
RNF216,PMS2_close	rs3779092	chr7:5679126	96.7	0.50	0.46	0.56	√
RNF216,PMS2_close	rs852394	chr7:5687974	95.7	0.50	0.53	0.63	√
RNF216,PMS2_close	rs4724712	chr7:5692548	100	0.50	0.57	0.28	√
RNF216,PMS2_close	rs852266	chr7:5694525	98.9	0.10	0.11	1.00	√
RNF216,PMS2_close	rs852262	chr7:5696742	100	0.02	0.02	1.00	√
RNF216,PMS2_close	rs1468996	chr7:5700983	97.8	0.47	0.46	0.94	√
PMS2	rs1805321	chr7:5993514	100	0.00	0.00	1.00	√
PMS2	rs2286680	chr7:6002033	0	-	-	-	-
JTV1,PMS2_close	rs2009115	chr7:6019082	98.9	0.14	0.15	1.00	√
JTV1,PMS2_close	rs3779107	chr7:6022455	98.9	0.14	0.15	1.00	√
JTV1,PMS2_close	rs1860459	chr7:6023145	100	0.18	0.17	1.00	√
XPO5,POLH_close	rs699937	chr6:43648904	98.9	0.41	0.37	0.54	√
POLH	rs6458343	chr6:43654356	98.9	0.10	0.11	1.00	√
POLH	rs2307456	chr6:43673546	100	0.01	0.01	1.00	√
POLH	rs9333548	chr6:43680445	100	0.00	0.00	1.00	√
POLH	rs9296419	chr6:43689563	0	-	-	-	-
POLH	rs9333555	chr6:43689913	0	-	-	-	-
POLH	rs6941583	chr6:43690069	0	-	-	-	-
RAD50_close	rs2522410	chr5:131913216	100	0.00	0.00	1.00	√
RAD50	rs4526098	chr5:131920878	100	0.00	0.00	1.00	√
RAD50	rs2244012	chr5:131929124	100	0.38	0.39	0.95	√
RAD50	rs2706348	chr5:131933709	98.9	0.37	0.39	1.00	√
RAD50	rs2230017	chr5:131943473	100	0.00	0.00	1.00	√
RAD50	rs2252775	chr5:131946343	98.9	0.37	0.39	1.00	√
RAD50	rs1047380	chr5:131958512	100	0.00	0.00	1.00	√
RAD50	rs1047382	chr5:131959284	100	0.00	0.00	1.00	√
RAD50	rs3187395	chr5:131972260	100	0.01	0.01	1.00	√
RAD50	rs1047386	chr5:131972768	100	0.00	0.00	1.00	√
RAD50	rs1047387	chr5:131979669	100	0.00	0.00	1.00	√
RAD50	rs2237060	chr5:131998784	100	0.44	0.42	0.93	√
RAD50	rs2240032	chr5:132005026	100	0.38	0.37	1.00	√
RAD50_close	rs2158177	chr5:132011957	100	0.36	0.36	1.00	√
RAD51_close	rs2412545	chr15:38766943	100	0.48	0.49	1.00	√
RAD51	rs2619681	chr15:38776313	100	0.24	0.27	0.30	√
RAD51	rs4924496	chr15:38782627	0	-	-	-	-
RAD51	rs7174493	chr15:38785766	100	0.00	0.00	1.00	√
RAD51	rs2412546	chr15:38793815	98.9	0.50	0.54	0.58	√
RAD51	rs11858337	chr15:38795572	98.9	0.50	0.54	0.61	√

Supplementary Table III-4, continued.

Gene	SNP	chr:location	% Genotyped	PredHET	ObsHET	HWpval	QC pass
RAD51	rs957603	chr15:38796960	98.9	0.48	0.53	0.48	√
RAD51	rs11070291	chr15:38803850	100	0.50	0.53	0.64	√
RAD51	rs11544205	chr15:38809068	0	-	-	-	-
RAD51	rs2229876	chr15:38809343	0	-	-	-	-
RAD51	rs1056742	chr15:38810585	100	0.00	0.00	1.00	√
FAM82A2,RAD51_close	rs4924501	chr15:38814286	100	0.50	0.57	0.28	√
FAM82A2,RAD51_close	rs11558809	chr15:38815668	97.8	0.01	0.01	1.00	√
RAD52,WNK1	rs1060499	chr12:890527	98.9	0.27	0.19	0.017	√
RAD52	rs11226	chr12:892074	0	-	-	-	-
RAD52	rs1051669	chr12:892713	97.8	0.32	0.31	0.98	√
RAD52	rs10744729	chr12:894855	98.9	0.50	0.47	0.77	√
RAD52	rs4766370	chr12:897623	98.9	0.50	0.46	0.62	√
RAD52	rs9634161	chr12:900192	100	0.26	0.20	0.061	√
RAD52	rs7312883	chr12:905165	100	0.00	0.00	1.00	√
RAD52	rs7487683	chr12:906303	100	0.03	0.03	1.00	√
RAD52	rs4766377	chr12:909003	98.9	0.33	0.31	0.67	√
RAD52	rs1131839	chr12:909294	100	0.00	0.00	1.00	√
RAD52	rs11571421	chr12:909550	0	-	-	-	-
RAD52	rs10774471	chr12:912282	0	-	-	-	-
RAD52	rs1833095	chr12:916469	100	0.39	0.39	1.00	√
RAD52	rs7311151	chr12:920515	100	0.49	0.47	0.83	√
RAD52	rs11064602	chr12:923999	0	-	-	-	-
RAD52	rs2887531	chr12:926148	100	0.40	0.39	1.00	√
RAD52	rs3748522	chr12:928949	0	-	-	-	-
RAD52_close	rs10849605	chr12:934699	96.7	0.49	0.44	0.43	√
RAD54B	rs2046666	chr8:95457324	100	0.47	0.45	0.68	√
RAD54B	rs2470740	chr8:95468727	97.8	0.48	0.44	0.59	√
RAD54B	rs2046663	chr8:95476626	100	0.16	0.13	0.26	√
RAD54B	rs3019149	chr8:95499755	100	0.38	0.36	0.73	√
LOC100128414,RAD54B	rs3136421	chr8:95514505	100	0.10	0.11	1.00	√
LOC100128414,RAD54B	rs3019279	chr8:95517081	100	0.45	0.51	0.31	√
RAD54B	rs2921385	chr8:95528524	98.9	0.48	0.47	1.00	√
RAD54B	rs2930968	chr8:95531871	100	0.45	0.48	0.82	√
RAD54B	rs1372048	chr8:95536417	100	0.39	0.38	0.95	√
RAD54B	rs2919661	chr8:95539687	100	0.00	0.00	1.00	√
RAD54B_close	rs1992371	chr8:95564238	96.7	0.47	0.49	0.78	√
SPO11_close	rs6014975	chr20:55328836	0	-	-	-	-
SPO11_close	rs12624637	chr20:55335323	0	-	-	-	-
SPO11	rs3736832	chr20:55341676	98.9	0.00	0.00	1.00	√
SPO11	rs6099553	chr20:55344526	100	0.48	0.41	0.27	√
SPO11	rs1467581	chr20:55349761	97.8	0.02	0.02	1.00	√
SPO11	rs2236330	chr20:55351421	0	-	-	-	-
TUBGCP4,TP53BP1_close	rs2244746	chr15:41482375	0	-	-	-	-
TP53BP1	rs2242067	chr15:41486595	0	-	-	-	-
TP53BP1	rs1058298	chr15:41488222	97.8	0.42	0.41	1.00	√
TP53BP1	rs11554564	chr15:41499999	100	0.00	0.00	1.00	√
TP53BP1	rs2230449	chr15:41500136	100	0.00	0.00	1.00	√
TP53BP1	rs2242069	chr15:41500926	100	0.27	0.21	0.085	√
TP53BP1	rs542898	chr15:41506684	100	0.42	0.41	0.96	√
TP53BP1	rs3803339	chr15:41511824	0	-	-	-	-
TP53BP1	rs2602141	chr15:41511938	100	0.42	0.41	0.96	√
TP53BP1	rs536313	chr15:41517290	95.7	0.41	0.39	0.78	√
TP53BP1	rs2256238	chr15:41526951	98.9	0.00	0.00	1.00	√
TP53BP1	rs2467739	chr15:41527488	97.8	0.13	0.08	0.0085	√
TP53BP1	rs2467741	chr15:41529918	98.9	0.43	0.40	0.62	√
TP53BP1	rs690367	chr15:41535596	0	-	-	-	-
TP53BP1	rs694725	chr15:41543502	98.9	0.42	0.44	0.85	√

Supplementary Table III-4, continued.

Gene	SNP	chr:location	% Genotyped	PredHET	ObsHET	HWpval	QC pass
TP53BP1	rs689647	chr15:41549488	100	0.15	0.10	0.022	√
TP53BP1	rs560191	chr15:41555066	100	0.43	0.44	1.00	√
TP53BP1	rs689754	chr15:41563187	100	0.14	0.07	0.00060	-
TP53BP1	rs2439850	chr15:41572146	98.9	0.15	0.10	0.022	√
TP53BP1	rs495175	chr15:41585976	0	-	-	-	-
TP53BP1_close	rs1869258	chr15:41590913	100	0.42	0.41	1.00	√
MAP1A,TP53BP1_close	rs523156	chr15:41599135	100	0.41	0.42	1.00	√
XRCC2_close	rs6962238	chr7:151970041	100	0.08	0.07	0.29	√
XRCC2	rs3218536	chr7:151976940	97.8	0.11	0.11	1.00	√
XRCC2	rs3111471	chr7:151993943	100	0.50	0.52	0.86	√
XRCC2	rs3218408	chr7:151998549	100	0.36	0.37	1.00	√
XRCC2_close	rs2040639	chr7:152006121	98.9	0.50	0.47	0.74	√
XRCC2_close	rs6464268	chr7:152012083	100	0.21	0.22	1.00	√
XRCC2_close	rs479215	chr7:152408312	0	-	-	-	-
XRCC2_close	rs10227264	chr7:152409893	98.9	0.30	0.24	0.11	√
XRCC2_close	rs684088	chr7:152415847	100	0.48	0.55	0.25	√
XRCC2_close	rs513586	chr7:152424532	100	0.13	0.12	0.73	√
XRCC2_close	rs2018083	chr7:152437650	98.9	0.45	0.39	0.26	√
KLC1,XRCC3_close	rs2273175	chr14:103229894	100	0.46	0.42	0.52	√
KLC1,XRCC3	rs861539	chr14:103235506	0	-	-	-	-
KLC1,XRCC3	rs709399	chr14:103237298	0	-	-	-	-
XRCC3	rs3212057	chr14:103243218	0	-	-	-	-
XRCC3	rs3212038	chr14:103247939	0	-	-	-	-
ZFYVE21,XRCC3_close	rs941474	chr14:103259614	95.7	0.50	0.51	1.00	√

Supplementary Table III-5. Target gene resequencing coverage by gene.

Gene	# of targets (exons)	# of 2x Covered Targets	Exons not 2x covered	Target bases	1x covered bases	2x covered bases	% 1x covered bases	% 2x covered bases	Territory bases	1x territory bases covered	2x territory bases covered	% 1x territory bases covered	% 2x territory bases covered
AID	5	5	-	2,791	2,545	2,369	91.2%	84.9%	267,936	244,258	228,141	91.2%	85.2%
APEX1	7	7	-	1,953	1,953	1,934	100.0%	99.0%	187,488	184,553	180,785	98.4%	96.4%
ERCC1	12	5	2-4, 7-9, 12	1,687	1,255	921	74.4%	54.6%	161,952	120,868	107,383	74.6%	66.3%
MLH1	19	17	10, 12	2,522	2,434	2,263	96.5%	89.7%	242,112	230,143	215,311	95.1%	88.9%
MRE11	21	19	18, 20	5,331	5,330	5,240	100.0%	98.3%	511,776	506,764	496,390	99.0%	97.0%
MSH2	16	12	2, 4, 7, 16	3,145	2,635	2,501	83.8%	79.5%	301,920	249,097	237,490	82.5%	78.7%
NBS1	17	16	1	4,671	4,398	4,325	94.2%	92.6%	448,416	415,851	405,753	92.7%	90.5%
RAD50	26	23	7, 11, 23	5,950	5,681	5,582	95.5%	93.8%	571,200	536,411	522,613	93.9%	91.5%
RAD52	18	15	1, 2, 18	4,104	3,284	3,135	80.0%	76.4%	393,984	350,473	335,790	89.0%	85.2%

Supplemental Table III-6. Genetic association of single markers with IgAD and CVID. Bold values are used to highlight significant p-values (P<0.05) and allele frequencies of SNPs detected solely in patient groups. Statistics for the combined IgAD/CVID cohort are bold if p-values are more significant in the combined group compared to IgAD and CVID individually.

			Controls	IgAD			CVID			IgAD & CVID combined		
Gene	SNP	Associated allele[a]	Allele frequency	Allele frequency	p-value	Corrected p-value[b]	Allele frequency	p-value	Corrected p-value[b]	Allele frequency	p-value	Corrected p-value[b]
AID[i]	rs2028373	A	112/182	175/280	0.84	1.00	64/96	0.40	0.84	239/376	0.64	0.98
AID[i]	rs2580874	T	73/184	151/280	**0.0026**	**0.010**	52/96	**0.021**	0.073	203/376	**0.0015**	**0.0055**
AID[i]	rs1561559	G	2/184	15/280	**0.017**	0.068	2/94	0.49	0.93	17/374	**0.034**	0.13
AID close[i]	rs714629	C	73/178	142/274	**0.024**	0.086	49/90	**0.037**	0.11	191/364	**0.012**	**0.044**
BLM[h]	rs7184015	G	138/184	207/278	0.90	1.00	76/96	0.44	1.00	283/374	0.86	1.00
BLM[h]	rs8034371	T	40/184	65/276	0.65	1.00	24/96	0.54	1.00	89/372	0.57	1.00
BLM[h]	rs6496724	G	50/182	86/280	0.46	1.00	29/94	0.56	1.00	115/374	0.43	1.00
BLM[h]	rs7183841	C	37/178	59/280	0.94	1.00	20/96	0.99	1.00	79/376	0.95	1.00
BLM[h]	rs6496725	C	38/182	67/276	0.40	1.00	21/96	0.85	1.00	88/372	0.46	1.00
BLM[h]	rs3784782	C	142/182	227/280	0.42	1.00	75/96	0.98	1.00	302/376	0.53	1.00
BLM[h]	rs2518968	C	99/176	161/274	0.60	1.00	59/96	0.41	1.00	220/370	0.48	1.00
BLM[h]	rs3784780	G	148/184	229/280	0.72	1.00	75/94	0.90	1.00	304/374	0.81	1.00
BLM[h]	rs3815003	G	49/184	83/278	0.45	1.00	28/94	0.58	1.00	111/372	0.43	1.00
BLM[h]	rs8036601	C	112/180	175/272	0.65	1.00	58/94	0.93	1.00	233/366	0.74	1.00
BLM[h]	rs4932363	A	9/184	20/280	0.33	1.00	7/94	0.39	1.00	27/374	0.29	1.00
BLM[h]	rs7175811	C	106/184	171/278	0.40	1.00	57/96	0.78	1.00	228/374	0.45	1.00
BLM[h]	rs2229035	-	0/182	0/274	--	--	0/96	--	--	0/370	--	--
BLM[h]	rs7182287	C	129/168	218/276	0.59	1.00	71/94	0.82	1.00	289/370	0.73	1.00
BLM[h]	rs1801256	T	0/180	**1/276**	0.42	1.00	0/96	--	1.00	1/372	0.49	1.00
BLM[h]	rs2270132	C	73/182	114/276	0.80	1.00	36/96	0.67	1.00	150/372	0.96	1.00
BLM[h]	rs2073919	G	138/182	222/280	0.38	1.00	72/96	0.88	1.00	294/376	0.53	1.00
BLM[h]	rs7167216	A	12/184	26/278	0.28	1.00	5/94	0.69	1.00	31/372	0.45	1.00
BLM[h]	rs414634	C	131/182	208/278	0.50	1.00	68/94	0.95	1.00	276/372	0.58	1.00
DMC1_close[h]	rs1946990	A	114/184	185/280	0.37	1.00	65/96	0.34	1.00	250/376	0.29	1.00
DMC1[h]	rs5750616	C	45/184	76/278	0.49	1.00	29/96	0.30	1.00	105/374	0.36	1.00
DMC1[h]	rs1980455	A	114/182	184/280	0.50	1.00	66/96	0.31	1.00	250/376	0.37	1.00
DMC1[h]	rs1292811	A	8/184	14/280	0.75	1.00	6/92	0.44	1.00	20/372	0.60	1.00
DMC1[h]	rs2227914	-	0/182	0/278	--	--	0/96	--	--	0/374	--	--
DMC1[h]	rs5757130	G	110/180	184/278	0.27	1.00	66/96	0.21	1.00	250/374	0.19	1.00
DMC1[h]	rs4821792	G	176/184	269/280	0.82	1.00	91/96	0.75	1.00	360/376	0.96	1.00
DMC1[h]	rs5757133	G	115/182	192/280	0.23	1.00	72/96	**0.046**	0.83	264/376	0.10	0.97
DMC1[h]	rs5757135	G	110/182	185/280	0.22	1.00	65/96	0.23	1.00	250/376	0.16	1.00
DMC1[h]	rs5757141	C	115/178	194/278	0.25	1.00	72/94	**0.042**	0.80	266/372	0.10	0.97
DMC1 close[h]	rs8140617	C	121/184	193/274	0.29	1.00	75/96	**0.032**	0.72	268/370	0.11	0.98
DMC1 close[h]	rs1569492	G	110/182	179/280	0.45	1.00	65/96	0.23	1.00	244/376	0.31	1.00
ERCC1_close	rs1046282	G	32/182	67/278	0.10	nd	23/94	0.17	nd	90/372	0.08	nd
ERCC1	rs3212977	-	0/184	0/276	--	--	0/94	--	--	0/370	--	--
ERCC1	rs11615	C	61/184	107/280	0.27	nd	35/92	0.42	nd	142/372	0.25	nd
ERCC1_close	rs1319052	T	52/178	110/280	**0.028**	nd	39/96	0.06	nd	149/376	**0.017**	nd
EXO1_close	rs10802995	T	72/184	122/280	0.34	nd	40/96	0.68	nd	162/376	0.37	nd
EXO1	rs1635518	T	86/184	131/274	0.82	nd	44/94	0.99	nd	175/368	0.86	nd
EXO1	rs4149864	-	0/182	0/276	--	--	0/90	--	--	0/366	--	--
EXO1	rs1635515	T	70/180	111/276	0.78	nd	50/94	**0.023**	nd	161/370	0.30	nd
EXO1	rs2526700	G	92/182	142/280	0.97	nd	52/94	0.45	nd	194/374	0.77	nd
EXO1	rs4149910	-	0/184	0/278	--	--	0/96	--	--	0/374	--	--
EXO1	rs1776133	C	73/180	109/272	0.92	nd	48/94	0.10	nd	157/366	0.60	nd
EXO1	rs2526698	A	92/184	142/278	0.82	nd	52/96	0.51	nd	194/374	0.68	nd
EXO1	rs735943	A	76/182	111/276	0.74	nd	49/92	0.07	nd	160/368	0.70	nd
EXO1	rs851781	C	92/184	148/280	0.55	nd	53/96	0.41	nd	201/376	0.44	nd
EXO1	rs4149963	A	16/182	31/278	0.41	nd	2/96	**0.031**	nd	33/374	0.99	nd
EXO1	rs4149966	G	1/184	3/280	0.55	nd	0/96	0.47	nd	3/376	0.74	nd
EXO1	rs4149967	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--
EXO1	rs2526697	T	104/184	165/278	0.55	nd	47/96	0.23	nd	212/374	0.97	nd
EXO1	rs1047840	C	116/184	181/276	0.58	nd	70/94	0.06	nd	251/370	0.26	nd
EXO1	rs12122770	C	3/184	18/280	**0.015**	nd	3/94	0.40	nd	21/374	**0.029**	nd
EXO1	rs4149978	G	183/184	277/278	0.77	nd	94/94	0.47	nd	371/372	0.61	nd
EXO1	rs1776148	C	108/182	173/272	0.36	nd	61/96	0.50	nd	234/368	0.33	nd
EXO1	rs1635498	C	4/184	6/278	0.99	nd	4/96	0.34	nd	10/374	0.72	nd
EXO1	rs9350	A	24/184	50/280	0.17	nd	17/94	0.26	nd	67/374	0.14	nd
EXO1	rs4150018	G	66/184	127/280	**0.043**	nd	42/94	0.15	nd	169/374	**0.036**	nd
H2AFX	rs1784304	T	44/184	77/280	0.39	nd	23/96	0.99	nd	100/376	0.50	nd
H2AFX	rs7350	A	61/184	93/278	0.95	nd	37/96	0.37	nd	130/374	0.71	nd
H2AFX_close	rs643788	G	68/184	102/274	0.95	nd	40/94	0.37	nd	142/368	0.71	nd
H2AFX_close	rs649870	G	66/182	106/280	0.73	nd	44/96	0.12	nd	150/376	0.41	nd

Supplemental Table III-6, continued

Gene	SNP	Associated allele[a]	Controls Allele frequency	IgAD Allele frequency	IgAD p-value	IgAD Corrected p-value[b]	CVID Allele frequency	CVID p-value	CVID Corrected p-value[b]	IgAD & CVID combined Allele frequency	IgAD & CVID combined p-value	IgAD & CVID combined Corrected p-value[b]
LIG1	rs11668325	G	183/184	279/280	0.76	nd	96/96	0.47	nd	375/376	0.61	nd
LIG1	rs274883	G	24/184	44/280	0.43	nd	14/96	0.72	nd	58/376	0.45	nd
LIG1	rs156633	A	95/176	156/278	0.66	nd	49/94	0.77	nd	205/372	0.80	nd
LIG1	rs3731003	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--
LIG1	rs156641	C	106/180	163/278	0.96	nd	60/96	0.56	nd	223/374	0.87	nd
LIG1	rs4987068	C	178/184	274/280	0.46	nd	95/96	0.26	nd	369/376	0.30	nd
LIG1	rs2304136	C	9/182	20/278	0.33	nd	6/96	0.65	nd	26/374	0.36	nd
LIG1	rs2288878	G	99/182	158/278	0.61	nd	49/94	0.72	nd	207/372	0.78	nd
LIG1	rs3730933	A	182/184	279/280	0.34	nd	95/96	0.97	nd	374/376	0.46	nd
LIG1	rs3730913	A	25/184	45/280	0.46	nd	15/96	0.64	nd	60/376	0.46	nd
LIG1	rs3730911	-	0/182	0/278	--	--	0/96	--	--	0/374	--	--
LIG1	rs11879148	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--
LIG1	rs274860	A	9/184	16/278	0.69	nd	4/96	0.78	nd	20/374	0.82	nd
LIG1	rs12981963	A	0/184	**1/280**	0.42	nd	0/96	--	--	1/376	0.48	nd
LIG1	rs4987181	G	183/184	280/280	0.22	nd	96/96	0.47	nd	376/376	0.15	nd
LIG1_close	rs274869	C	85/182	131/280	0.99	nd	44/94	0.99	nd	175/374	0.98	nd
LIG1_close	rs274873	G	84/182	133/280	0.78	nd	47/94	0.54	nd	180/374	0.66	nd
MLH1_close	rs1055095	A	93/180	152/276	0.48	nd	57/94	0.16	nd	209/370	0.29	nd
MLH1	rs1800734	C	140/182	216/278	0.85	nd	85/96	**0.019**	nd	301/374	0.33	nd
MLH1	rs9852378	A	78/184	132/280	0.31	nd	47/96	0.29	nd	179/376	0.24	nd
MLH1	rs11541859	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--
MLH1	rs4647224	T	79/180	132/280	0.49	nd	46/94	0.43	nd	178/374	0.41	nd
MLH1	rs4647250	C	75/178	130/278	0.33	nd	45/94	0.36	nd	175/372	0.28	nd
MLH1	rs4647257	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--
MLH1	rs1558528	A	79/180	131/280	0.54	nd	48/96	0.33	nd	179/376	0.41	nd
MLH1	rs2286939	C	73/180	132/278	0.15	nd	47/96	0.18	nd	179/374	0.11	nd
MLH1	rs2286940	A	76/180	132/280	0.30	nd	48/96	0.22	nd	180/376	0.21	nd
MLH1	rs9311150	-	0/184	0/276	--	--	0/96	--	--	0/372	--	--
MLH1	rs2241031	T	83/184	134/280	0.56	nd	48/96	0.44	nd	182/376	0.46	nd
MLH1	rs1800149	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--
MLH3	rs175047	G	83/178	121/274	0.61	nd	53/92	0.09	nd	174/366	0.84	nd
MLH3	rs108622	T	97/184	159/280	0.39	nd	43/96	0.21	nd	202/376	0.82	nd
MLH3	rs175052	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--
MLH3	rs175057	A	88/184	125/280	0.50	nd	56/96	0.09	nd	181/376	0.94	nd
MLH3	rs2098252	G	95/184	154/278	0.43	nd	39/92	0.15	nd	193/370	0.91	nd
MLH3	rs175067	-	0/180	0/280	--	--	0/96	--	--	0/376	--	--
MLH3	rs735452	G	87/180	123/278	0.39	nd	58/96	0.06	nd	181/374	0.99	nd
MLH3	rs149220	A	0/184	**1/278**	0.42	nd	0/96	--	--	1/374	0.48	nd
MLH3	rs3742780	C	95/180	146/278	0.96	nd	59/94	0.11	nd	205/372	0.61	nd
MLH3	rs10136948	G	96/182	148/280	0.98	nd	60/96	0.12	nd	208/376	0.57	nd
MLH3	rs175081	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--
MLH3	rs175082	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--
MLH3_close	rs10142770	C	96/182	147/278	0.98	nd	58/94	0.16	nd	205/372	0.60	nd
MRE11[g]	rs2155209	G	61/178	94/276	0.96	1.00	36/94	0.51	1.00	130/370	0.84	1.00
MRE11[g]	rs1805362	-	0/182	0/280	--	--	0/96	--	--	0/376	--	--
MRE11[g]	rs661957	T	59/180	91/270	0.84	1.00	36/92	0.30	0.97	127/362	0.59	1.00
MRE11[g]	rs518276	G	114/184	181/280	0.56	1.00	66/94	0.17	0.90	247/374	0.34	0.99
MRE11[g]	rs682213	T	98/180	164/278	0.34	0.99	62/96	0.10	0.56	226/374	0.18	0.89
MRE11[g]	rs569143	C	104/182	174/280	0.28	0.98	65/96	0.09	0.48	239/376	0.14	0.66
MRE11[g]	rs654718	A	121/184	192/280	0.53	1.00	69/96	0.30	0.97	261/376	0.38	0.99
MRE11[g]	rs1805367	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--
MRE11[g]	rs1061945	-	0/184	0/278	--	--	0/96	--	--	0/374	--	--
MRE11[g]	rs497763	G	97/184	165/280	0.19	0.80	62/96	0.06	0.35	227/376	0.08	0.49
MRE11 close[g]	rs472344	T	95/180	168/280	0.13	0.66	60/92	**0.050**	0.31	228/372	0.06	0.37
MSH2[c,d]	rs12612908	A	0/184	**1/280**	0.42	0.42	0/96	--	--	1/376	0.48	1.00
MSH2[c,d]	rs1863332	T	160/180	258/278	0.15	0.15	85/96	0.93	1.00	343/374	0.28	0.98
MSH2[c,d]	rs4987188	T	1/182	7/278	0.11	0.11	0/94	0.47	1.00	7/372	0.22	0.97
MSH2[c,d]	rs4952887	T	14/184	31/280	0.22	0.22	18/96	**0.0054**	0.062	49/376	0.06	0.49
MSH2[c,d]	rs2347794	A	60/182	117/278	**0.049**	**0.049**	47/94	**0.0059**	0.070	164/372	**0.012**	0.17
MSH2[c,d]	rs3771274	T	62/184	118/278	0.06	0.059	49/96	**0.0049**	0.053	167/374	**0.013**	0.18
MSH2[c,d]	rs6726691	C	45/184	85/280	0.17	0.17	32/96	0.11	0.71	117/376	0.10	0.75
MSH2[c,d]	rs1981928	T	47/182	81/280	0.47	0.47	29/94	0.38	1.00	110/374	0.38	1.00
MSH2[c,d]	rs3771275	G	64/184	117/276	0.10	0.10	48/94	**0.0088**	0.098	165/370	**0.027**	0.30
MSH2[c,d]	rs3771276	A	59/182	116/276	**0.038**	**0.038**	48/96	**0.0042**	**0.0498**	164/372	**0.0085**	0.13
MSH2[c,d]	rs6729015	A	61/184	122/278	**0.021**	**0.021**	49/96	**0.0036**	**0.045**	171/374	**0.0046**	0.061
MSH2[c,d]	rs3771281	A	56/184	116/280	**0.016**	**0.016**	43/96	**0.017**	0.21	159/376	**0.0068**	0.079
MSH2[c,d]	rs3764959	G	47/178	79/274	0.57	0.57	30/96	0.39	1.00	109/370	0.46	1.00
MSH2[c,d]	rs2042649	G	14/182	25/278	0.62	0.62	7/96	0.90	1.00	32/374	0.73	1.00
MSH2[c,d]	rs1802577	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--

Supplemental Table III-6, continued

Gene	SNP	Associated allele[a]	Controls Allele frequency	IgAD Allele frequency	IgAD p-value	IgAD Corrected p-value[b]	CVID Allele frequency	CVID p-value	CVID Corrected p-value[b]	IgAD & CVID combined Allele frequency	IgAD & CVID combined p-value	IgAD & CVID combined Corrected p-value[b]
MSH3 close[d]	rs844370	G	39/182	73/280	0.26	1.00	18/94	0.66	1.00	91/374	0.45	1.00
MSH3[d]	rs380691	G	58/184	91/280	0.83	1.00	35/96	0.41	1.00	126/376	0.64	1.00
MSH3[d]	rs1650670	C	40/178	75/276	0.26	1.00	19/96	0.61	1.00	94/372	0.47	1.00
MSH3[d]	rs1650666	A	40/180	75/278	0.25	1.00	19/96	0.64	1.00	94/374	0.45	1.00
MSH3[d]	rs1650663	C	42/180	74/276	0.40	1.00	17/94	0.32	1.00	91/370	0.75	1.00
MSH3[d]	rs1677649	T	44/180	74/276	0.57	1.00	19/94	0.43	1.00	93/370	0.86	1.00
MSH3[d]	rs6864493	C	39/184	72/280	0.26	0.72	34/96	**0.010**	0.17	106/376	0.08	0.72
MSH3[d]	rs3776968	T	58/184	108/278	0.11	0.71	39/96	0.13	0.89	147/374	0.07	0.71
MSH3[d]	rs26282	G	136/184	203/278	0.83	1.00	77/96	0.24	0.99	280/374	0.81	1.00
MSH3[d]	rs33013	A	53/182	88/280	0.60	1.00	35/96	0.21	0.98	123/376	0.39	1.00
MSH3[d]	rs6151838	T	165/184	254/280	0.71	1.00	84/94	0.94	1.00	338/374	0.79	1.00
MSH3[d]	rs42290	T	26/184	46/280	0.50	1.00	10/96	0.38	1.00	56/376	0.81	1.00
MSH3[d]	rs3797896	G	164/180	263/280	0.25	0.83	93/96	0.07	0.68	356/376	0.11	0.83
MSH3[d]	rs245391	G	43/182	80/280	0.24	0.58	38/94	**0.0037**	0.070	118/374	0.05	0.58
MSH3[d]	rs32991	T	99/184	152/276	0.79	1.00	63/96	0.06	0.62	215/372	0.37	1.00
MSH3[d]	rs26910	T	22/178	38/270	0.60	0.96	23/94	**0.011**	0.18	61/364	0.18	0.96
MSH3[d]	rs3776978	A	21/184	38/280	0.49	1.00	13/96	0.60	1.00	51/376	0.48	1.00
MSH3[d]	rs184967	G	155/184	235/280	0.93	1.00	86/96	0.22	0.99	321/376	0.72	1.00
MSH3[d]	rs245341	A	42/178	79/278	0.26	1.00	20/96	0.60	1.00	99/374	0.47	1.00
MSH3[d]	rs27887	T	46/182	80/280	0.44	1.00	20/94	0.46	1.00	100/374	0.71	1.00
MSH3 close[d]	rs32964	C	47/184	79/280	0.53	1.00	20/96	0.38	1.00	99/376	0.84	1.00
MSH4 close[e]	rs1498311	C	132/184	204/280	0.79	1.00	72/96	0.56	1.00	276/376	0.68	1.00
MSH4[e]	rs1146646	T	40/178	72/280	0.43	1.00	29/94	0.13	0.92	101/374	0.25	0.99
MSH4[e]	rs5745325	C	128/182	203/280	0.61	1.00	71/96	0.52	1.00	274/376	0.53	1.00
MSH4[e]	rs5745327	A	163/182	255/278	0.43	1.00	87/96	0.78	1.00	342/374	0.47	1.00
MSH4[e]	rs5745329	T	0/180	**1/278**	0.42	1.00	**1/96**	0.17	0.97	**2/374**	0.33	1.00
MSH4[e]	rs1146652	G	170/184	262/280	0.62	1.00	91/96	0.45	1.00	353/376	0.50	1.00
MSH4[e]	rs1144342	C	163/182	257/280	0.42	1.00	87/96	0.78	1.00	344/376	0.46	1.00
MSH4[e]	rs5745390	G	38/182	67/280	0.44	1.00	26/96	0.24	1.00	93/376	0.31	1.00
MSH4[e]	rs1565717	A	38/184	64/278	0.55	1.00	24/94	0.36	1.00	88/372	0.43	1.00
MSH4[e]	rs5745458	A	142/182	221/278	0.70	1.00	74/96	0.86	1.00	295/374	0.82	1.00
MSH4[e]	rs5745459	A	177/184	274/280	0.29	1.00	93/94	0.20	0.99	367/374	0.17	0.97
MSH4[e]	rs1001160	A	131/182	201/278	0.94	1.00	69/94	0.80	1.00	270/372	0.88	1.00
MSH4[e]	rs5745543	A	127/178	196/274	0.97	1.00	67/92	0.80	1.00	263/366	0.90	1.00
MSH4[e]	rs5745545	A	131/184	201/280	0.89	1.00	70/96	0.76	1.00	271/376	0.83	1.00
MSH4[e]	rs5745549	A	6/180	14/280	0.39	1.00	5/94	0.43	1.00	19/374	0.35	1.00
MSH4 close[e]	rs946163	G	132/184	203/280	0.86	1.00	70/94	0.63	1.00	273/374	0.75	1.00
MSH5 close[e]	rs805304	G	68/182	170/280	**9.2E-07**	**0E-00**[f]	37/94	0.75	1.00	207/374	**6.9E-05**	**0.00**
MSH5 close[e]	rs400547	A	10/184	24/278	0.20	0.97	6/96	0.78	1.00	30/374	0.27	1.00
MSH5[e]	rs3131382	G	172/184	271/280	0.09	0.81	90/96	0.93	1.00	361/376	0.19	0.98
MSH5[e]	rs409558	G	30/184	56/280	0.32	1.00	17/96	0.77	1.00	73/376	0.37	1.00
MSH5[e]	rs28381349	T	4/184	9/278	0.50	1.00	3/96	0.63	1.00	12/374	0.49	1.00
MSH5[e]	rs707915	T	10/184	24/280	0.20	0.97	5/96	0.94	1.00	29/376	0.32	1.00
MSH5[e]	rs2075788	G	30/184	56/280	0.32	1.00	17/96	0.77	1.00	73/376	0.37	1.00
MSH5[e]	rs3117572	G	142/182	251/278	**2.7E-04**	**0.01**	76/96	0.83	1.00	327/374	**0.0042**	**0.09**
MSH5[e]	rs2299851	T	20/184	31/280	0.95	1.00	10/96	0.91	1.00	41/376	0.99	1.00
MSH5[e]	rs3131379	A	27/182	87/278	**6.4E-05**	**0.00**	14/92	0.93	1.00	101/370	**0.0011**	**0.02**
MSH5[e]	rs3131378	G	26/182	89/280	**2.1E-05**	**0.00**	18/96	0.33	1.00	107/376	**2.3E-04**	**0.01**
MSH5[e]	rs707939	C	129/180	226/280	**0.024**	**0.38**	66/96	0.61	1.00	292/376	0.12	0.90
MSH5[e]	rs3117577	C	26/180	88/278	**3.2E-05**	**0.00**	15/96	0.79	1.00	103/374	**6.4E-04**	**0.02**
MSH5[e]	rs707938	C	63/180	154/276	**1.4E-05**	**0.00**	35/96	0.81	1.00	189/372	**4.7E-04**	**0.01**
MSH5[e]	rs1802127	C	180/184	267/274	0.79	1.00	95/96	0.50	1.00	362/370	0.99	1.00
MSH5_close[e]	rs707937	G	147/182	232/278	0.46	1.00	83/96	0.23	1.00	315/374	0.31	1.00
MSH5_close[e]	rs707936	T	12/184	25/280	0.35	1.00	7/96	0.81	1.00	32/376	0.41	1.00
MSH6[c]	rs3136245	T	39/184	71/280	0.30	0.98	18/94	0.69	1.00	89/374	0.49	1.00
MSH6[c]	rs1878484	A	182/184	277/280	0.99	1.00	93/94	0.99	1.00	370/374	0.99	1.00
MSH6[c]	rs1800932	C	26/180	45/280	0.64	1.00	18/96	0.35	1.00	63/376	0.49	1.00
MSH6[c]	rs3211299	-	0/178	0/274	--	--	0/96	--	--	0/370	--	--
MSH6[c]	rs1800938	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--
MSH6[c]	rs3136334	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--
MSH6[c]	rs2020912	A	182/184	275/278	0.99	1.00	96/96	0.31	0.99	371/374	0.74	1.00
MSH6[c]	rs2020911	T	63/184	114/278	0.14	0.85	37/96	0.48	1.00	151/374	0.16	0.91
MUS81_close	rs635375	G	47/178	91/280	0.17	nd	28/94	0.55	nd	119/374	0.19	nd
MUS81_close	rs652021	G	49/182	91/280	0.20	nd	31/96	0.35	nd	122/376	0.18	nd
MUS81_close	rs665306	T	58/174	99/268	0.44	nd	33/96	0.86	nd	132/364	0.51	nd
MUS81	rs630303	C	54/184	91/280	0.47	nd	28/94	0.94	nd	119/374	0.55	nd
MUS81	rs765593	-	0/174	0/274	--	--	0/96	--	--	0/370	--	--
MUS81_close	rs659824	A	81/178	135/274	0.43	nd	43/90	0.72	nd	178/364	0.46	nd
MUS81_close	rs630394	C	57/182	93/276	0.60	nd	28/94	0.79	nd	121/370	0.74	nd

Supplemental Table III-6, continued

Gene	SNP	Associated allele[a]	Controls Allele frequency	IgAD Allele frequency	IgAD p-value	IgAD Corrected p-value[b]	CVID Allele frequency	CVID p-value	CVID Corrected p-value[b]	IgAD & CVID combined Allele frequency	IgAD & CVID combined p-value	IgAD & CVID combined Corrected p-value[b]
NBS1[g]	rs3026268	-	0/182	0/280	--	--	0/96	--	--	0/376	--	--
NBS1[g]	rs769420	-	0/184	0/280	--	--	0/92	--	--	0/372	--	--
NBS1[g]	rs1805794	C	112/178	183/280	0.60	1.00	61/92	0.58	1.00	244/372	0.54	1.00
PMS2_close	rs710939	C	3/184	8/280	0.40	nd	2/96	0.79	nd	10/376	0.45	nd
PMS2_close	rs852520	A	95/184	168/278	0.06	nd	60/96	0.08	nd	228/374	**0.036**	nd
PMS2_close	rs852516	T	11/184	25/280	0.25	nd	11/96	0.11	nd	36/376	0.15	nd
PMS2_close	rs852417	G	118/184	192/278	0.27	nd	62/96	0.94	nd	254/374	0.37	nd
PMS2_close	rs852413	A	3/184	9/280	0.29	nd	2/96	0.79	nd	11/376	0.36	nd
PMS2_close	rs3779092	G	87/178	151/278	0.26	nd	46/94	0.99	nd	197/372	0.37	nd
PMS2_close	rs852394	T	79/176	127/276	0.81	nd	43/94	0.89	nd	170/370	0.82	nd
PMS2_close	rs4724712	A	84/184	128/280	0.99	nd	46/94	0.60	nd	174/374	0.85	nd
PMS2_close	rs852266	C	10/182	25/280	0.17	nd	10/96	0.13	nd	35/376	0.12	nd
PMS2_close	rs852262	G	2/184	8/278	0.20	nd	3/96	0.22	nd	11/374	0.17	nd
PMS2_close	rs1468996	T	67/180	113/276	0.43	nd	37/94	0.73	nd	150/370	0.45	nd
PMS2	rs1805321	T	0/184	**5/280**	0.068	nd	**1/96**	0.17	nd	**6/376**	0.085	nd
PMS2_close	rs2009115	A	14/182	42/278	**0.017**	nd	12/96	0.19	nd	54/374	**0.023**	nd
PMS2_close	rs3779107	G	14/182	44/280	**0.011**	nd	12/96	0.19	nd	56/376	**0.016**	nd
PMS2_close	rs1860459	A	18/184	25/280	0.76	nd	13/96	0.34	nd	38/376	0.90	nd
POLH_close	rs699937	T	52/182	86/274	0.52	nd	25/94	0.73	nd	111/368	0.70	nd
POLH	rs6458343	A	10/182	12/278	0.56	nd	9/96	0.22	nd	21/374	0.95	nd
POLH	rs2307456	G	183/184	280/280	0.22	nd	96/96	0.47	nd	376/376	0.15	nd
POLH	rs9333548	-	0/184	0/278	--	--	0/94	--	--	0/372	--	--
RAD50 close[g]	rs2522410	A	0/184	**2/280**	0.25	0.98	**1/96**	0.17	0.88	**3/376**	0.22	0.96
RAD50[g]	rs4526098	C	0/184	**1/280**	0.42	1.00	0/96	**--**	--	1/376	0.48	1.00
RAD50[g]	rs2244012	T	138/184	222/280	0.28	0.98	74/96	0.70	1.00	296/376	0.32	0.99
RAD50[g]	rs2706348	G	137/182	217/272	0.26	0.98	74/96	0.74	1.00	291/368	0.31	0.98
RAD50[g]	rs2230017	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--
RAD50[g]	rs2252775	A	137/182	219/278	0.38	0.99	74/96	0.74	1.00	293/374	0.42	1.00
RAD50[g]	rs1047380	-	0/184	0/276	--	--	0/94	--	--	0/370	--	--
RAD50[g]	rs1047382	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--
RAD50[g]	rs3187395	G	183/184	278/278	0.22	0.93	96/96	0.47	1.00	374/374	0.15	0.86
RAD50[g]	rs1047386	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--
RAD50[g]	rs1047387	-	0/184	0/278	--	--	0/96	--	--	0/374	--	--
RAD50[g]	rs2237060	G	59/184	117/280	**0.035**	0.23	48/96	**0.0034**	**0.029**	165/376	**0.0073**	0.057
RAD50[g]	rs2240032	C	138/184	221/278	0.26	0.98	72/94	0.77	1.00	293/372	0.32	0.99
RAD50 close[g]	rs2158177	A	141/184	227/280	0.25	0.95	73/94	0.85	1.00	300/374	0.33	0.99
RAD51 close[h]	rs2412545	G	109/184	185/280	0.14	0.99	56/96	0.88	1.00	241/376	0.26	1.00
RAD51[h]	rs2619681	T	25/184	47/280	0.35	1.00	13/96	0.99	1.00	60/376	0.46	1.00
RAD51[h]	rs7174493	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--
RAD51[h]	rs2412546	G	83/182	133/278	0.64	1.00	39/96	0.43	1.00	172/374	0.93	1.00
RAD51[h]	rs11858337	G	85/182	137/276	0.54	1.00	40/94	0.51	1.00	177/370	0.80	1.00
RAD51[h]	rs957603	T	72/182	119/276	0.45	1.00	35/94	0.71	1.00	154/370	0.64	1.00
RAD51[h]	rs11070291	A	83/184	138/280	0.38	1.00	41/96	0.70	1.00	179/376	0.58	1.00
RAD51[h]	rs1056742	-	0/184	0/278	--	--	0/96	--	--	0/374	--	--
RAD51_close[h]	rs4924501	G	84/184	140/278	0.32	1.00	41/94	0.75	1.00	181/372	0.50	1.00
RAD51 close[h]	rs11558809	C	179/180	280/280	0.21	1.00	96/96	0.46	1.00	376/376	0.15	1.00
RAD52[h]	rs1060499	G	29/182	48/276	0.68	1.00	25/94	**0.034**	0.74	73/370	0.28	1.00
RAD52[h]	rs1051669	G	144/180	223/280	0.93	1.00	82/96	0.27	1.00	305/376	0.75	1.00
RAD52[h]	rs10744729	G	83/182	138/276	0.36	1.00	54/94	0.06	0.91	192/370	0.16	1.00
RAD52[h]	rs4766370	A	82/182	137/280	0.42	1.00	54/96	0.08	0.94	191/376	0.20	1.00
RAD52[h]	rs9634161	C	28/184	54/280	0.26	1.00	27/96	**0.010**	0.34	81/376	0.08	0.93
RAD52[h]	rs7312883	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--
RAD52[h]	rs7487683	T	3/184	13/280	0.08	0.95	6/96	**0.038**	0.77	19/376	0.05	0.85
RAD52[h]	rs4766377	A	144/182	223/280	0.89	1.00	82/96	0.20	1.00	305/376	0.58	1.00
RAD52[h]	rs1131839	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--
RAD52[h]	rs1833095	C	48/184	75/280	0.87	1.00	25/96	0.99	1.00	100/376	0.90	1.00
RAD52[h]	rs7311151	G	77/184	116/280	0.93	1.00	43/96	0.64	1.00	159/376	0.92	1.00
RAD52[h]	rs2887531	A	134/184	204/280	0.99	1.00	72/96	0.70	1.00	276/376	0.88	1.00
RAD52 close[h]	rs10849605	A	75/178	123/276	0.61	1.00	59/94	**0.0012**	0.060	182/370	0.12	0.99
RAD54B[h]	rs2046666	C	71/184	143/280	**0.0083**	0.27	46/96	0.13	0.99	189/376	**0.0092**	0.32
RAD54B[h]	rs2470740	T	72/180	144/276	**0.0109**	0.36	46/96	0.21	1.00	190/372	**0.015**	0.44
RAD54B[h]	rs2046663	A	16/184	32/278	0.33	1.00	11/96	0.46	1.00	43/374	0.31	1.00
RAD54B[h]	rs3019149	A	47/184	97/280	**0.038**	0.76	34/96	0.08	0.96	131/376	**0.026**	0.64
RAD54B[h]	rs3136421	T	10/184	21/280	0.38	1.00	7/96	0.54	1.00	28/376	0.37	1.00
RAD54B[h]	rs3019279	G	63/184	134/276	**0.0024**	0.094	47/94	**0.011**	0.36	181/370	**0.0010**	**0.046**
RAD54B[h]	rs2921385	C	71/182	145/276	**0.0045**	0.17	48/96	0.08	0.95	193/372	**0.0044**	0.16
RAD54B[h]	rs2930968	A	64/184	132/280	**0.0084**	0.28	47/94	**0.014**	0.44	179/374	**0.0034**	0.14
RAD54B[h]	rs1372048	G	49/184	95/278	0.09	0.96	35/96	0.09	0.96	130/374	0.05	0.86
RAD54B[h]	rs2919661	-	0/184	0/278	--	--	0/92	--	1.00	0/370	--	--
RAD54B close[h]	rs1992371	T	66/178	131/280	**0.041**	0.78	42/96	0.28	1.00	173/376	**0.047**	0.82

Supplemental Table III-6, continued

Gene	SNP	Associated allele[a]	Controls Allele frequency	IgAD Allele frequency	IgAD p-value	IgAD Corrected p-value[b]	CVID Allele frequency	CVID p-value	CVID Corrected p-value[b]	IgAD & CVID combined Allele frequency	IgAD & CVID combined p-value	IgAD & CVID combined Corrected p-value[b]
SPO11	rs3736832	-	0/182	0/280	--	--	0/94	--	--	0/374	--	--
SPO11	rs6099553	G	112/184	175/280	0.72	nd	67/94	0.09	nd	242/374	0.38	nd
SPO11	rs1467581	T	2/180	8/274	0.20	nd	6/96	**0.015**	nd	14/370	0.08	nd
TP53BP1	rs1058298	T	53/180	86/280	0.77	nd	28/94	0.95	nd	114/374	0.80	nd
TP53BP1	rs11554564	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--
TP53BP1	rs2230449	-	0/184	0/280	--	--	0/96	--	--	0/376	--	--
TP53BP1	rs2242069	C	29/184	55/280	0.29	nd	17/96	0.68	nd	72/376	0.33	nd
TP53BP1	rs542898	A	56/184	88/280	0.82	nd	28/96	0.83	nd	116/376	0.92	nd
TP53BP1	rs2602141	T	128/184	194/280	0.95	nd	68/96	0.83	nd	262/376	0.98	nd
TP53BP1	rs536313	T	50/176	85/280	0.66	nd	28/96	0.89	nd	113/376	0.69	nd
TP53BP1	rs2256238	-	0/182	0/276	--	--	0/96	--	--	0/372	--	--
TP53BP1	rs2467739	A	13/180	37/280	**0.044**	nd	10/96	0.36	nd	47/376	0.06	nd
TP53BP1	rs2467741	C	56/182	93/280	0.58	nd	29/96	0.92	nd	122/376	0.69	nd
TP53BP1	rs694725	C	54/182	84/278	0.90	nd	27/96	0.79	nd	111/374	1.00	nd
TP53BP1	rs689647	T	15/184	34/280	0.17	nd	10/96	0.53	nd	44/376	0.20	nd
TP53BP1	rs560191	G	58/184	92/280	0.76	nd	30/94	0.95	nd	122/374	0.79	nd
TP53BP1	rs2439850	G	15/182	34/278	0.18	nd	8/94	0.94	nd	42/372	0.27	nd
TP53BP1_close	rs1869258	G	54/184	87/280	0.69	nd	28/96	0.97	nd	115/376	0.76	nd
TP53BP1_close	rs523156	G	131/184	198/278	0.99	nd	70/94	0.56	nd	268/372	0.83	nd
XRCC2_close	rs6962238	G	8/184	25/280	0.06	nd	8/96	0.17	nd	33/376	0.06	nd
XRCC2	rs3218536	T	10/180	18/280	0.70	nd	6/96	0.81	nd	24/376	0.70	nd
XRCC2	rs3111471	T	88/184	149/280	0.26	nd	52/96	0.31	nd	201/376	0.21	nd
XRCC2	rs3218408	G	44/184	72/278	0.63	nd	24/96	0.84	nd	96/374	0.65	nd
XRCC2_close	rs2040639	C	85/182	149/280	0.17	nd	50/94	0.31	nd	199/374	0.15	nd
XRCC2_close	rs6464268	T	162/184	250/280	0.68	nd	82/96	0.53	nd	332/376	0.93	nd
XRCC2_close	rs10227264	G	34/182	55/280	0.80	nd	28/94	**0.036**	nd	83/374	0.34	nd
XRCC2_close	rs684088	A	109/184	162/280	0.77	nd	61/96	0.48	nd	223/376	0.99	nd
XRCC2_close	rs513586	C	171/184	266/278	0.20	nd	87/96	0.50	nd	353/374	0.50	nd
XRCC2_close	rs2018083	T	121/182	189/276	0.66	nd	64/88	0.30	nd	253/364	0.47	nd
XRCC3_close	rs2273175	G	67/184	109/278	0.54	nd	31/94	0.57	nd	140/372	0.78	nd
XRCC3_close	rs941474	A	87/176	136/274	0.97	nd	49/96	0.80	nd	185/370	0.90	nd

a) as determined in the combined IgAD/CVID cohort
b) p-value corrected for multiple testing bias by 100,000 permutations of all markers within pathways as described in Materials and Methods
c) Permutation tests performed on markers of the MutSα complex - p-values reported for MSH2 are from this permutation set
d) Permutation tests performed on markers of the MutSβ complex
e) Permutation tests performed on markers of the MutSγ complex
f) no chi-squared permutations exceeded those observed for SNP rs805304
g) Permutation tests performed on markers of the MRN complex
h) Permutation tests performed on markers of the RAD52 extended epistasis group
i) Permutation tests performed on markers in AID

Supplementary Table III-7. Haplotype block structure obtained from control genotypes.

Gene	Haplotype block	SNPs in block					
BLM	1	rs8034371	rs6496724	rs7183841			
BLM	2	rs3815003	rs8036601	rs7175811			
BLM	3	rs7182287	rs2270132	rs2073919			
DMC1	4	rs1946990	rs5750616	rs1980455			
DMC1	5	rs5757130	rs5757133	rs5757135	rs5757141	rs8140617	rs1569492
ERCC1	6	rs11615	rs1319052				
EXO1	7	rs1635515	rs2526700	rs1776133	rs2526698	rs735943	rs851781
H2AFX	8	rs643788	rs649870				
LIG1	9	rs156633	rs156641	rs2288878			
LIG1	10	rs274869	rs274873				
MLH1	11	rs9852378 rs2241031	rs4647224	rs4647250	rs1558528	rs2286939	rs2286940
MLH3	12	rs175047 rs10136948	rs108622 rs10142770	rs175057	rs2098252	rs735452	rs3742780
MRE11	13	rs2155209	rs661957				
MRE11	14	rs682213	rs569143				
MRE11	15	rs497763	rs472344				
MSH2	16	rs2347794 rs6729015	rs3771274 rs3771281	rs6726691	rs1981928	rs3771275	rs3771276
MSH3	17	rs1650670 rs26282	rs1650666 rs33013	rs1650663 rs6151838	rs1677649	rs6864493	rs3776968
MSH3	18	rs245391	rs32991	rs26910			
MSH3	19	rs245341	rs27887	rs32964			
MSH4	20	rs5745325 rs5745458	rs5745327 rs1001160	rs1146652 rs5745543	rs1144342 rs5745545	rs5745390 rs946163	rs1565717
MSH5	21	rs2075788 rs707938	rs3117572	rs2299851	rs3131379	rs3131378	rs3117577
MUS81	22	rs635375	rs652021	rs665306	rs630303	rs659824	
RAD50	23	rs2244012	rs2706348	rs2252775	rs2237060	rs2240032	rs2158177
RAD51	24	rs2412545 rs4924501	rs2619681	rs2412546	rs11858337	rs957603	rs11070291
RAD52	25	rs10744729	rs4766370	rs9634161	rs4766377	rs1833095	
RAD54B	26	rs2046666 rs2921385	rs2470740	rs2046663	rs3019149	rs3136421	rs3019279
TP53BP1	27	rs542898 rs689647	rs2602141 rs560191	rs536313 rs2439850	rs2467739 rs1869258	rs2467741 rs523156	rs694725
XRCC2	28	rs3111471	rs3218408				

Supplemental Table III-8. Haplotype association for SNPs genotyped in IgAD and CVID. Significant p-values (<0.05) are highlighted in bold.

	IgAD				CVID				Combined CVID & IgAD			
	Hap freq.	Case, control frequencies	p-value	Corrected p-value [a]	Hap freq.	Case, control frequencies	p-value	Corrected p-value [a]	Hap freq.	Case, control frequencies	p-value	Corrected p-value [a]
BLM [g]												
Block 1												
ATT	0.702	0.693, 0.716	0.59	1.00	0.705	0.685, 0.716	0.59	1.00	0.699	0.691, 0.716	0.54	1.00
TGC	0.206	0.211, 0.200	0.78	1.00	0.203	0.208, 0.200	0.87	1.00	0.207	0.210, 0.200	0.78	1.00
AGT	0.065	0.071, 0.055	0.49	1.00	0.059	0.065, 0.055	0.76	1.00	0.065	0.070, 0.055	0.52	1.00
TGT	0.022	0.025, 0.017	0.58	1.00	0.021	0.031, 0.016	0.43	1.00	0.023	0.026, 0.017	0.48	1.00
TTT	--	--	--	--	0.005	0.011, 0.001	0.24	1.00	0.002	0.003, 0.001	0.57	1.00
ATC	0.002	0.000, 0.006	0.20	1.00	0.004	0.000, 0.006	0.45	1.00	0.002	0.000, 0.006	0.14	1.00
AGC	0.002	0.000, 0.005	0.25	1.00	0.003	0.000, 0.005	0.49	1.00	0.002	0.000, 0.005	0.18	1.00
BLM [g]												
Block 2												
ACC	0.575	0.583, 0.563	0.68	1.00	0.571	0.583, 0.564	0.77	1.00	0.576	0.583, 0.564	0.67	1.00
GTT	0.262	0.268, 0.253	0.71	1.00	0.267	0.291, 0.254	0.51	1.00	0.267	0.274, 0.253	0.60	1.00
ATT	0.091	0.073, 0.119	0.09	0.97	0.105	0.081, 0.118	0.35	1.00	0.089	0.075, 0.118	0.09	0.97
ACT	0.038	0.033, 0.046	0.50	1.00	0.041	0.033, 0.046	0.60	1.00	0.037	0.033, 0.046	0.46	1.00
GCC	0.014	0.020, 0.007	0.25	1.00	0.004	0.000, 0.006	0.50	1.00	0.012	0.015, 0.006	0.40	1.00
ATC	0.009	0.012, 0.006	0.53	1.00	0.008	0.011, 0.006	0.66	1.00	0.01	0.011, 0.006	0.54	1.00
GCT	0.008	0.008, 0.007	0.88	1.00	0.005	0.002, 0.006	0.57	1.00	0.007	0.006, 0.007	0.98	1.00
GTC	0.002	0.003, 0.000	0.45	1.00	--	--	--	--	0.002	0.003, 0.000	0.50	1.00
BLM [g]												
Block 3												
CCG	0.409	0.415, 0.400	0.75	1.00	0.386	0.363, 0.398	0.57	1.00	0.4	0.401, 0.399	0.96	1.00
CAG	0.365	0.375, 0.350	0.59	1.00	0.364	0.387, 0.352	0.56	1.00	0.37	0.379, 0.351	0.53	1.00
TAA	0.224	0.207, 0.250	0.28	1.00	0.244	0.238, 0.248	0.85	1.00	0.226	0.214, 0.249	0.36	1.00
TCA	--	--	--	--	0.006	0.012, 0.002	0.26	1.00	0.003	0.004, 0.001	0.58	1.00
TAG	0.002	0.004, 0.000	0.47	1.00	--	--	--	--	0.002	0.003, 0.000	0.53	1.00
DMC1 [g]												
Block 4												
AAA	0.375	0.379, 0.369	0.83	1.00	0.373	0.380, 0.369	0.87	1.00	0.376	0.379, 0.369	0.82	1.00
GAC	0.347	0.332, 0.369	0.41	1.00	0.346	0.302, 0.369	0.26	1.00	0.339	0.324, 0.369	0.29	1.00
ACA	0.26	0.271, 0.245	0.53	1.00	0.259	0.287, 0.245	0.44	1.00	0.265	0.275, 0.245	0.45	1.00
AAC	0.009	0.011, 0.006	0.56	1.00	0.007	0.011, 0.006	0.65	1.00	0.009	0.011, 0.006	0.55	1.00
GAA	0.007	0.004, 0.011	0.33	1.00	0.009	0.006, 0.011	0.66	1.00	0.006	0.004, 0.011	0.33	1.00
GCA	0.002	0.004, 0.000	0.42	1.00	0.005	0.015, 0.000	0.10	0.97	0.005	0.007, 0.000	0.27	1.00
DMC1 [g]												
Block 5												
GGGCCG	0.601	0.625, 0.565	0.20	1.00	0.596	0.655, 0.565	0.14	1.00	0.61	0.633, 0.565	0.12	1.00
AAATTA	0.308	0.296, 0.326	0.50	1.00	0.282	0.198, 0.325	**0.024**	0.47	0.289	0.271, 0.326	0.18	1.00
AGACCA	0.036	0.036, 0.038	0.91	1.00	0.046	0.063, 0.038	0.33	1.00	0.041	0.043, 0.038	0.79	1.00
GAGCCG	0.017	0.014, 0.022	0.54	1.00	0.018	0.011, 0.022	0.49	1.00	0.016	0.013, 0.022	0.45	1.00
GGGCCA	0.017	0.022, 0.011	0.40	1.00	0.011	0.010, 0.011	0.97	1.00	0.016	0.019, 0.011	0.50	1.00
AAATCA	0.007	0.004, 0.011	0.36	1.00	0.014	0.021, 0.011	0.51	1.00	0.009	0.008, 0.011	0.75	1.00
AAACTA	--	--	--	--	0.007	0.021, 0.000	0.060	0.89	0.004	0.005, 0.000	0.34	1.00
AGATCA	0.002	0.004, 0.000	0.43	1.00	0.003	0.010, 0.000	0.19	1.00	0.004	0.005, 0.000	0.33	1.00
GGATCG	--	--	--	--	0.004	0.010, 0.000	0.17	1.00	0.002	0.003, 0.000	0.48	1.00
AAATTG	0.002	0.000, 0.005	0.22	1.00	0.004	0.000, 0.005	0.47	1.00	0.002	0.000, 0.005	0.16	1.00
AGACCG	0.002	0.000, 0.006	0.21	1.00	0.004	0.000, 0.006	0.46	1.00	0.002	0.000, 0.006	0.15	1.00
GAATTA	0.002	0.000, 0.005	0.22	1.00	0.004	0.000, 0.006	0.47	1.00	0.002	0.000, 0.005	0.15	1.00
GGGTCG	0.002	0.000, 0.005	0.22	1.00	0.004	0.001, 0.006	0.56	1.00	0.002	0.000, 0.005	0.17	1.00
AGATTA	0.002	0.000, 0.006	0.23	1.00	0.002	0.000, 0.004	0.56	1.00	0.002	0.000, 0.006	0.16	1.00
GGATTA	--	--	--	--	0.001	0.000, 0.002	0.67	1.00	--	--	--	--
ERCC1												
Block 6												
TC	0.616	0.589, 0.657	0.14	nd	0.632	0.583, 0.657	0.22	nd	0.61	0.587, 0.657	0.11	nd
CT	0.337	0.364, 0.297	0.13	nd	0.326	0.384, 0.296	0.14	nd	0.345	0.369, 0.297	0.093	nd
CC	0.025	0.018, 0.035	0.26	nd	0.027	0.011, 0.035	0.24	nd	0.023	0.017, 0.035	0.17	nd
TT	0.022	0.029, 0.012	0.21	nd	0.015	0.023, 0.011	0.47	nd	0.022	0.028, 0.012	0.23	nd

Supplemental Table III-8, continued.

	IgAD				CVID				Combined CVID & IgAD			
	Hap freq.	Case, control frequencies	p-value	Corrected p-value [a]	Hap freq.	Case, control frequencies	p-value	Corrected p-value [a]	Hap freq.	Case, control frequencies	p-value	Corrected p-value [a]
EXO1												
Block 7												
CTTGGT	0.476	0.468, 0.489	0.65	nd	0.464	0.417, 0.489	0.25	nd	0.468	0.457, 0.489	0.47	nd
TGCAAC	0.394	0.396, 0.391	0.91	nd	0.432	0.510, 0.391	0.056	nd	0.415	0.427, 0.391	0.41	nd
CGTAGC	0.101	0.111, 0.087	0.41	nd	0.064	0.021, 0.087	**0.032**	nd	0.088	0.089, 0.087	0.95	nd
CTTGGC	0.011	0.014, 0.005	0.37	nd	0.007	0.010, 0.005	0.64	nd	0.011	0.013, 0.005	0.39	nd
TTCAAC	0.004	0.007, 0.000	0.26	nd	--	--	--	--	0.004	0.005, 0.000	0.33	nd
CGCAAC	0.004	0.000, 0.011	0.079	nd	0.007	0.000, 0.011	0.31	nd	0.004	0.000, 0.011	**0.043**	nd
TGCAAT	0.004	0.000, 0.011	0.082	nd	0.007	0.000, 0.011	0.31	nd	0.004	0.000, 0.011	**0.045**	nd
CGTGGC	0.002	0.000, 0.005	0.22	nd	0.007	0.010, 0.005	0.64	nd	0.004	0.003, 0.005	0.61	nd
TTTGGT	--	--	--	--	0.004	0.010, 0.000	0.17	nd	0.002	0.003, 0.000	0.48	nd
CTTGAT	0.002	0.004, 0.000	0.43	nd	--	--	--	--	0.002	0.003, 0.000	0.49	nd
CTCGGT	--	--	--	--	0.004	0.010, 0.000	0.17	nd	--	--	--	--
TGTAAT	--	--	--	--	0.002	0.006, 0.000	0.31	nd	--	--	--	--
TGTAGT	--	--	--	--	0.002	0.005, 0.000	0.35	nd	--	--	--	--
H2AFX												
Block 8												
AA	0.623	0.621, 0.625	0.94	nd	0.596	0.541, 0.625	0.18	nd	0.609	0.601, 0.625	0.59	nd
GG	0.368	0.375, 0.359	0.72	nd	0.382	0.427, 0.358	0.27	nd	0.378	0.388, 0.359	0.50	nd
AG	0.004	0.004, 0.005	0.78	nd	0.015	0.032, 0.006	0.085	nd	0.009	0.011, 0.006	0.53	nd
GA	0.004	0.000, 0.011	0.084	nd	0.007	0.000, 0.011	0.31	nd	0.004	0.000, 0.011	**0.046**	nd
LIG1												
Block 9												
GCA	0.437	0.428, 0.450	0.64	nd	0.452	0.457, 0.450	0.90	nd	0.44	0.436, 0.450	0.74	nd
ATG	0.4	0.405, 0.393	0.81	nd	0.382	0.354, 0.397	0.48	nd	0.393	0.392, 0.397	0.91	nd
ACG	0.145	0.156, 0.128	0.41	nd	0.127	0.135, 0.124	0.79	nd	0.142	0.151, 0.124	0.41	nd
GTG	0.007	0.007, 0.006	0.83	nd	0.007	0.010, 0.006	0.66	nd	0.007	0.008, 0.006	0.75	nd
GCG	0.004	0.000, 0.011	0.081	nd	0.015	0.021, 0.012	0.53	nd	0.007	0.005, 0.011	0.46	nd
ATA	0.004	0.000, 0.011	0.087	nd	0.008	0.011, 0.007	0.79	nd	0.004	0.003, 0.007	0.44	nd
ACA	--	--	--	--	0.007	0.011, 0.004	0.49	nd	0.003	0.003, 0.004	0.84	nd
GTA	0.002	0.004, 0.000	0.47	nd	--	--	--	--	0.002	0.003, 0.000	0.53	nd
LIG1												
Block 10												
TA	0.526	0.518, 0.538	0.67	nd	0.528	0.510, 0.538	0.66	nd	0.523	0.516, 0.538	0.62	nd
CG	0.461	0.461, 0.462	0.98	nd	0.464	0.469, 0.462	0.92	nd	0.462	0.463, 0.462	0.99	nd
TG	0.009	0.014, 0.000	0.11	nd	0.007	0.021, 0.000	**0.049**	nd	0.011	0.016, 0.000	0.087	nd
CA	0.004	0.007, 0.000	0.27	nd	--	--	--	--	0.004	0.005, 0.000	0.34	nd
MLH1 [d]												
Block 11												
GCTCTGC	0.526	0.511, 0.549	0.42	1.00	0.525	0.490, 0.543	0.39	1.00	0.520	0.505, 0.549	0.33	1.00
ATCACAT	0.431	0.461, 0.385	0.11	0.68	0.414	0.479, 0.380	0.11	0.46	0.437	0.465, 0.379	0.054	0.53
ATTATAT	0.008	0.000, 0.020	**0.018**	**0.036**	0.013	0.000, 0.020	0.16	0.64	0.006	0.000, 0.020	**0.0062**	**0.014**
GCTCTGT	0.004	0.004, 0.005	0.76	1.00	--	--	--	--	0.005	0.003, 0.011	0.21	0.97
GCTATAT	--	--	--	--	0.004	0.000, 0.005	0.47	1.00	0.004	0.003, 0.005	0.60	1.00
ATTACAT	0.005	0.004, 0.008	0.55	1.00	0.005	0.000, 0.007	0.43	1.00	0.004	0.003, 0.008	0.40	1.00
ACTCTGC	0.004	0.007, 0.000	0.25	1.00	--	--	--	--	0.004	0.005, 0.000	0.33	1.00
GCCCCAT	0.004	0.007, 0.000	0.25	1.00	--	--	--	--	0.004	0.005, 0.000	0.32	1.00
ATCATGT	0.004	0.000, 0.011	0.079	0.32	0.007	0.000, 0.011	0.31	1.00	0.004	0.000, 0.011	**0.042**	0.25
GTCACAT	0.004	0.000, 0.011	0.081	0.38	0.007	0.000, 0.011	0.31	1.00	0.004	0.000, 0.011	**0.043**	0.28
ACTCCAT	--	--	--	--	--	--	--	--	0.002	0.003, 0.000	0.49	1.00
GCTATGC	--	--	--	--	0.004	0.010, 0.000	0.17	0.94	0.002	0.003, 0.000	0.49	1.00
GTTACGT	0.002	0.004, 0.000	0.42	1.00	--	--	--	--	0.002	0.003, 0.000	0.49	1.00
GTTCTGC	0.002	0.004, 0.000	0.42	1.00	--	--	--	--	0.002	0.003, 0.000	0.48	1.00
ATTACGT	0.002	0.000, 0.005	0.22	0.95	0.004	0.000, 0.005	0.47	1.00	0.002	0.000, 0.005	0.15	0.86
GCCATGC	0.001	0.000, 0.003	0.37	1.00	0.002	0.000, 0.003	0.60	1.00	--	--	--	--
GTCATGC	0.001	0.000, 0.003	0.39	1.00	0.002	0.000, 0.003	0.62	1.00	--	--	--	--
GCTCTAT	--	--	--	--	0.011	0.010, 0.011	0.97	1.00	--	--	--	--
ACTACAT	--	--	--	--	0.004	0.010, 0.000	0.17	0.92	--	--	--	--

Supplemental Table III-8, continued.

	IgAD				CVID				Combined CVID & IgAD			
	Hap freq.	Case, control frequencies	p-value	Corrected p-value [a]	Hap freq.	Case, control frequencies	p-value	Corrected p-value [a]	Hap freq.	Case, control frequencies	p-value	Corrected p-value [a]
MLH3												
Block 12												
GCAAGCGC	0.444	0.432, 0.462	0.53	nd	0.486	0.531, 0.462	0.27	nd	0.461	0.457, 0.467	0.83	nd
CTGGAGAT	0.459	0.461, 0.456	0.93	nd	0.417	0.344, 0.455	0.073	nd	0.438	0.431, 0.454	0.60	nd
CTGGACGC	0.065	0.079, 0.043	0.13	nd	0.044	0.042, 0.045	0.90	nd	0.062	0.069, 0.046	0.28	nd
GTAGGCGC	0.006	0.011, 0.000	0.16	nd	0.005	0.016, 0.000	0.089	nd	0.011	0.016, 0.000	0.087	nd
CTGAAGAT	0.007	0.007, 0.005	0.81	nd	0.002	0.000, 0.003	0.60	nd	0.005	0.005, 0.005	1.00	nd
CTGGGGAT	--	--	--	--	0.008	0.021, 0.002	0.090	nd	0.004	0.005, 0.002	0.56	nd
CTGGAGAC	0.004	0.004, 0.005	0.76	nd	0.004	0.000, 0.005	0.47	nd	0.004	0.003, 0.006	0.60	nd
CTGAGGAT	--	--	--	--	0.004	0.010, 0.000	0.17	nd	0.002	0.003, 0.000	0.49	nd
GCAGGCGC	--	--	--	--	0.004	0.010, 0.000	0.17	nd	0.002	0.003, 0.000	0.49	nd
GCAAACGC	--	--	--	--	0.002	0.006, 0.000	0.30	nd	0.002	0.003, 0.000	0.49	nd
CTGGAGGT	0.002	0.004, 0.000	0.42	nd	--	--	--	--	0.002	0.003, 0.000	0.49	nd
GTAAACGC	0.002	0.004, 0.000	0.42	nd	--	--	--	--	0.002	0.003, 0.000	0.49	nd
CTAAAGGT	0.002	0.000, 0.005	0.22	nd	0.004	0.000, 0.005	0.47	nd	0.002	0.000, 0.005	0.15	nd
CTAAGCGC	0.002	0.000, 0.005	0.22	nd	0.004	0.000, 0.005	0.47	nd	0.002	0.000, 0.005	0.15	nd
GCAAAGAT	0.002	0.000, 0.005	0.22	nd	0.004	0.000, 0.005	0.47	nd	0.002	0.000, 0.005	0.15	nd
CTGGGCGC	0.002	0.000, 0.005	0.22	nd	0.003	0.000, 0.004	0.53	nd	0.001	0.000, 0.004	0.24	nd
GCGGGCGC	0.001	0.000, 0.003	0.38	nd	0.002	0.000, 0.003	0.61	nd	--	--	--	--
GCGGGGGC	0.001	0.000, 0.003	0.38	nd	0.002	0.000, 0.003	0.61	nd	--	--	--	--
GTAAGCGC	--	--	--	--	0.005	0.016, 0.000	0.089	nd	--	--	--	--
CCAAACGC	--	--	--	--	0.002	0.004, 0.000	0.37	nd	--	--	--	--
GTGAAGAT	--	--	--	--	0.001	0.000, 0.002	0.65	nd	--	--	--	--
MRE11 [c]												
Block 13												
AC	0.651	0.654, 0.646	0.86	1.00	0.636	0.617, 0.646	0.63	1.00	0.646	0.645, 0.646	0.97	1.00
GT	0.335	0.338, 0.331	0.88	1.00	0.348	0.383, 0.331	0.39	1.00	0.343	0.349, 0.331	0.67	1.00
GC	0.009	0.004, 0.017	0.15	0.90	0.011	0.000, 0.017	0.20	0.96	0.008	0.003, 0.017	0.075	0.72
AT	0.005	0.004, 0.006	0.76	1.00	0.004	0.000, 0.006	0.48	1.00	0.004	0.003, 0.006	0.61	1.00
MRE11 [c]												
Block 14												
TC	0.564	0.585, 0.532	0.25	1.00	0.571	0.646, 0.532	0.067	0.63	0.578	0.601, 0.532	0.12	0.87
CG	0.392	0.375, 0.418	0.36	1.00	0.385	0.323, 0.417	0.12	0.81	0.38	0.361, 0.418	0.20	0.99
CC	0.038	0.036, 0.040	0.85	1.00	0.037	0.032, 0.040	0.74	1.00	0.036	0.035, 0.039	0.79	1.00
TG	0.007	0.004, 0.011	0.34	1.00	0.008	0.000, 0.011	0.31	1.00	0.006	0.003, 0.011	0.21	0.99
MRE11 [c]												
Block 15												
GT	0.558	0.586, 0.516	0.14	0.88	0.56	0.645, 0.516	**0.039**	0.27	0.573	0.601, 0.516	0.057	0.62
AA	0.422	0.396, 0.462	0.16	0.93	0.421	0.343, 0.462	0.057	0.57	0.409	0.383, 0.462	0.075	0.72
AT	0.013	0.014, 0.011	0.78	1.00	0.011	0.011, 0.011	0.97	1.00	0.013	0.013, 0.011	0.83	1.00
GA	0.007	0.004, 0.011	0.33	1.00	0.007	0.000, 0.011	0.33	1.00	0.006	0.003, 0.011	0.21	0.99
MSH2 [b, c]												
Block 16												
GCGACGCG	0.575	0.543, 0.625	0.079	0.40	0.568	0.458, 0.625	**0.0075**	**0.025**	0.558	0.521, 0.633	**0.013**	0.066
ATCTGAAA	0.258	0.285, 0.217	0.10	0.74	0.25	0.311, 0.217	0.085	0.53	0.27	0.293, 0.222	0.077	0.57
ATGAGAAA	0.095	0.115, 0.065	0.075	0.39	0.086	0.126, 0.065	0.085	0.53	0.102	0.118, 0.067	0.059	0.47
ATGAGAAG	0.011	0.007, 0.016	0.32	1.00	0.025	0.041, 0.016	0.21	0.99	0.016	0.015, 0.017	0.91	1.00
GCGACGAG	0.015	0.018, 0.011	0.54	1.00	0.007	0.000, 0.011	0.31	1.00	0.013	0.013, 0.011	0.82	1.00
ATGACGCG	0.004	0.004, 0.005	0.76	1.00	0.011	0.021, 0.005	0.23	0.99	0.007	0.008, 0.006	0.75	1.00
GCGACAAG	0.004	0.004, 0.005	0.76	1.00	0.007	0.010, 0.005	0.64	1.00	0.005	0.005, 0.006	0.98	1.00
ATCTGAAG	0.005	0.004, 0.005	0.83	1.00	0.004	0.001, 0.005	0.58	1.00	0.004	0.003, 0.006	0.71	1.00
ATCAGAAA	0.004	0.007, 0.000	0.25	1.00	--	--	--	--	0.004	0.005, 0.000	0.33	1.00
GCGAGGCG	0.002	0.004, 0.000	0.41	1.00	0.004	0.010, 0.000	0.16	0.70	0.004	0.005, 0.000	0.32	1.00
GTCAGAAA	--	--	--	--	0.004	0.010, 0.000	0.17	0.94	0.002	0.003, 0.000	0.49	1.00
GTCTGACG	0.004	0.000, 0.011	0.080	0.65	0.007	0.000, 0.011	0.31	1.00	0.002	0.000, 0.006	0.15	0.75
GCCACGCA	0.002	0.004, 0.000	0.41	1.00	--	--	--	--	0.002	0.003, 0.000	0.48	1.00
GCGACACA	0.002	0.004, 0.000	0.42	1.00	--	--	--	--	0.002	0.003, 0.000	0.49	1.00
GTCACGCG	0.002	0.004, 0.000	0.42	1.00	--	--	--	--	0.002	0.003, 0.000	0.49	1.00
ACGACGCA	0.002	0.000, 0.005	0.22	0.98	0.004	0.000, 0.005	0.47	1.00	0.002	0.000, 0.006	0.15	0.75
ATGTGGAA	0.002	0.000, 0.005	0.22	0.98	0.004	0.000, 0.005	0.47	1.00	0.002	0.000, 0.006	0.15	0.75
GCGTGGCG	0.002	0.000, 0.005	0.22	0.98	0.004	0.000, 0.005	0.47	1.00	0.002	0.000, 0.006	0.15	0.75
GTCAGGCA	0.002	0.000, 0.005	0.22	0.98	0.004	0.000, 0.005	0.47	1.00	0.002	0.000, 0.006	0.15	0.75
GTGTGGCG	0.002	0.000, 0.005	0.22	0.98	0.004	0.000, 0.005	0.47	1.00	0.002	0.000, 0.006	0.15	0.75
ACCTGGCG	0.001	0.000, 0.003	0.39	1.00	0.002	0.000, 0.003	0.61	1.00	--	--	--	--
ACGAGAAA	0.001	0.000, 0.003	0.38	1.00	0.002	0.000, 0.003	0.61	1.00	--	--	--	--
ACGTGAAA	0.001	0.000, 0.003	0.39	1.00	0.002	0.000, 0.003	0.61	1.00	--	--	--	--
GCCTGGCG	0.001	0.000, 0.003	0.38	1.00	0.002	0.000, 0.003	0.61	1.00	--	--	--	--
GCCACGAG	--	--	--	--	0.002	0.005, 0.000	0.33	1.00	--	--	--	--
GCCAGGAG	--	--	--	--	0.002	0.005, 0.000	0.33	1.00	--	--	--	--

Supplemental Table III-8, continued.

	IgAD				CVID				Combined CVID & IgAD			
	Hap freq.	Case, control frequencies	p-value	Corrected p-value [a]	Hap freq.	Case, control frequencies	p-value	Corrected p-value [a]	Hap freq.	Case, control frequencies	p-value	Corrected p-value [a]
MSH3 [c]												
Block 17												
TGTCTCGGT	0.292	0.286, 0.304	0.68	1.00	0.311	0.332, 0.301	0.60	1.00	0.3	0.298, 0.302	0.93	1.00
TGTCCTGAT	0.221	0.246, 0.186	0.13	0.97	0.224	0.301, 0.183	**0.026**	0.35	0.236	0.262, 0.184	**0.042**	0.55
CACTTCTGT	0.147	0.162, 0.125	0.27	1.00	0.115	0.094, 0.125	0.44	1.00	0.138	0.145, 0.124	0.51	1.00
TGTCTTGGT	0.127	0.125, 0.131	0.84	1.00	0.115	0.084, 0.131	0.24	1.00	0.118	0.112, 0.130	0.55	1.00
CACTTCTGC	0.093	0.089, 0.099	0.74	1.00	0.096	0.093, 0.097	0.91	1.00	0.093	0.091, 0.098	0.79	1.00
TGTCTCGAT	0.072	0.057, 0.096	0.12	0.97	0.073	0.031, 0.095	0.052	0.76	0.065	0.051, 0.095	**0.047**	0.77
TGTCCCGGT	0.006	0.002, 0.012	0.19	1.00	0.012	0.011, 0.012	0.91	1.00	0.007	0.004, 0.013	0.29	1.00
CACTTTTGT	0.006	0.010, 0.001	0.25	1.00	--	--	--	--	0.005	0.008, 0.001	0.32	1.00
TGTTTCTGT	0.006	0.000, 0.015	**0.044**	0.49	0.01	0.001, 0.014	0.26	1.00	0.005	0.000, 0.015	**0.022**	0.32
TGTTCCTGT	--	--	--	--	0.007	0.010, 0.005	0.67	1.00	0.004	0.004, 0.005	0.75	1.00
TGTCCCGAT	0.004	0.001, 0.010	0.13	0.97	0.007	0.001, 0.010	0.39	1.00	0.004	0.000, 0.010	0.086	0.94
CGTCTCTGT	0.004	0.004, 0.005	0.76	1.00	0.004	0.000, 0.005	0.47	1.00	0.004	0.003, 0.005	0.61	1.00
CGCTTCGGT	0.004	0.004, 0.005	0.77	1.00	0.001	0.000, 0.002	0.69	1.00	0.004	0.003, 0.005	0.61	1.00
TATCTCGGT	--	--	--	--	--	--	--	--	0.003	0.004, 0.000	0.37	1.00
CGTCCTGAT	--	--	--	--	0.004	0.010, 0.000	0.17	0.98	0.002	0.003, 0.000	0.48	1.00
TATTTCTGC	--	--	--	--	0.004	0.000, 0.005	0.50	1.00	0.002	0.000, 0.005	0.15	1.00
TGTCCTGAC	--	--	--	--	0.004	0.011, 0.000	0.19	1.00	0.002	0.003, 0.000	0.52	1.00
TGTTCCGAT	--	--	--	--	0.004	0.010, 0.000	0.17	1.00	0.002	0.003, 0.000	0.49	1.00
TGTCTCTGT	0.003	0.000, 0.007	0.16	0.99	0.005	0.000, 0.007	0.40	1.00	0.002	0.000, 0.007	0.10	0.96
TGTTTCGAT	0.002	0.004, 0.000	0.42	1.00	--	--	--	--	0.002	0.003, 0.000	0.48	1.00
TATCTTGGT	--	--	--	--	--	--	--	--	0.001	0.002, 0.000	0.58	1.00
TATCCTGAT	0.004	0.007, 0.000	0.25	1.00	--	--	--	--	0.001	0.002, 0.000	0.55	1.00
CATTTCTGC	0.001	0.000, 0.003	0.34	1.00	--	--	--	--	--	--	--	--
CGCCTCTGC	0.001	0.002, 0.000	0.57	1.00	--	--	--	--	--	--	--	--
TGCCTCTGC	0.001	0.002, 0.000	0.57	1.00	--	--	--	--	--	--	--	--
CACTTCGGT	--	--	--	--	0.003	0.000, 0.004	0.58	1.00	--	--	--	--
TATTTCGGT	--	--	--	--	0.003	0.010, 0.000	0.17	1.00	--	--	--	--
MSH3 [c]												
Block 18												
TGA	0.451	0.448, 0.456	0.87	1.00	0.416	0.342, 0.455	0.068	0.83	0.433	0.422, 0.456	0.44	1.00
TTA	0.282	0.266, 0.307	0.34	1.00	0.291	0.259, 0.308	0.39	1.00	0.278	0.264, 0.307	0.28	1.00
GTT	0.137	0.143, 0.128	0.64	1.00	0.171	0.252, 0.129	**0.010**	0.14	0.157	0.171, 0.128	0.19	1.00
GTA	0.126	0.142, 0.103	0.22	1.00	0.116	0.145, 0.101	0.27	1.00	0.13	0.143, 0.103	0.18	1.00
GGA	0.003	0.001, 0.006	0.34	1.00	0.005	0.002, 0.007	0.57	1.00	0.003	0.001, 0.006	0.26	1.00
MSH3 [c]												
Block 19												
GCT	0.722	0.711, 0.739	0.51	1.00	0.75	0.771, 0.739	0.56	1.00	0.73	0.726, 0.739	0.75	1.00
ATC	0.261	0.282, 0.228	0.20	1.00	0.218	0.198, 0.228	0.56	1.00	0.25	0.261, 0.228	0.40	1.00
GTC	0.009	0.000, 0.022	**0.014**	0.20	0.014	0.000, 0.022	0.14	0.95	0.007	0.000, 0.022	**0.0043**	0.075
GTT	0.002	0.004, 0.000	0.42	1.00	0.004	0.010, 0.000	0.17	1.00	0.004	0.005, 0.000	0.33	1.00
GCC	0.001	0.000, 0.003	0.40	1.00	0.007	0.010, 0.006	0.64	1.00	0.004	0.003, 0.005	0.60	1.00
ACT	0.002	0.004, 0.000	0.43	1.00	--	--	--	--	0.003	0.004, 0.000	0.41	1.00
ATT	0.002	0.000, 0.005	0.22	1.00	0.007	0.010, 0.005	0.64	1.00	0.003	0.001, 0.005	0.38	1.00
ACC	0.001	0.000, 0.003	0.37	1.00	--	--	--	--	--	--	--	--
MSH4 [f]												
Block 20												
CAGCTGAAAAG	0.384	0.393, 0.370	0.63	1.00	0.366	0.364, 0.370	0.93	1.00	0.38	0.385, 0.372	0.75	1.00
TAGCTGACTGA	0.264	0.257, 0.275	0.66	1.00	0.261	0.229, 0.279	0.37	1.00	0.259	0.251, 0.277	0.51	1.00
CAGCGAGAAAG	0.196	0.196, 0.195	0.99	1.00	0.193	0.187, 0.196	0.85	1.00	0.195	0.195, 0.196	0.96	1.00
CTATTGAAAAG	0.067	0.064, 0.071	0.79	1.00	0.064	0.052, 0.071	0.54	1.00	0.065	0.062, 0.071	0.67	1.00
CTGTTGAAAAG	0.022	0.018, 0.028	0.48	1.00	0.033	0.042, 0.028	0.55	1.00	0.025	0.024, 0.028	0.78	1.00
CAGCGAAAAAG	0.02	0.029, 0.006	0.076	0.62	0.018	0.042, 0.006	**0.031**	0.34	0.024	0.033, 0.006	**0.050**	0.65
CAGCTGGAAAG	0.008	0.004, 0.015	0.20	0.98	0.015	0.011, 0.017	0.69	1.00	0.009	0.006, 0.015	0.25	1.00
TAGCTGACTGG	0.004	0.007, 0.000	0.25	1.00	--	--	--	--	0.005	0.008, 0.000	0.22	1.00
CAGCGGAAAAG	0.004	0.007, 0.000	0.27	1.00	0.004	0.010, 0.000	0.19	1.00	0.005	0.008, 0.000	0.27	1.00
CAGCTGACTGA	0.004	0.007, 0.000	0.25	1.00	0.004	0.010, 0.000	0.17	0.99	0.005	0.008, 0.000	0.23	1.00
TAGCTGAAAAG	0.004	0.004, 0.006	0.76	1.00	0.004	0.000, 0.006	0.47	1.00	0.004	0.003, 0.006	0.60	1.00
TAGCTGACAGA	0.002	0.000, 0.005	0.22	1.00	0.007	0.010, 0.005	0.64	1.00	0.004	0.003, 0.005	0.61	1.00
CAGCGAGAAGG	--	--	--	--	0.004	0.010, 0.000	0.17	0.99	0.002	0.003, 0.000	0.48	1.00
CAGCTGAATAG	--	--	--	--	0.004	0.011, 0.000	0.19	1.00	0.002	0.003, 0.000	0.51	1.00
TAGCTGGCTGA	0.003	0.004, 0.002	0.75	1.00	--	--	--	--	0.002	0.003, 0.002	0.80	1.00
CAGCGAGCAAG	0.002	0.004, 0.000	0.42	1.00	--	--	--	--	0.002	0.003, 0.000	0.48	1.00
CAGCGGACTGA	0.002	0.004, 0.000	0.41	1.00	--	--	--	--	0.002	0.003, 0.000	0.48	1.00
TAGCTGAATGA	0.002	0.004, 0.000	0.42	1.00	--	--	--	--	0.002	0.003, 0.000	0.49	1.00
CAGCTGAAAGG	0.002	0.000, 0.005	0.22	1.00	0.004	0.000, 0.005	0.47	1.00	0.002	0.000, 0.005	0.15	0.97
CAGTTGAAAAG	0.002	0.000, 0.005	0.21	0.99	0.004	0.000, 0.006	0.46	1.00	0.002	0.000, 0.006	0.15	0.96
CTACTGAAAAG	0.002	0.000, 0.005	0.22	1.00	0.004	0.000, 0.005	0.47	1.00	0.002	0.000, 0.005	0.15	0.98
CAGCGGGAAAG	0.002	0.000, 0.005	0.23	1.00	0.003	0.000, 0.005	0.48	1.00	0.002	0.000, 0.005	0.17	1.00
TAGCTAAAAAG	0.001	0.000, 0.003	0.38	1.00	--	--	--	--	--	--	--	--
TAGCTAGAAAG	0.001	0.000, 0.003	0.39	1.00	--	--	--	--	--	--	--	--
TAGCGAGCTGG	--	--	--	--	0.004	0.010, 0.000	0.17	0.99	--	--	--	--
TAGCGGGCAAA	--	--	--	--	0.002	0.005, 0.000	0.33	1.00	--	--	--	--
TAGCGGGCTAA	--	--	--	--	0.002	0.005, 0.000	0.33	1.00	--	--	--	--

Supplemental Table III-8, continued.

	IgAD				CVID				Combined CVID & IgAD			
	Hap freq.	Case, control frequencies	p-value	Corrected p-value [a]	Hap freq.	Case, control frequencies	p-value	Corrected p-value [a]	Hap freq.	Case, control frequencies	p-value	Corrected p-value [a]
MSH5 [f]												
Block 21												
TGCGATT	0.363	0.328, 0.418	**0.049**	0.46	0.412	0.399, 0.418	0.76	1.00	0.37	0.345, 0.422	0.077	0.78
TGCAGCC	0.243	0.307, 0.146	**7.5E-05**	**0.0010**	0.146	0.146, 0.145	0.99	1.00	0.227	0.266, 0.147	**0.0017**	**0.023**
TACGATT	0.139	0.091, 0.212	**2.0E-04**	**0.0033**	0.211	0.208, 0.212	0.94	1.00	0.15	0.121, 0.209	**0.0066**	0.095
GGTGATC	0.104	0.105, 0.103	0.95	1.00	0.102	0.100, 0.103	0.95	1.00	0.104	0.104, 0.104	0.98	1.00
GGCGATC	0.078	0.089, 0.060	0.25	1.00	0.065	0.073, 0.060	0.67	1.00	0.077	0.085, 0.061	0.31	1.00
TGCGATC	0.055	0.058, 0.050	0.72	1.00	0.044	0.031, 0.050	0.46	1.00	0.051	0.051, 0.051	0.98	1.00
TGCGGTT	0.002	0.004, 0.000	0.42	1.00	0.01	0.028, 0.000	**0.024**	0.28	0.007	0.011, 0.000	0.17	1.00
TGTGATT	0.005	0.004, 0.006	0.74	1.00	0.002	0.000, 0.003	0.60	1.00	0.004	0.003, 0.006	0.58	1.00
TGCAGCT	0.003	0.005, 0.000	0.35	1.00	--	--	--	--	0.003	0.004, 0.000	0.43	1.00
TGCGGCC	--	--	--	--	0.004	0.010, 0.000	0.17	1.00	0.002	0.003, 0.000	0.49	1.00
GACGATT	0.002	0.004, 0.000	0.43	1.00	--	--	--	--	0.002	0.003, 0.000	0.50	1.00
TACGACT	0.002	0.004, 0.000	0.43	1.00	--	--	--	--	0.002	0.003, 0.000	0.49	1.00
GGTAGCC	0.001	0.002, 0.000	0.52	1.00	--	--	--	--	0.001	0.002, 0.000	0.58	1.00
TACAACT	0.001	0.000, 0.003	0.38	1.00	0.002	0.000, 0.003	0.62	1.00	--	--	--	--
TACAATT	0.001	0.000, 0.003	0.38	1.00	0.002	0.000, 0.003	0.60	1.00	--	--	--	--
TATGATT	--	--	--	--	0.002	0.000, 0.003	0.64	1.00	--	--	--	--
GGTGGTC	--	--	--	--	0.001	0.003, 0.000	0.42	1.00	--	--	--	--
MUS81 [h]												
Block 22												
AACTG	0.504	0.489, 0.527	0.42	0.95	0.52	0.510, 0.525	0.80	1.00	0.504	0.494, 0.525	0.48	0.99
GGTCA	0.303	0.325, 0.269	0.20	0.61	0.269	0.276, 0.266	0.85	1.00	0.297	0.313, 0.266	0.26	0.91
AACTA	0.148	0.152, 0.142	0.75	1.00	0.14	0.139, 0.141	0.96	1.00	0.147	0.149, 0.141	0.81	1.00
AATTA	0.018	0.018, 0.018	0.97	1.00	0.018	0.019, 0.017	0.91	1.00	0.018	0.018, 0.017	0.94	1.00
AATTG	0.016	0.016, 0.017	0.91	1.00	0.02	0.024, 0.017	0.72	1.00	0.018	0.018, 0.017	0.95	1.00
AATCA	0.011	0.000, 0.027	**0.0053**	0.014	0.018	0.000, 0.027	0.11	0.49	0.009	0.000, 0.027	**0.0014**	0.0054
GGTTA	--	--	--	--	0.007	0.011, 0.005	0.63	1.00	0.004	0.003, 0.005	0.61	1.00
AGTCA	--	--	--	--	0.004	0.011, 0.000	0.17	0.74	0.002	0.003, 0.000	0.50	1.00
GGCCA	--	--	--	--	0.004	0.011, 0.000	0.19	0.83	0.002	0.003, 0.000	0.51	1.00
RAD50 [c]												
Block 23												
TGAGCA	0.365	0.398, 0.313	0.064	0.50	0.367	0.476, 0.310	**0.0062**	**0.041**	0.386	0.422, 0.311	**0.011**	0.097
TGATCA	0.4	0.384, 0.426	0.37	1.00	0.368	0.253, 0.429	**0.0038**	**0.024**	0.373	0.346, 0.428	0.061	0.63
CACTTG	0.19	0.177, 0.210	0.38	1.00	0.2	0.183, 0.209	0.60	1.00	0.189	0.179, 0.210	0.37	1.00
CACTTA	0.02	0.018, 0.022	0.73	1.00	0.026	0.032, 0.022	0.61	1.00	0.022	0.022, 0.022	0.97	1.00
TGATCG	--	--	--	--	0.007	0.021, 0.000	0.056	0.54	0.004	0.005, 0.000	0.33	1.00
TGCGCA	0.004	0.007, 0.000	0.25	1.00	--	--	--	--	0.004	0.005, 0.000	0.33	1.00
CACGCG	--	--	--	--	0.005	0.010, 0.003	0.40	1.00	0.003	0.003, 0.003	0.99	1.00
CACGTG	0.004	0.002, 0.007	0.36	1.00	0.006	0.003, 0.008	0.64	1.00	0.003	0.001, 0.007	0.30	1.00
CACTCG	0.004	0.000, 0.011	0.08	0.65	0.006	0.000, 0.008	0.40	1.00	0.003	0.000, 0.008	0.084	0.77
TGATTA	0.002	0.000, 0.005	0.25	0.99	0.007	0.010, 0.005	0.61	1.00	0.003	0.003, 0.005	0.64	1.00
CGCTTG	--	--	--	--	--	--	--	--	0.002	0.004, 0.000	0.43	1.00
TGAGTG	--	--	--	--	0.003	0.010, 0.000	0.17	0.89	0.002	0.003, 0.000	0.48	1.00
CAATCA	0.002	0.004, 0.000	0.42	1.00	--	--	--	--	0.002	0.003, 0.000	0.48	1.00
TGAGCG	0.002	0.004, 0.000	0.42	1.00	--	--	--	--	0.002	0.003, 0.000	0.50	1.00
TGATTG	0.002	0.000, 0.006	0.19	0.97	0.004	0.001, 0.006	0.52	1.00	0.002	0.000, 0.006	0.16	0.94
CGCGTG	0.004	0.007, 0.000	0.25	1.00	--	--	--	--	0.001	0.002, 0.000	0.56	1.00
RAD51 [g]												
Block 24												
GCGGTAG	0.387	0.402, 0.365	0.43	1.00	0.36	0.343, 0.368	0.68	1.00	0.379	0.387, 0.364	0.60	1.00
ACAACGT	0.351	0.328, 0.386	0.21	1.00	0.396	0.416, 0.386	0.62	1.00	0.362	0.351, 0.386	0.42	1.00
GTAACGT	0.151	0.161, 0.136	0.46	1.00	0.128	0.114, 0.136	0.62	1.00	0.144	0.149, 0.136	0.68	1.00
GCGGCAG	0.061	0.066, 0.053	0.57	1.00	0.054	0.063, 0.050	0.66	1.00	0.062	0.065, 0.054	0.61	1.00
GCAGTAG	0.012	0.014, 0.008	0.55	1.00	0.006	0.010, 0.003	0.45	1.00	0.012	0.013, 0.010	0.72	1.00
GCAACGT	0.004	0.004, 0.005	0.76	1.00	0.014	0.031, 0.006	0.085	0.96	0.009	0.011, 0.005	0.54	1.00
ACGGTAG	0.006	0.004, 0.011	0.34	1.00	0.007	0.000, 0.011	0.31	1.00	0.005	0.003, 0.011	0.21	1.00
GCGGTAT	0.004	0.004, 0.005	0.76	1.00	0.004	0.000, 0.005	0.47	1.00	0.004	0.003, 0.005	0.61	1.00
GTAATGG	0.003	0.004, 0.000	0.38	1.00	--	--	--	--	0.002	0.003, 0.000	0.44	1.00
ACAGCGT	0.002	0.003, 0.000	0.43	1.00	--	--	--	--	0.002	0.003, 0.000	0.49	1.00
GCGATAG	0.002	0.004, 0.000	0.42	1.00	--	--	--	--	0.002	0.003, 0.000	0.49	1.00
ACAACGG	0.001	0.002, 0.000	0.56	1.00	--	--	--	--	0.002	0.003, 0.000	0.48	1.00
GCAGCAG	0.003	0.000, 0.008	0.15	1.00	0.009	0.000, 0.013	0.26	1.00	0.002	0.000, 0.007	0.13	1.00
GTAACAG	--	--	--	--	0.004	0.010, 0.000	0.17	1.00	0.002	0.003, 0.000	0.48	1.00
GTAATGT	--	--	--	--	0.004	0.011, 0.000	0.19	1.00	0.002	0.003, 0.000	0.50	1.00
ACGACGT	0.002	0.000, 0.005	0.22	1.00	0.004	0.000, 0.005	0.48	1.00	0.002	0.000, 0.005	0.16	1.00
ACGGCGT	0.002	0.000, 0.005	0.25	1.00	0.004	0.000, 0.005	0.47	1.00	0.002	0.000, 0.005	0.17	1.00
GCGGCGG	0.002	0.000, 0.005	0.22	1.00	0.004	0.000, 0.005	0.47	1.00	0.002	0.000, 0.005	0.15	1.00
GCGGTGG	0.002	0.000, 0.005	0.22	1.00	0.004	0.000, 0.005	0.47	1.00	0.002	0.000, 0.005	0.15	1.00
GTAACGG	0.002	0.003, 0.000	0.46	1.00	--	--	--	--	0.001	0.002, 0.000	0.53	1.00
ACGACGG	0.001	0.002, 0.000	0.58	1.00	--	--	--	--	--	--	--	--

Supplemental Table III-8, continued.

	IgAD				CVID				Combined CVID & IgAD			
	Hap freq.	Case, control frequencies	p-value	Corrected p-value [a]	Hap freq.	Case, control frequencies	p-value	Corrected p-value [a]	Hap freq.	Case, control frequencies	p-value	Corrected p-value [a]
RAD52 [g]												
Block 25												
TTTAC	0.257	0.255, 0.260	0.90	1.00	0.256	0.248, 0.260	0.83	1.00	0.255	0.253, 0.260	0.86	1.00
TTTAT	0.262	0.248, 0.283	0.40	1.00	0.247	0.178, 0.283	0.054	0.86	0.248	0.231, 0.283	0.18	1.00
GATGT	0.202	0.200, 0.205	0.90	1.00	0.184	0.146, 0.204	0.23	1.00	0.192	0.186, 0.205	0.60	1.00
GACAT	0.171	0.184, 0.151	0.37	1.00	0.193	0.271, 0.152	**0.017**	0.37	0.186	0.202, 0.151	0.14	1.00
GATAT	0.096	0.096, 0.094	0.93	1.00	0.108	0.134, 0.094	0.30	1.00	0.104	0.109, 0.094	0.59	1.00
GACAC	0.006	0.009, 0.001	0.24	1.00	--	--	--	--	0.007	0.010, 0.001	0.21	1.00
GTTGT	0.004	0.004, 0.006	0.76	1.00	--	--	--	--	0.004	0.003, 0.005	0.60	1.00
GTTAC	0.002	0.004, 0.000	0.42	1.00	--	--	--	--	0.002	0.003, 0.000	0.48	1.00
GTCAT	--	--	--	--	0.004	0.011, 0.000	0.16	1.00	0.002	0.003, 0.000	0.48	1.00
GATAC	--	--	--	--	0.005	0.012, 0.001	0.20	1.00	--	--	--	--
TTTGT	--	--	--	--	0.004	0.000, 0.006	0.48	1.00	--	--	--	--
RAD54B [g]												
Block 26												
TAGGGCT	0.508	0.457, 0.587	**0.0062**	0.15	0.55	0.479, 0.587	0.086	0.96	0.503	0.463, 0.587	**0.0057**	0.16
CTGAGGC	0.292	0.331, 0.233	**0.023**	0.52	0.262	0.317, 0.233	0.13	0.99	0.296	0.326, 0.233	**0.023**	0.56
CTAGTGC	0.06	0.064, 0.054	0.66	1.00	0.059	0.069, 0.054	0.63	1.00	0.062	0.066, 0.054	0.58	1.00
CTGGGCC	0.05	0.047, 0.054	0.73	1.00	0.043	0.021, 0.054	0.19	1.00	0.045	0.040, 0.054	0.46	1.00
CTAGGGC	0.041	0.050, 0.027	0.23	1.00	0.032	0.042, 0.027	0.51	1.00	0.041	0.048, 0.027	0.25	1.00
TAGGGGC	0.007	0.007, 0.005	0.81	1.00	0.014	0.031, 0.005	0.084	0.95	0.01	0.013, 0.005	0.42	1.00
TTGAGGC	0.008	0.006, 0.011	0.55	1.00	0.007	0.000, 0.011	0.30	1.00	0.007	0.004, 0.011	0.34	1.00
CTGAGGT	0.002	0.000, 0.005	0.23	1.00	0.009	0.016, 0.005	0.36	1.00	0.006	0.006, 0.006	1.00	1.00
TAGGGGT	0.007	0.011, 0.000	0.16	1.00	--	--	--	--	0.005	0.008, 0.000	0.23	1.00
CTGGGCT	0.007	0.007, 0.006	0.87	1.00	0.004	0.000, 0.006	0.46	1.00	0.005	0.005, 0.006	0.94	1.00
TAGAGGC	0.003	0.004, 0.000	0.37	1.00	--	--	--	--	0.003	0.004, 0.000	0.39	1.00
CTGATGC	0.003	0.004, 0.000	0.36	1.00	--	--	--	--	0.002	0.003, 0.000	0.44	1.00
CTGGTCC	0.002	0.003, 0.000	0.52	1.00	--	--	--	--	0.002	0.002, 0.000	0.58	1.00
CTGGTGC	0.002	0.004, 0.000	0.42	1.00	--	--	--	--	0.002	0.003, 0.000	0.48	1.00
TTGGGCC	0.002	0.004, 0.000	0.42	1.00	--	--	--	--	0.002	0.003, 0.000	0.45	1.00
CAGAGGC	--	--	--	--	0.004	0.010, 0.000	0.17	1.00	0.002	0.003, 0.000	0.49	1.00
TTGAGCC	--	--	--	--	0.004	0.010, 0.000	0.17	1.00	0.002	0.002, 0.000	0.52	1.00
CAAGGGC	0.002	0.000, 0.005	0.22	1.00	0.004	0.000, 0.005	0.47	1.00	0.002	0.000, 0.005	0.15	1.00
TAGAGCT	0.002	0.000, 0.005	0.22	1.00	0.004	0.000, 0.005	0.47	1.00	0.002	0.000, 0.005	0.16	1.00
TTGGGCT	0.002	0.000, 0.005	0.24	1.00	0.004	0.000, 0.005	0.47	1.00	0.002	0.000, 0.005	0.17	1.00
CTAGTGT	--	--	--	--	0.001	0.004, 0.000	0.37	1.00	--	--	--	--
TP53BP1												
Block 27												
GTCGTTCCCTG	0.636	0.626, 0.652	0.56	nd	0.661	0.677, 0.652	0.68	nd	0.645	0.642, 0.652	0.81	nd
AGTGCCCGCGC	0.194	0.185, 0.206	0.57	nd	0.2	0.187, 0.206	0.71	nd	0.193	0.187, 0.206	0.58	nd
AGTACCTGGGC	0.084	0.100, 0.060	0.13	nd	0.061	0.062, 0.060	0.93	nd	0.079	0.088, 0.060	0.24	nd
GTCGTTCGCTG	0.019	0.021, 0.016	0.69	nd	0.018	0.021, 0.016	0.76	nd	0.02	0.022, 0.016	0.68	nd
GTCGCTCCCTG	0.019	0.025, 0.011	0.28	nd	0.007	0.000, 0.011	0.31	nd	0.016	0.019, 0.011	0.48	nd
AGTACCTGGGG	0.011	0.014, 0.005	0.37	nd	0.014	0.031, 0.005	0.084	nd	0.014	0.019, 0.005	0.21	nd
GTCATTCCCTG	0.006	0.010, 0.000	0.17	nd	--	--	--	--	0.005	0.008, 0.000	0.23	nd
AGCGTTCCCTG	0.004	0.004, 0.005	0.76	nd	0.004	0.000, 0.005	0.47	nd	0.004	0.003, 0.005	0.61	nd
AGTACTTGGGG	0.004	0.004, 0.005	0.76	nd	0.004	0.000, 0.005	0.47	nd	0.004	0.003, 0.005	0.61	nd
GTCGTTCCCGG	0.004	0.004, 0.005	0.77	nd	0.004	0.000, 0.005	0.47	nd	0.004	0.003, 0.005	0.61	nd
GTCACTCCCTG	--	--	--	--	0.004	0.010, 0.000	0.17	nd	0.002	0.002, 0.000	0.50	nd
ATCACCTGGGG	0.002	0.004, 0.000	0.42	nd	--	--	--	--	0.002	0.003, 0.000	0.48	nd
ATCGTTCCCTC	0.002	0.004, 0.000	0.42	nd	--	--	--	--	0.002	0.003, 0.000	0.48	nd
AGCGCCCGCGC	0.002	0.000, 0.006	0.21	nd	0.004	0.000, 0.006	0.46	nd	0.002	0.000, 0.006	0.15	nd
AGTACCTGGTC	0.002	0.000, 0.006	0.21	nd	0.004	0.000, 0.006	0.47	nd	0.002	0.000, 0.006	0.15	nd
AGTGCCCGCTG	0.002	0.000, 0.005	0.22	nd	0.004	0.000, 0.005	0.47	nd	0.002	0.000, 0.005	0.15	nd
GTCGTTCCCGC	0.002	0.000, 0.005	0.22	nd	0.004	0.000, 0.005	0.47	nd	0.002	0.000, 0.005	0.15	nd
GTCGTTCCGTG	0.002	0.000, 0.005	0.22	nd	0.004	0.000, 0.005	0.47	nd	0.002	0.000, 0.005	0.16	nd
AGTGCCTGGTC	0.002	0.000, 0.005	0.23	nd	0.003	0.000, 0.005	0.48	nd	0.002	0.000, 0.005	0.16	nd
AGTGCTTGCGC	--	--	--	--	0.002	0.005, 0.000	0.33	nd	--	--	--	--
AGTGCTTGGGC	--	--	--	--	0.002	0.005, 0.000	0.33	nd	--	--	--	--
XRCC2												
Block 28												
CT	0.486	0.463, 0.521	0.23	nd	0.5	0.458, 0.522	0.31	nd	0.481	0.462, 0.521	0.19	nd
TT	0.264	0.279, 0.240	0.35	nd	0.257	0.292, 0.239	0.34	nd	0.269	0.283, 0.240	0.28	nd
TG	0.247	0.253, 0.238	0.72	nd	0.243	0.250, 0.239	0.84	nd	0.247	0.252, 0.238	0.73	nd
CG	0.003	0.004, 0.001	0.49	nd	--	--	--	--	0.003	0.003, 0.001	0.55	nd

a) p-value corrected for multiple testing bias by 100,000 permutations of all markers within pathways as described in Materials and Methods
b) Permutation tests performed on haplotypes (in blocks) of the MutSα complex - p-values reported for MSH2 are from this permutation set
c) Permutation tests performed on haplotypes (in blocks) of the MutSβ complex
d) Permutation tests performed on haplotypes (in blocks) of the MutLα complex
e) Permutation tests performed on haplotypes (in blocks) of the MRN complex
f) Permutation tests performed on haplotypes (in blocks) of the MutSγ complex
g) Permutation tests performed on haplotypes (in blocks) of the RAD52 extended epistasis group
h) Permutation tests performed on haplotypes (in blocks) in MUS81

Supplementary Table III-9. SNPs identified by resequencing.

Gene	SNP ID	Detection category [a]	Variation [b]	Position (Build 36)	Classification	Novel [c]
AID	Broad13929182	1		Chr12:8646213	3'UTR	√
AID	Broad13929186	3	rs11324989	Chr12:8646692	3'UTR	-
AID	Broad13929196	1		Chr12:8647701	3'UTR	√
AID	Broad13929198	1	rs12307097	Chr12:8647838	3'UTR	-
AID	Broad13929199	1	rs11046349	Chr12:8648058	3'UTR	-
AID	Broad13929201	1	rs2028373	Chr12:8648748	SYN	-
AID	Broad13929204	1		Chr12:8649100	SYN	√
AID	Broad13929205	1		Chr12:8649295	SYN	√
AID	Broad13929208	1	rs2518144	Chr12:8650712	Intron	-
AID	Broad13929209	1		Chr12:8650860	SYN	√
APEX1	Broad13929213	1	rs1760944	Chr14:19992989	5' flanking	-
APEX1	Broad13929215	1		Chr14:19993136	5'UTR	√
APEX1	Broad13929216	1	rs3136814	Chr14:19993137	5'UTR	-
APEX1	Broad13929217	1	rs11622131	Chr14:19993163	5'UTR	-
APEX1	Broad13929218	1	rs41561214	Chr14:19993256	5'UTR	-
APEX1	Broad13929219	1	rs2307490	Chr14:19993380	5'UTR	-
APEX1	Broad13929226	1	rs2307485	Chr14:19993571	Intron	-
APEX1	Broad13929230	1	rs1048945	Chr14:19994007	Q51H	-
APEX1	Broad13929231	1	rs17111967	Chr14:19994165	Intron	-
APEX1	Broad13929232	1	rs41541122	Chr14:19994632	Intron	-
APEX1	Broad13929233	1		Chr14:19994991	SYN	√
APEX1	Broad13929234	1	rs1130409	Chr14:19994994	D148E	-
ERCC1	Broad13929242	1		Chr19:50608418	3' flanking	√
ERCC1	Broad13929243	1	rs3212978	Chr19:50608427	3' flanking	-
ERCC1	Broad13929252	1		Chr19:50608997	Intron	√
ERCC1	Broad13931087	1	rs41558212	Chr19:50610085	Intron	-
ERCC1	Broad13931088	1	rs41561512	Chr19:50610095	Intron	-
ERCC1	Broad13929279	1	rs3212961	Chr19:50614163	Intron	-
ERCC1	Broad13929284	1		Chr19:50614306	Intron	√
ERCC1	Broad13929287	1	rs3212955	Chr19:50615336	Intron	-
ERCC1	Broad13929290	1	rs11615	Chr19:50615493	SYN	-
ERCC1	Broad13929293	1		Chr19:50615648	Intron	√
ERCC1	Broad13929299	1	rs3212948	Chr19:50616202	Intron	-
ERCC1	Broad13929309	1	rs3212947	Chr19:50616372	SYN	-
ERCC1	Broad13961070	1		Chr19:50616564	Intron	√
ERCC1	Broad13929319	3		Chr19:50616574	Intron	√
ERCC1	Broad13929331	1	rs2298881	Chr19:50618756	Intron	-
ERCC1	Broad13929334	1	rs41559012	Chr19:50618890	5'UTR	-
ERCC1	Broad13929335	2	rs3212931	Chr19:50619155	5' flanking	-
MLH1	Broad11527987	1	rs1800734	Chr3:37009950	5' flanking	-
MLH1	Broad13929352	1	rs56198082	Chr3:37010015	5'UTR	-
MLH1	Broad12535601	1		Chr3:37010036	5'UTR	√
MLH1	Broad11528025	1		Chr3:37013005	Intron	√
MLH1	Broad11528256	1	rs4234259	Chr3:37023637	Intron	-
MLH1	Broad11528349	1	rs4647255	Chr3:37025258	Intron	-
MLH1	Broad11528512	1	rs1799977	Chr3:37028572	I219V	-
MLH1	Broad13929360	1		Chr3:37030988	S247A	√
MLH1	Broad11528611	1		Chr3:37033922	Intron	√
MLH1	Broad11528699	1		Chr3:37036724	Intron	√

Supplementary Table III-9, continued.

Gene	SNP ID	Detection category [a]	Variation [b]	Position (Build 36)	Classification	Novel [c]
MLH1	Broad11528740	1	rs2286939	Chr3:37037044	Intron	-
MLH1	Broad13369339	1		Chr3:37042319	Q409P	√
MLH1	Broad11529160	1	rs2286940	Chr3:37045110	Intron	-
MLH1	Broad11529172	1	rs41562513	Chr3:37045441	Intron	-
MLH1	Broad12536365	1		Chr3:37058728	Intron	√
MLH1	Broad11529222	1	rs9876116	Chr3:37058744	Intron	-
MLH1	Broad11529268	1	SWISS:VAR_004462	Chr3:37064134-5	K618A	-
MLH1	Broad13771867	1		Chr3:37065027	G638R	√
MLH1	Broad11529371	1	rs1800146	Chr3:37065074	SYN	-
MLH1	Broad11529408	1	rs2241031	Chr3:37065278	Intron	-
MLH1	Broad13929365	1	rs35831931	Chr3:37067023	V716M	-
MLH1	Broad13929366	1		Chr3:37067057	H727L	√
MLH1	Broad11529639	3		Chr3:37067429	3' flanking	√
MRE11	Broad13928899	1	rs13447762	Chr11:93790066	3' flanking	-
MRE11	Broad13928901	1	rs2155209	Chr11:93790438	3'UTR	-
MRE11	Broad13928907	1		Chr11:93791006	3'UTR	√
MRE11	Broad13928908	1		Chr11:93791163	3'UTR	√
MRE11	Broad13928909	2		Chr11:93791193	3'UTR	√
MRE11	Broad13928911	1		Chr11:93791711	3'UTR	√
MRE11	Broad13932853	3		Chr11:93792254	3'UTR	√
MRE11	Broad13928918	3	rs591959	Chr11:93792279	3'UTR	-
MRE11	Broad13928919	1	rs11020777	Chr11:93792281	3'UTR	-
MRE11	Broad13928920	1		Chr11:93792283	3'UTR	√
MRE11	Broad13928924	1	rs13447749	Chr11:93792428	3'UTR	-
MRE11	Broad13928926	1		Chr11:93792508	3'UTR	√
MRE11	Broad13928927	1		Chr11:93792725	3'UTR	√
MRE11	Broad13928931	1		Chr11:93802894	Intron	√
MRE11	Broad13928932	1		Chr11:93802911	Intron	√
MRE11	Broad13928946	1	rs1014666	Chr11:93818773	Intron	-
MRE11	Broad13928947	1		Chr11:93818823	Intron	√
MRE11	Broad13928950	1	rs13447696	Chr11:93819873	Intron	-
MRE11	Broad13928953	1	rs13447695	Chr11:93819914	Intron	-
MRE11	Broad13928968	1		Chr11:93832054	Intron	√
MRE11	Broad13928970	1		Chr11:93832242	E494K	√
MRE11	Broad13928971	1	rs61749249	Chr11:93832247	A492D	-
MRE11	Broad13928974	1	rs529126	Chr11:93833951	Intron	-
MRE11	Broad13928975	1	rs641936	Chr11:93836908	Intron	-
MRE11	Broad13928976	1	rs640627	Chr11:93837216	Intron	-
MRE11	Broad13928977	1		Chr11:93837270	Intron	√
MRE11	Broad13928978	1		Chr11:93840496	Intron	√
MRE11	Broad13928989	1	rs610611	Chr11:93843516	Intron	-
MRE11	Broad13928991	1		Chr11:93844302	Intron	√
MRE11	Broad13928992	1	rs13447623	Chr11:93848996	Intron	-
MRE11	Broad13928997	1	rs535801	Chr11:93851696	Intron	-
MRE11	Broad13928998	1	rs680695	Chr11:93851802	Intron	-
MRE11	Broad13929006	1		Chr11:93863879	Intron	√
MRE11	Broad13929007	1	rs496797	Chr11:93865455	Intron	-
MRE11	Broad13929008	1	rs497763	Chr11:93865568	Intron	-
MRE11	Broad13929010	1	rs1805363	Chr11:93866600	5'UTR	-
MRE11	Broad13929012	1	rs11020802	Chr11:93866773	5' flanking	-

Supplementary Table III-9, continued.

Gene	SNP ID	Detection category [a]	Variation [b]	Position (Build 36)	Classification	Novel [c]
MSH2	Broad11460725	1	rs2303425	Chr2:47483717	5' flanking	-
MSH2	Broad11460793	1	rs2303426	Chr2:47484054	Intron	-
MSH2	Broad11460829	1	rs3815865	Chr2:47484143	Intron	-
MSH2	Broad13929340	2	rs17217758	Chr2:47490569	Intron	-
MSH2	Broad13761443	1		Chr2:47490629	Intron	√
MSH2	Broad11461066	1	rs17217765	Chr2:47490651	Intron	-
MSH2	Broad11461385	2		Chr2:47494993	T292S	√
MSH2	Broad11461569	1	rs4987188	Chr2:47496961	G322D	-
MSH2	Broad13882428	2		Chr2:47497248	Intron	√
MSH2	Broad11461724	1	rs1981929	Chr2:47526073	Intron	-
MSH2	Broad11461796	1	rs17224444	Chr2:47526404	Intron	-
MSH2	Broad11461840	1		Chr2:47543409	Intron	√
MSH2	Broad11461842	1	rs6741393	Chr2:47543424	Intron	-
MSH2	Broad11461912	1	rs3771278	Chr2:47543908	Intron	-
MSH2	Broad11461914	1	rs3771279	Chr2:47543912	Intron	-
MSH2	Broad11461917	1	rs3771280	Chr2:47543915	Intron	-
MSH2	Broad11461954	1	rs3732182	Chr2:47547210	Intron	-
MSH2	Broad11461995	1	rs12998837	Chr2:47547292	Intron	-
MSH2	Broad11462036	1	rs3732183	Chr2:47547463	Intron	-
MSH2	Broad11462049	1	rs10183143	Chr2:47547541	Intron	-
MSH2	Broad13882449	2		Chr2:47547627	Intron	√
MSH2	Broad11462123	1	rs17218363	Chr2:47551762	Intron	-
MSH2	Broad11462137	1	rs3764959	Chr2:47551812	Intron	-
MSH2	Broad11462151	1	rs3764960	Chr2:47551888	Intron	-
MSH2	Broad13929344	2	rs17218439	Chr2:47555606	Intron	-
MSH2	Broad12520396	3	rs17218446	Chr2:47555955	Intron	-
MSH2	Broad11462187	1	rs2303428	Chr2:47557004	Intron	-
MSH2	Broad11462232	1		Chr2:47557183	A727S	√
MSH2	Broad11462291	1	rs4583514	Chr2:47557389	Intron	-
MSH2	Broad11462544	3		Chr2:47563653	3'UTR	√
MSH2	Broad11462575	1	rs17225060	Chr2:47563818	3'UTR	-
MSH2	Broad13929347	1	rs56116962	Chr2:47563935	3' flanking	-
MSH2	Broad13929348	1		Chr2:47563979	3' flanking	√
MSH2	Broad13929349	1		Chr2:47563984	3' flanking	√
NBS1	Broad13929663	1	rs14448	Chr8:91015009	3'UTR	-
NBS1	Broad13929665	1	rs9995	Chr8:91015232	3'UTR	-
NBS1	Broad13929666	1	rs13312986	Chr8:91015294	3'UTR	-
NBS1	Broad13929671	1	rs1063054	Chr8:91015777	3'UTR	-
NBS1	Broad13929678	1	rs2735383	Chr8:91016445	3'UTR	-
NBS1	Broad13929681	1		Chr8:91016585	3'UTR	√
NBS1	Broad13929683	1	rs1063053	Chr8:91016713	3'UTR	-
NBS1	Broad13929687	1	rs13312971	Chr8:91018273	Intron	-
NBS1	Broad13929688	1	rs13312970	Chr8:91018342	Intron	-
NBS1	Broad13929691	1	rs13312969	Chr8:91018601	Intron	-
NBS1	Broad13929696	1	rs7840099	Chr8:91024534	Intron	-
NBS1	Broad13929697	2	rs3736639	Chr8:91024800	Intron	-
NBS1	Broad13929698	1	rs1061302	Chr8:91027598	SYN	-
NBS1	Broad13929699	1	rs2308962	Chr8:91027706	Intron	-
NBS1	Broad13931239	1		Chr8:91034914	D527Y	√
NBS1	Broad13929705	1		Chr8:91036822	L421S	√
NBS1	Broad13929706	1		Chr8:91036882	P401R	√
NBS1	Broad13929707	1	rs709816	Chr8:91036887	SYN	-

Supplementary Table III-9, continued.

Gene	SNP ID	Detection category [a]	Variation [b]	Position (Build 36)	Classification	Novel [c]
NBS1	Broad13929709	1	rs1805786	Chr8:91037038	Intron	-
NBS1	Broad13929712	1	rs1805818	Chr8:91040038	Intron	-
NBS1	Broad13929714	1	rs2234744	Chr8:91040111	Intron	-
NBS1	Broad13929717	1	rs1805824	Chr8:91046082	Intron	-
NBS1	Broad13929719	1	rs1805826	Chr8:91051732	Intron	-
NBS1	Broad13929722	1	rs769418	Chr8:91051979	Intron	-
NBS1	Broad13929723	1		Chr8:91051990	Intron	√
NBS1	Broad13929727	1	rs3026271	Chr8:91052428	Intron	-
NBS1	Broad13929728	1		Chr8:91052451	Intron	√
NBS1	Broad13929730	1		Chr8:91052493	Intron	√
NBS1	Broad13929731	1	SWISS:VAR_025798	Chr8:91052651	V210F	-
NBS1	Broad13929733	1	rs1805794	Chr8:91059655	E185Q	-
NBS1	Broad13929735	1		Chr8:91059895	Intron	√
NBS1	Broad13929737	1	rs61754795	Chr8:91062237	SYN	-
NBS1	Broad13929738	1	rs1805797	Chr8:91063059	Intron	-
NBS1	Broad13929741	1		Chr8:91063375	5' UTR	√
NBS1	Broad13931251	1		Chr8:91063456	Intron	√
NBS1	Broad13929743	1	rs1063045	Chr8:91064195	SYN	-
RAD50	Broad13929372	1		Chr5:131920564	5'UTR	√
RAD50	Broad13929374	1		Chr5:131920632	5'UTR	√
RAD50	Broad13929375	1	rs4526098	Chr5:131920878	5'UTR	-
RAD50	Broad13929379	3		Chr5:131923031	Intron	√
RAD50	Broad13929380	1		Chr5:131939324	Intron	√
RAD50	Broad13929384	1		Chr5:131942745	Intron	√
RAD50	Broad13929385	1		Chr5:131943036	P165H	√
RAD50	Broad13929386	1	rs17166050	Chr5:131943112	Intron	-
RAD50	Broad13929387	1		Chr5:131943178	Intron	√
RAD50	Broad13929388	1	rs2522403	Chr5:131943216	Intron	-
RAD50	Broad13931118	1	rs28903091	Chr5:131951609	R327H	-
RAD50	Broad13929401	1		Chr5:131952240	Intron	√
RAD50	Broad13929402	1		Chr5:131952340	Q372X	√
RAD50	Broad13929408	1	rs56798121	Chr5:131953073	Intron	-
RAD50	Broad13929409	1	rs2706362	Chr5:131953086	Intron	-
RAD50	Broad13929419	1	rs2706377	Chr5:131967396	Intron	-
RAD50	Broad13929420	1	rs12187537	Chr5:131967803	Intron	-
RAD50	Broad13931139	1	rs2522391	Chr5:131968240	Intron	-
RAD50	Broad13929431	1		Chr5:131972175	Intron	√
RAD50	Broad13929439	1		Chr5:131979756	Intron	√
RAD50	Broad13929440	1	rs35191767	Chr5:131979770	Intron	-
RAD50	Broad13929441	1	rs2301713	Chr5:131979895	Intron	-
RAD50	Broad13931152	4		Chr5:131981506	Intron	√
RAD50	Broad13929446	1		Chr5:131981657	Intron	√
RAD50	Broad13929448	1		Chr5:131981726	R1077Q	√
RAD50	Broad13929449	1	rs2074369	Chr5:132001562	Intron	-
RAD50	Broad13929450	1		Chr5:132001904	Intron	√
RAD50	Broad13929451	1	rs7737470	Chr5:132001962	Intron	-
RAD50	Broad13929454	3		Chr5:132004328	SYN	√
RAD52	Broad13929014	1	rs10849584	Chr12:891420	3' flanking	-
RAD52	Broad13929016	1	rs56131726	Chr12:891449	3' flanking	-
RAD52	Broad13929026	4		Chr12:891624	3'UTR	√
RAD52	Broad13929027	1	rs11571487	Chr12:891643	3'UTR	-
RAD52	Broad13929032	1	rs7963551	Chr12:891776	3'UTR	-
RAD52	Broad13929033	1	rs1051672	Chr12:891818	3'UTR	-

Supplementary Table III-9, continued.

Gene	SNP ID	Detection category [a]	Variation [b]	Position (Build 36)	Classification	Novel [c]
RAD52	Broad13961032	3		Chr12:891826	3'UTR	√
RAD52	Broad13929046	2	rs11226	Chr12:892074	3'UTR	-
RAD52	Broad13929054	2	rs11571478	Chr12:892173	3'UTR	-
RAD52	Broad13929064	1		Chr12:892350	3'UTR	√
RAD52	Broad13929066	2		Chr12:892365	3'UTR	√
RAD52	Broad13929067	1	rs7310449	Chr12:892376	3'UTR	-
RAD52	Broad13929069	1		Chr12:892423	3'UTR	√
RAD52	Broad13929070	1	rs7301931	Chr12:892445	3'UTR	-
RAD52	Broad13929072	1		Chr12:892493	3'UTR	√
RAD52	Broad13929073	1		Chr12:892494	3'UTR	√
RAD52	Broad13929074	1		Chr12:892496	3'UTR	√
RAD52	Broad13929075	1		Chr12:892497	3'UTR	√
RAD52	Broad13929076	1		Chr12:892498	3'UTR	√
RAD52	Broad13929077	1		Chr12:892500	3'UTR	√
RAD52	Broad13929078	1		Chr12:892501	3'UTR	√
RAD52	Broad13929079	1		Chr12:892502	3'UTR	√
RAD52	Broad13929080	1		Chr12:892503	3'UTR	√
RAD52	Broad13929081	1	rs11571475	Chr12:892613	3'UTR	-
RAD52	Broad13929083	1	rs1051669	Chr12:892713	3'UTR	-
RAD52	Broad13929084	1	rs11571474	Chr12:892714	3'UTR	-
RAD52	Broad13929085	1	rs4987208	Chr12:892830	3'UTR	-
RAD52	Broad13929086	1	rs6413436	Chr12:892940	3'UTR	-
RAD52	Broad13929090	1	rs4987207	Chr12:893479	3'UTR	-
RAD52	Broad13929091	2		Chr12:893723	Intron	√
RAD52	Broad13929099	1	rs28910277	Chr12:893931	3'UTR	-
RAD52	Broad13929123	1	rs11571446	Chr12:906078	Intron	-
RAD52	Broad13929124	1	rs7487683	Chr12:906303	G180R	-
RAD52	Broad13929126	1		Chr12:906486	Intron	√
RAD52	Broad13929128	1		Chr12:906551	Intron	√
RAD52	Broad13929132	1	rs4766377	Chr12:909003	Intron	-
RAD52	Broad13929133	1	rs11571425	Chr12:909027	Intron	-
RAD52	Broad13929134	1	rs11571424	Chr12:909115	Intron	-
RAD52	Broad13929137	1	rs11571420	Chr12:909596	Intron	-
RAD52	Broad13929139	1	rs11571409	Chr12:910480	Intron	-
RAD52	Broad13929141	4		Chr12:910558	Intron	√
RAD52	Broad13929143	1	rs7303748	Chr12:910634	Intron	-
RAD52	Broad13929165	2	rs3748522	Chr12:928949	Intron	-
RAD52	Broad13929167	2	rs3748523	Chr12:928999	Intron	-

a) Description of detection categories

Category	Confidence	Description
1	High	One or more PolyPhred calls at PolyPhred SNP Score 99 (regardless of local read quality)
2	High	One or more PolyPhred calls at PolyPhred SNP Score 60 **and** one or more PolyDhan calls
3	High	One or more PolyPhred calls with >= PolyPhred SNP Score 95 with average local read quality >= q30
4	High	One or more PolyPhred calls with >= PolyPhred SNP Score 90 with average local read quality >= q40
5	Low	All other calls

b) SNP ID from dbSNP or SwissProt where available

c) A check (√) denotes a novel SNP.

Supplemental Table III-10. Genetic association of SNPs identified by reseqencing. Non-synonymous SNPs are depicted as the amino acid changes they encode; synonymous SNPs, SYN; Intronic SNPs, Intron; SNPs located in the 5' untranslated region, 5' UTR; SNPs located in the 3' untranslated region, 3' UTR. Significant p values (<0.05) are highlighted in bold. Bold allele frequency values indicate that the SNP is detected in only that group.

			Controls	IgAD		CVID	
Gene	SNP	Allele	Allele frequency	Allele frequency	p-value	Allele frequency	p-value
APEX1	rs11622131	5' UTR	15/1892	9/650	0.18	0/226	0.18
APEX1	rs41561214	5' UTR	59/1902	18/658	0.64	3/228	0.13
APEX1	rs2307490	5' UTR	3/1872	2/642	0.46	2/230	**0.037**
APEX1	rs2307485	Intron (5)	1/1868	2/640	0.10	0/230	0.73
APEX1	rs1048945	Q51H	60/1840	12/626	0.085	9/212	0.45
MLH1	rs56198082	5' UTR	9/1904	6/658	0.20	1/230	0.94
MLH1	Broad12535601	5' UTR	8/1876	6/644	0.14	1/230	0.99
MLH1	rs1799977	I219V	563/1904	187/656	0.61	79/222	0.065
MLH1	Broad13929360	S247A	0/1876	2/642	**0.016**	0/230	--
MLH1	Broad13369339	Q409P	0/1902	0/656	--	**1/230**	**0.0040**
MLH1	Broad11529268	K618A	13/1882	5/646	0.83	2/230	0.76
MLH1	Broad13771867	G638R	0/1872	**1/642**	0.088	0/230	--
MLH1	rs1800146	SYN	17/1876	10/648	0.17	2/228	0.97
MLH1	rs35831931	V716M	3/1904	1/658	0.98	1/230	0.36
MLH1	Broad13929366	H727L	1/1872	0/642	0.56	1/228	0.075
MRE11	Broad13928907	3' UTR	2/1866	2/636	0.26	3/228	**0.00040**
MRE11	Broad13928908	3' UTR	1/1898	0/654	0.56	1/230	0.074
MRE11	Broad13928911	3' UTR	2/1878	1/642	0.75	0/230	0.62
MRE11	rs13447749	3' UTR	106/1870	25/602	0.15	13/216	0.83
MRE11	Broad13928926	3' UTR	0/1864	0/636	--	**1/228**	**0.0042**
MRE11	Broad13928970	A492D	2/1898	3/652	0.077	1/230	0.21
MRE11	Broad13928971	E494K	8/1876	0/642	0.098	2/230	0.36
MSH2	Broad11461385	T292S	1/1890	2/652	0.10	0/230	0.73
MSH2	Broad11462232	A727S	0/1888	**1/650**	0.088	0/230	--
MSH2	rs17225060	3' UTR	2/1892	2/646	0.26	1/228	0.21
NBS1	Broad13931239	D527Y	0/1844	**1/598**	0.079	0/212	--
NBS1	Broad13929705	L421S	1/1884	0/644	0.56	1/230	0.076
NBS1	Broad13929706	P401R	0/1892	**2/646**	**0.016**	0/230	--
NBS1	Broad13929731	V210F	4/1858	2/638	0.66	0/230	0.48
NBS1	rs1805794	E185Q	625/1892	207/650	0.58	77/224	0.69
NBS1	Broad13929737	SYN	1/1872	1/642	0.43	1/230	0.077
NBS1	Broad13931251	Intron	0/1814	**1/602**	0.083	0/226	--
RAD50	Broad13929372	5' UTR	0/1842	**1/640**	0.090	0/230	--
RAD50	Broad13929374	5' UTR	0/1900	**1/658**	0.089	0/230	--
RAD50	rs4526098	5' UTR	3/1868	3/640	0.17	0/230	0.54
RAD50	Broad13929385	P165H	0/1842	**1/640**	0.0897	0/230	--
RAD50	rs28903091	R327H	4/1826	1/584	0.83	0/214	0.49
RAD50	Broad13929402	Q372X	0/1900	0/654	--	**1/230**	**0.0040**
RAD50	Broad13929448	R1077Q	0/1856	**1/596**	0.078	0/216	--
RAD50	Broad13929454	SYN	0/1900	**1/652**	0.088	0/230	--
RAD52	rs7487683	G180R	74/1876	17/644	0.13	9/230	0.98

CHAPTER IV: THESIS SUMMARY AND CONCLUDING DISCUSSION

Failure to mount an effective antibody-mediated immune response is the hallmark of primary immunodeficiency syndromes including the hyper-IgM syndromes, CVID, and IgAD. This failure could be due to defects in a number of different pathways. First, mutations in the DNA metabolism machinery responsible for CSR, such as AID and UNG2, manifest as antibody deficiency. Second, defects in signaling pathways that act upstream of CSR, such as CD40LG binding of CD40, can also lead to impaired antibody production. Finally, disruption of signaling pathways necessary for B cell differentiation into antibody-secreting plasma cells, such as CD19, can also result in decreased serum antibody levels and immunodeficiency. The fact that these mutations occur in multiple pathways highlights the multi-factoral nature of primary antibody-deficiency syndromes as a whole.

A genetic cause(s) has yet to be defined for the vast majority of primary antibody deficiency syndromes. Those that have an identified cause are extremely rare disorders. For instance Hyper-IgM2, Hyper-IgM3, and Hyper-IgM1, due to mutations in AID, CD40, and CD40LG, have an estimated prevalence of approximately 1:2,000,000, 1:2,000,000, and 1:1,000,000 (males), respectively [ImmunoDeficiency Resource (IDR), http://bioinf.uta.fi/xml/idr/ff/FF17.xml & ./FF16.xml]. It is estimated that approximately 10% of CVID cases are attributable to mutations in TACI, and a handful of others (likely <1%) can be explained by deficiencies of ICOS or CD19. In contrast, no genetic causes have been directly demonstrated for IgAD, the most common immunodeficiency with a frequency of approximately 1:600 people of European descent (Koistinen 1975).

The overall goal of this thesis has been to identify and characterize novel genetic variations that associate with the relatively innocuous primary antibody deficiency syndromes CVID and IgAD. Towards this goal, genetic association studies were performed using markers in multiple candidate pathways. To look for putative causal alleles linked with associated markers, candidate genes were subsequently sequenced in CVID/IgAD patients. Using this

method, disease-correlated non-synonymous variations were detected in the B cell surface receptor TACI, the homologous recombination protein MSH5, and the DNA metabolism pathways of DNA mismatch repair and the multi-functional MRE11-RAD50-NBS1 complex.

The genetic associations and enrichment of novel mutant alleles we observed in CVID/IgAD patients suggests that these genes are important for proper antibody production. However, without functional confirmation of the impact of these mutations, we cannot conclude that they are disease-causing alleles or even genetic indicators of disease. The reporting of these alleles should serve as the foundation for future studies to characterize their potential impact on CSR.

TACI as a regulator of the B cell response

Overall, the mutations identified in TACI are present in nearly 10% of patients with CVID, making it the most prevalent primary immunodeficiency-associated gene (Pan-Hammarström and Hammarström 2008; Schaffer et al. 2007; Yong et al. 2008). It is important to note, however, that disease alleles of TACI are also present in healthy control populations (albeit at lower frequency) [see **Chapter II, Table II-1, Supplemental Table II-2** and (Pan-Hammarström et al. 2007)]. Furthermore, in family based studies, we noted that some family members carried TACI alleles, but had normal serum antibody levels [(**Chapter II, Supplemental Figure II-1**) (Pan-Hammarström et al. 2007) and also reported by (Warnatz et al. 2009)]. The detection of disease-associated TACI alleles in healthy controls and in healthy relatives of proband CVID/IgAD cases suggests that these disease alleles may exert incomplete penetrance. This phenomenon would fit with the normal disease progression of CVID/IgAD as there is variability in the age of disease onset. Both diseases are occasionally found intermixed within the same family, and in some instances, IgAD has been shown to progress to the more severe CVID in individuals, suggesting that the two diseases may represent differential disease

states of the same condition.

TACI has been suggested to play a paradoxical role in B cell biology, by both stimulating and attenuating the B cell response. Corroborating a link with immune-deficiency in humans, mouse studies indicate that TACI signaling stimulates CSR and increases antibody production (Castigli et al. 2004; Castigli et al. 2005; Litinskiy et al. 2002). This would suggest a stimulatory role in B cell activation, however TACI signaling has also been shown to exert a suppressive effect on B cell activation, which is thought to counteract BAFFR signaling (Mackay et al. 1999; Seshasayee et al. 2003; Stohl 2005). It has recently been reported that in addition to T cell-independent B cell activation, TACI can cooperate with T-dependent signaling to enhance differentiation into non-dividing plasma cells (Castigli et al. 2007; Mantchev et al. 2007). This suggests that prolonged TACI signaling may also enhance differentiation to plasma cells in the T-independent response. This would reduce the overall number of SHM events that occur outside of the controlled environment of the germinal center, where apoptosis can be induced in auto-reactive B cells (Hannum et al. 1996). Thus, by pushing activated peripheral B cells to the plasma cell stage (stimulating the B cell response), TACI signaling may limit the creation of auto-reactive B cells by unchecked SHM outside of the germinal center (attenuating the B cell response).

Genetic variations in multiple DNA repair pathways confer risk for CVID/IgAD

Following our identification of MSH5 alleles associated with CVID, we hypothesized that subtle variations in other DNA metabolism pathways may be enriched in CVID and IgAD. This hypothesis was supported by the contribution of DNA repair gene mutants to other CSR deficiencies (ie, AID and UNG2). We additionally hypothesized that these mutations would be relatively subtle, as severe disruption of certain general DNA repair factors confers a more serious chromosomal instability phenotype with antibody deficiency as a secondary diagnosis.

These disorders include ataxia telangiectasia (AT) due to mutations in ATM (ataxia telangiectasia mutated), ataxia telangiectasia-like disorder (AT-LD) due to mutations in MRE11, Nijmegen breakage syndrome (NBS) due to mutations in NBS1, and PMS2-deficiency, due to mutations in PMS2 (Carney et al. 1998; Concannon and Gatti 1997; Lahdesmaki et al. 2004; Pan et al. 2002; Peron et al. 2008; Savitsky et al. 1995; Stewart et al. 1999; Varon et al. 1998).

Our identification of CVID/IgAD-specific mutations in MSH2, MLH1, RAD50, NBS1, and MRE11 supports our hypothesis that mutations in multiple DNA repair pathways may underpin a subset of primary immunodeficiency syndromes. Some patients carried mutant alleles in multiple DNA repair genes and sometimes even multiple DNA repair pathways. One CVID patient was heterozygous for mutations in MSH2 and MLH1. In one IgAD patient, we found mutations in RAD50 and MLH1, and an additional IgAD patient carried novel mutations in MSH2, MLH1, and NBS. This aggregation of disease alleles in affected individuals suggests that minor variations in multiple genes/pathways can exert an additive effect on the pathogenesis of CSR deficiency. Additionally, this may help to explain the variable penetrance observed for some disease alleles when evaluated independently.

Future experiments to follow up these studies

The advent of extremely high density microarrays and high throughput sequencing have opened the door for a number of genetic studies that would have been previously impossible to perform on the necessary scale to detect associated alleles in a multi-factoral disease. Currently, the most-straightforward means to screen for non-synonymous single nucleotide variations is using the recently developed RNA-seq technology. This would provide both whole-transcriptome gene expression profiling (as customized micro-arrays are used to capture target RNA) and single nucleotide variation information simultaneously.

To fully identify the network of pathways that underlie immunodeficiency syndromes relating to CSR, we need to begin looking at additional genomic variation beyond polymorphisms that directly encode for changes in the amino-acid sequence of a protein. While mutations in binding sites for *cis*-regulatory elements, which affect expression by controlling transcription, only account for a small subset of all disease alleles identified to date, they have only been examined in a small minority of human diseases. Additionally, conventional means of mapping and characterizing epigenetic DNA modifications are severely limited in their ability to comprehensively look at multiple factors across large portions of the genome. By using chromatin immunoprecipitation (ChIP) coupled with high-throughput sequencing (ChIP-Seq) the epigenetic states of both healthy and immune-deficient patients could be mapped. The most straightforward use of this technology would be to map differences in transcription factor binding sites, however there is a very narrow window of opportunity following stimulation of a B cell and completion of CSR (and it is not possible to synchronize a culture of B cells to undergo CSR simultaneously). Therefore this approach will not be feasible until a defined *in vitro* model system of CSR is established. In the meantime, a number of global genome maps have already been generated for many transcription factors using this technology [reviewed by (Farnham 2009)]. This information could be used to generate micro-arrays to capture the binding sites of specific (previously mapped) transcriptions factors, thus eliminating the need to catch the cell at just the right epigenetic state.

A final interesting tangent from this research would be the study of potential links between immune-deficiency and cancer. B cells in a significant number of CVID/IgAD patients exhibit radiation sensitivity, consistent with a role of DNA repair pathways in CSR (Palanduz et al. 1998; Vorechovsky et al. 1993). An increased incidence of cancer is also observed in some CVID populations, further supporting a link between the two conditions (Mellemkjaer et al. 2002; Zenone and Souillet 1997). Examining the functional consequences of the DNA repair

alleles identified in individuals with CVID/IgAD will provide key insight into their potential roles in antibody deficiency as well as other widespread conditions, such as cancer.

REFERENCES

Agrawal A. and D. G. Schatz (1997). RAG1 and RAG2 form a stable postcleavage synaptic complex with DNA containing signal ends in V(D)J recombination. Cell **89**(1): 43-53.

Alper C. A., D. Marcus-Bagley, Z. Awdeh, M. S. Kruskall, G. S. Eisenbarth, S. J. Brink, A. J. Katz, R. Stein, D. H. Bing, E. J. Yunis and P. H. Schur (2000). Prospective analysis suggests susceptibility genes for deficiencies of IgA and several other immunoglobulins on the [HLA-B8, SC01, DR3] conserved extended haplotype. Tissue Antigens **56**(3): 207-216.

Arakawa H. and J. M. Buerstedde (2004). Immunoglobulin gene conversion: insights from bursal B cells and the DT40 cell line. Dev Dyn **229**(3): 458-464.

Arakawa H., J. Hauschild and J. M. Buerstedde (2002). Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. Science **295**(5558): 1301-1306.

Aruffo A., M. Farrington, D. Hollenbaugh, X. Li, A. Milatovich, S. Nonoyama, J. Bajorath, L. S. Grosmaire, R. Stenkamp, M. Neubauer and et al. (1993). The CD40 ligand, gp39, is defective in activated T cells from patients with X-linked hyper-IgM syndrome. Cell **72**(2): 291-300.

Ban C., M. Junop and W. Yang (1999). Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. Cell **97**(1): 85-97.

Barnetson R. A., N. Cartwright, A. van Vliet, N. Haq, K. Drew, S. Farrington, N. Williams, J. Warner, H. Campbell, M. E. Porteous and M. G. Dunlop (2008). Classification of ambiguous mutations in DNA mismatch repair genes identified in a population-based study of colorectal cancer. Hum Mutat **29**(3): 367-374.

Barreto V., B. Reina-San-Martin, A. R. Ramiro, K. M. McBride and M. C. Nussenzweig (2003). C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. Mol Cell **12**(2): 501-508.

Barrett J. C., B. Fry, J. Maller and M. J. Daly (2005). Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics **21**(2): 263-265.

Batista F. D. and N. E. Harwood (2009). The who, how and where of antigen presentation to B cells. Nat Rev Immunol **9**(1): 15-27.

Blunt T., N. J. Finnie, G. E. Taccioli, G. C. Smith, J. Demengeot, T. M. Gottlieb, R. Mizuta, A. J. Varghese, F. W. Alt, P. A. Jeggo and et al. (1995). Defective DNA-dependent protein kinase activity is linked to V(D)J recombination and DNA repair defects associated with the murine scid mutation. Cell **80**(5): 813-823.

Bransteitter R., P. Pham, M. D. Scharff and M. F. Goodman (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.

Briere F., J. M. Bridon, D. Chevet, G. Souillet, F. Bienvenu, C. Guret, H. Martinez-Valdez and J. Banchereau (1994). Interleukin 10 induces B lymphocytes from IgA-deficient patients to secrete IgA. J Clin Invest **94**(1): 97-104.

Buerstedde J. M. and S. Takeda (1991). Increased ratio of targeted to random integration after transfection of chicken B cell lines. Cell **67**(1): 179-188.

Burrows P. D. and M. D. Cooper (1997). IgA deficiency. Adv Immunol **65**: 245-276.

Cambier J. C., C. M. Pleiman and M. R. Clark (1994). Signal transduction by the B cell antigen receptor and its coreceptors. Annu Rev Immunol **12**: 457-486.

Cancro M. P. (2004). The BLyS family of ligands and receptors: an archetype for niche-specific homeostatic regulation. Immunol Rev **202**: 237-249.

Cao M., L. Chen, X. X. Shan and S. Q. Zhang (2005). Immunological effects of refolded human soluble BAFF synthesized in Escherichia coli on murine B lymphocytes in vitro and in vivo. Jpn J Physiol **55**(4): 221-227.

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

Carter R. H. and D. T. Fearon (1992). CD19: lowering the threshold for antigen receptor stimulation of B lymphocytes. Science **256**(5053): 105-107.

Carvalho Neves Forte W., F. Ferreira De Carvalho Junior, N. Damaceno, F. Vidal Perez, C. Gonzales Lopes and R. A. Mastroti (2000). Evolution of IgA deficiency to IgG subclass deficiency and common variable immunodeficiency. Allergol Immunopathol (Madr) **28**(1): 18-20.

Cascalho M., J. Wong, C. Steinberg and M. Wabl (1998). Mismatch repair co-opted by hypermutation. Science **279**(5354): 1207-1210.

Castigli E. and R. S. Geha (2006). Molecular basis of common variable immunodeficiency. J Allergy Clin Immunol **117**(4): 740-746; quiz 747.

Castigli E., S. Scott, F. Dedeoglu, P. Bryce, H. Jabara, A. K. Bhan, E. Mizoguchi and R. S. Geha (2004). Impaired IgA class switching in APRIL-deficient mice. Proc Natl Acad Sci U S A **101**(11): 3903-3908.

Castigli E., S. Wilson, L. Garibyan, R. Rachid, F. Bonilla, L. Schneider, M. Morra, J. Curran and R. Geha (2007). Reexamining the role of TACI coding variants in common variable immunodeficiency and selective IgA deficiency. Nat Genet **39**(4): 430-431.

Castigli E., S. A. Wilson, A. Elkhal, E. Ozcan, L. Garibyan and R. S. Geha (2007). Transmembrane activator and calcium modulator and cyclophilin ligand interactor enhances CD40-driven plasma cell differentiation. J Allergy Clin Immunol **120**(4): 885-891.

Castigli E., S. A. Wilson, L. Garibyan, R. Rachid, F. Bonilla, L. Schneider and R. S. Geha (2005). TACI is mutant in common variable immunodeficiency and IgA deficiency. Nat Genet **37**(8): 829-834.

Castigli E., S. A. Wilson, S. Scott, F. Dedeoglu, S. Xu, K. P. Lam, R. J. Bram, H. Jabara and R. S. Geha (2005). TACI and BAFF-R mediate isotype switching in B cells. J Exp Med **201**(1): 35-39.

Chaudhuri J. and F. W. Alt (2004). Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. Nat Rev Immunol **4**(7): 541-552.

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

Concannon P. and R. A. Gatti (1997). Diversity of ATM gene mutations detected in patients with ataxia-telangiectasia. Hum Mutat **10**(2): 100-107.

Conticello S. G. (2008). The AID/APOBEC family of nucleic acid mutators. Genome Biol **9**(6): 229.

Cook G. P. and I. M. Tomlinson (1995). The human immunoglobulin VH repertoire. Immunol Today **16**(5): 237-242.

Craxton A., K. E. Draves, A. Gruppi and E. A. Clark (2005). BAFF regulates B cell survival by downregulating the BH3-only family member Bim via the ERK pathway. J Exp Med **202**(10): 1363-1374.

Criswell L. A., K. A. Pfeiffer, R. F. Lum, B. Gonzales, J. Novitzke, M. Kern, K. L. Moser, A. B. Begovich, V. E. Carlton, W. Li, A. T. Lee, W. Ortmann, T. W. Behrens and P. K. Gregersen (2005). Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. Am J Hum Genet **76**(4): 561-571.

Cunningham-Rundles C. (2001). Physiology of IgA and IgA deficiency. J Clin Immunol **21**(5): 303-309.

Cunningham-Rundles C. and C. Bodian (1999). Common variable immunodeficiency: clinical and immunological features of 248 patients. Clin Immunol **92**(1): 34-48.

Day E. S., T. G. Cachero, F. Qian, Y. Sun, D. Wen, M. Pelletier, Y. M. Hsu and A. Whitty (2005). Selectivity of BAFF/BLyS and APRIL for binding to the TNF family receptors BAFFR/BR3 and BCMA. Biochemistry **44**(6): 1919-1931.

de Bakker P. I., R. Yelensky, I. Pe'er, S. B. Gabriel, M. J. Daly and D. Altshuler (2005). Efficiency and power in genetic association studies. Nat Genet **37**(11): 1217-1223.

de Jager M., J. van Noort, D. C. van Gent, C. Dekker, R. Kanaar and C. Wyman (2001). Human Rad50/Mre11 is a flexible complex that can tether DNA ends. Mol Cell **8**(5): 1129-1135.

de Vries S. S., E. B. Baart, M. Dekker, A. Siezen, D. G. de Rooij, P. de Boer and H. te Riele (1999). Mouse MutS-like protein Msh5 is required for proper chromosome synapsis in male and female meiosis. Genes Dev **13**(5): 523-531.

DeFranco A. L., J. D. Richards, J. H. Blum, T. L. Stevens, D. A. Law, V. W. Chan, S. K. Datta, S. P. Foy, S. L. Hourihane, M. R. Gold and et al. (1995). Signal transduction by the B-cell antigen receptor. Ann N Y Acad Sci **766**: 195-201.

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

Di Noia J. M. and M. S. Neuberger (2004). Immunoglobulin gene conversion in chicken DT40 cells largely proceeds through an abasic site intermediate generated by excision of the uracil produced by AID-mediated deoxycytidine deamination. Eur J Immunol **34**(2): 504-508.

Di Renzo M., A. L. Pasqui and A. Auteri (2004). Common variable immunodeficiency: a review. Clin Exp Med **3**(4): 211-217.

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

Dunnick W., G. Z. Hertz, L. Scappino and C. Gritzmacher (1993). DNA sequences at immunoglobulin switch region recombination sites. Nucleic Acids Res **21**(3): 365-372.

Durandy A., S. Peron and A. Fischer (2006). Hyper-IgM syndromes. Curr Opin Rheumatol **18**(4): 369-376.

Durandy A., N. Taubenheim, S. Peron and A. Fischer (2007). Pathophysiology of B-cell intrinsic immunoglobulin class switch recombination deficiencies. Adv Immunol **94**: 275-306.

Early P., H. Huang, M. Davis, K. Calame and L. Hood (1980). An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH. Cell **19**(4): 981-992.

Edelmann W., P. E. Cohen, B. Kneitz, N. Winand, M. Lia, J. Heyer, R. Kolodner, J. W. Pollard and R. Kucherlapati (1999). Mammalian MutS homologue 5 is required for chromosome pairing in meiosis. Nat Genet **21**(1): 123-127.

Ehrenstein M. R. and M. S. Neuberger (1999). Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation. EMBO J **18**(12): 3484-3490.

Ehrenstein M. R., C. Rada, A. M. Jones, C. Milstein and M. S. Neuberger (2001). Switch junction sequences in PMS2-deficient mice reveal a microhomology-mediated mechanism of Ig class switch recombination. Proc Natl Acad Sci U S A **98**(25): 14553-14558.

Erdos M., A. Durandy and L. Marodi (2005). Genetically acquired class-switch recombination defects: the multi-faced hyper-IgM syndrome. Immunol Lett **97**(1): 1-6.

Espanol T., M. Catala, M. Hernandez, I. Caragol and J. M. Bertran (1996). Development of a common variable immunodeficiency in IgA-deficient patients. Clin Immunol Immunopathol **80**(3 Pt 1): 333-335.

Farnham P. J. (2009). Insights from genomic profiling of transcription factors. Nat Rev Genet.

Fearon D. T. and R. M. Locksley (1996). The instructive role of innate immunity in the acquired immune response. Science **272**(5258): 50-53.

Ferrari S., S. Giliani, A. Insalaco, A. Al-Ghonaium, A. R. Soresina, M. Loubser, M. A. Avanzini, M. Marconi, R. Badolato, A. G. Ugazio, Y. Levy, N. Catalan, A. Durandy, A. Tbakhi, L. D. Notarangelo and A. Plebani (2001). Mutations of CD40 gene cause an autosomal recessive form of immunodeficiency with hyper IgM. Proc Natl Acad Sci U S A **98**(22): 12614-12619.

Frey S., B. Bertocci, F. Delbos, L. Quint, J. C. Weill and C. A. Reynaud (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.

Fuellen G., M. Spitzer, P. Cullen and S. Lorkowski (2005). Correspondence of function and phylogeny of ABC proteins based on an automated analysis of 20 model protein data sets. Proteins **61**(4): 888-899.

Fujimoto M., Y. Fujimoto, J. C. Poe, P. J. Jansen, C. A. Lowell, A. L. DeFranco and T. F. Tedder (2000). CD19 regulates Src family protein tyrosine kinase activation in B lymphocytes through processive amplification. Immunity **13**(1): 47-57.

Garside P., E. Ingulli, R. R. Merica, J. G. Johnson, R. J. Noelle and M. K. Jenkins (1998). Visualization of specific B and T lymphocyte interactions in the lymph node. Science **281**(5373): 96-99.

Gorelik L., K. Gilbride, M. Dobles, S. L. Kalled, D. Zandman and M. L. Scott (2003). Normal B cell homeostasis requires B cell activation factor production by radiation-resistant cells. J Exp Med **198**(6): 937-945.

Grimbacher B., A. Hutloff, M. Schlesier, E. Glocker, K. Warnatz, R. Drager, H. Eibel, B. Fischer, A. A. Schaffer, H. W. Mages, R. A. Kroczek and H. H. Peter (2003). Homozygous loss of ICOS is associated with adult-onset common variable immunodeficiency. Nat Immunol **4**(3): 261-268.

Gross J. A., S. R. Dillon, S. Mudri, J. Johnston, A. Littau, R. Roque, M. Rixon, O. Schou, K. P. Foley, H. Haugen, S. McMillen, K. Waggie, R. W. Schreckhise, K. Shoemaker, T. Vu, M. Moore, A. Grossman and C. H. Clegg (2001). TACI-Ig neutralizes molecules critical for B cell development and autoimmune disease. impaired B cell maturation in mice lacking BLyS. Immunity **15**(2): 289-302.

Gross J. A., J. Johnston, S. Mudri, R. Enselman, S. R. Dillon, K. Madden, W. Xu, J. Parrish-Novak, D. Foster, C. Lofton-Day, M. Moore, A. Littau, A. Grossman, H. Haugen, K. Foley, H. Blumberg, K. Harrison, W. Kindsvogel and C. H. Clegg (2000). TACI and BCMA are receptors for a TNF homologue implicated in B-cell autoimmune disease. Nature **404**(6781): 995-999.

Guerrette S., S. Acharya and R. Fishel (1999). The interaction of the human MutL homologues in hereditary nonpolyposis colon cancer. J Biol Chem **274**(10): 6336-6341.

Guikema J. E., E. K. Linehan, D. Tsuchimoto, Y. Nakabeppu, P. R. Strauss, J. Stavnezer and C. E. Schrader (2007). APE1- and APE2-dependent DNA breaks in immunoglobulin class switch recombination. J Exp Med **204**(12): 3017-3026.

Hahne M., T. Kataoka, M. Schroter, K. Hofmann, M. Irmler, J. L. Bodmer, P. Schneider, T. Bornand, N. Holler, L. E. French, B. Sordat, D. Rimoldi and J. Tschopp (1998). APRIL, a new ligand of the tumor necrosis factor family, stimulates tumor cell growth. J Exp Med **188**(6): 1185-1190.

Hammarström L. and C. I. Smith (1983). HLA-A, B, C and DR antigens in immunoglobulin A deficiency. Tissue Antigens **21**(1): 75-79.

Hammarström L., I. Vorechovsky and D. Webster (2000). Selective IgA deficiency (SIgAD) and common variable immunodeficiency (CVID). Clin Exp Immunol **120**(2): 225-231.

Hannelius U., C. M. Lindgren, E. Melen, A. Malmberg, U. von Dobeln and J. Kere (2005). Phenylketonuria screening registry as a resource for population genetic studies. J Med Genet **42**(10): e60.

Hannum L. G., D. Ni, A. M. Haberman, M. G. Weigert and M. J. Shlomchik (1996). A disease-related rheumatoid factor autoantibody is not tolerized in a normal mouse: implications for the origins of autoantibodies in autoimmune disease. J Exp Med **184**(4): 1269-1278.

Harris R. S. and M. T. Liddament (2004). Retroviral restriction by APOBEC proteins. Nat Rev Immunol **4**(11): 868-877.

Harris R. S., J. E. Sale, S. K. Petersen-Mahrt and M. S. Neuberger (2002). AID is essential for immunoglobulin V gene conversion in a cultured B cell line. Curr Biol **12**(5): 435-438.

Honjo T., M. Muramatsu and S. Fagarasan (2004). AID: how does it aid antibody diversity? Immunity **20**(6): 659-668.

Hughes A. L. and M. Yeager (1998). Natural selection at major histocompatibility complex loci of vertebrates. Annu Rev Genet **32**: 415-435.

Hymowitz S. G., D. R. Patel, H. J. Wallweber, S. Runyon, M. Yan, J. Yin, S. K. Shriver, N. C. Gordon, B. Pan, N. J. Skelton, R. F. Kelley and M. A. Starovasnik (2005). Structures of APRIL-receptor complexes: like BCMA, TACI employs only a single cysteine-rich domain for high affinity ligand binding. J Biol Chem **280**(8): 7218-7227.

Imai K., G. Slupphaug, W. I. Lee, P. Revy, S. Nonoyama, N. Catalan, L. Yel, M. Forveille, B. Kavli, H. E. Krokan, H. D. Ochs, A. Fischer and A. Durandy (2003). Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. Nat Immunol **4**(10): 1023-1028.

Imai K., Y. Zhu, P. Revy, T. Morio, S. Mizutani, A. Fischer, S. Nonoyama and A. Durandy (2005). Analysis of class switch recombination and somatic hypermutation in patients affected with autosomal dominant hyper-IgM syndrome type 2. Clin Immunol **115**(3): 277-285.

Ingold K., A. Zumsteg, A. Tardivel, B. Huard, Q. G. Steiner, T. G. Cachero, F. Qiang, L. Gorelik, S. L. Kalled, H. Acha-Orbea, P. D. Rennert, J. Tschopp and P. Schneider (2005). Identification of proteoglycans as the APRIL-specific binding partners. J Exp Med **201**(9): 1375-1383.

Ishizaka A., M. Nakanishi, S. Yamada, Y. Sakiyama and S. Matsumoto (1989). Development of hypogammaglobulinaemia in a patient with common variable immunodeficiency. Eur J Pediatr **149**(3): 175-176.

Jain A., C. A. Ma, S. Liu, M. Brown, J. Cohen and W. Strober (2001). Specific missense mutations in NEMO result in hyper-IgM syndrome with hypohydrotic ectodermal dysplasia. Nat Immunol **2**(3): 223-228.

Jarmuz A., A. Chester, J. Bayliss, J. Gisbourne, I. Dunham, J. Scott and N. Navaratnam (2002). An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. Genomics **79**(3): 285-296.

Johnson M. L., L. G. Keeton, Z. B. Zhu, J. E. Volanakis, M. D. Cooper and H. W. Schroeder, Jr. (1997). Age-related changes in serum immunoglobulins in patients with familial IgA deficiency and common variable immunodeficiency (CVID). Clin Exp Immunol **108**(3): 477-483.

Kanegane H., K. Agematsu, T. Futatani, M. M. Sira, K. Suga, T. Sekiguchi, M. C. van Zelm and T. Miyawaki (2007). Novel mutations in a Japanese patient with CD19 deficiency. Genes Immun **8**(8): 663-670.

Kasahara Y., H. Kaneko, T. Fukao, T. Terada, T. Asano, K. Kasahara and N. Kondo (2003). Hyper-IgM syndrome with putative dominant negative mutation in activation-induced cytidine deaminase. J Allergy Clin Immunol **112**(4): 755-760.

Khare S. D., I. Sarosi, X. Z. Xia, S. McCabe, K. Miner, I. Solovyev, N. Hawkins, M. Kelley, D. Chang, G. Van, L. Ross, J. Delaney, L. Wang, D. Lacey, W. J. Boyle and H. Hsu (2000). Severe B cell hyperplasia and autoimmune disease in TALL-1 transgenic mice. Proc Natl Acad Sci U S A **97**(7): 3370-3375.

Kim N., G. Bozek, J. C. Lo and U. Storb (1999). Different mismatch repair deficiencies all have the same effects on somatic hypermutation: intact primary mechanism accompanied by secondary modifications. J Exp Med **190**(1): 21-30.

Klein U., Y. Tu, G. A. Stolovitzky, J. L. Keller, J. Haddad, Jr., V. Miljkovic, G. Cattoretti, A. Califano and R. Dalla-Favera (2003). Transcriptional analysis of the B cell germinal center reaction. Proc Natl Acad Sci U S A **100**(5): 2639-2644.

Kneitz B., P. E. Cohen, E. Avdievich, L. Zhu, M. F. Kane, H. Hou, Jr., R. D. Kolodner, R. Kucherlapati, J. W. Pollard and W. Edelmann (2000). MutS homolog 4 localization to meiotic chromosomes is required for chromosome pairing during meiosis in male and female mice. Genes Dev **14**(9): 1085-1097.

Kobayashi J., A. Antoccia, H. Tauchi, S. Matsuura and K. Komatsu (2004). NBS1 and its functional role in the DNA damage response. DNA Repair (Amst) **3**(8-9): 855-861.

Kohli M., C. Rago, C. Lengauer, K. W. Kinzler and B. Vogelstein (2004). Facile methods for generating human somatic cell gene knockouts using recombinant adeno-associated viruses. Nucleic Acids Res **32**(1): e3.

Koistinen J. (1975). Selective IgA deficiency in blood donors. Vox Sang **29**(3): 192-202.

Kokron C. M., P. R. Errante, M. T. Barros, G. V. Baracho, M. M. Camargo, J. Kalil and L. V. Rizzo (2004). Clinical and laboratory aspects of common variable immunodeficiency. An Acad Bras Cienc **76**(4): 707-726.

Kondo E., A. Horii and S. Fukushige (2001). The interacting domains of three MutL heterodimers in man: hMLH1 interacts with 36 homologous amino acid residues within hMLH3, hPMS1 and hPMS2. Nucleic Acids Res **29**(8): 1695-1702.

Kong Q. and N. Maizels (1999). PMS2-deficiency diminishes hypermutation of a lambda1 transgene in young but not older mice. Mol Immunol **36**(2): 83-91.

Kracker S., Y. Bergmann, I. Demuth, P. O. Frappart, G. Hildebrand, R. Christine, Z. Q. Wang, K. Sperling, M. Digweed and A. Radbruch (2005). Nibrin functions in Ig class-switch recombination. Proc Natl Acad Sci U S A **102**(5): 1584-1589.

Krokan H. E., R. Standal and G. Slupphaug (1997). DNA glycosylases in the base excision repair of DNA. Biochem J **325 (Pt 1)**: 1-16.

Kunz C., Y. Saito and P. Schar (2009). DNA Repair in mammalian cells: Mismatched repair: variations on a theme. Cell Mol Life Sci **66**(6): 1021-1038.

Kutukculer N., D. Moratto, Y. Aydinok, V. Lougaris, S. Aksoylar, A. Plebani, F. Genel and L. D. Notarangelo (2003). Disseminated cryptosporidium infection in an infant with hyper-IgM syndrome caused by CD40 deficiency. J Pediatr **142**(2): 194-196.

Lahdesmaki A., A. M. Taylor, K. H. Chrzanowska and Q. Pan-Hammarström (2004). Delineation of the role of the Mre11 complex in class switch recombination. J Biol Chem **279**(16): 16479-16487.

Lane P., A. Traunecker, S. Hubele, S. Inui, A. Lanzavecchia and D. Gray (1992). Activated human T cells express a ligand for the human B cell-associated antigen CD40 which participates in T cell-dependent activation of B lymphocytes. Eur J Immunol **22**(10): 2573-2578.

Larson E. D., W. J. Cummings, D. W. Bednarski and N. Maizels (2005). MRE11/RAD50 cleaves DNA in the AID/UNG-dependent pathway of immunoglobulin gene diversification. Mol Cell **20**(3): 367-375.

Lesley R., Y. Xu, S. L. Kalled, D. M. Hess, S. R. Schwab, H. B. Shu and J. G. Cyster (2004). Reduced competitiveness of autoantigen-engaged B cells due to increased dependence on BAFF. Immunity **20**(4): 441-453.

Li Z., S. J. Scherer, D. Ronai, M. D. Iglesias-Ussel, J. U. Peled, P. D. Bardwell, M. Zhuang, K. Lee, A. Martin, W. Edelmann and M. D. Scharff (2004). Examination of Msh6- and Msh3-deficient mice in class switching reveals overlapping and distinct roles of MutS homologues in antibody diversification. J Exp Med **200**(1): 47-59.

Lieber M. R. (2008). The mechanism of human nonhomologous DNA end joining. J Biol Chem **283**(1): 1-5.

Litinskiy M. B., B. Nardelli, D. M. Hilbert, B. He, A. Schaffer, P. Casali and A. Cerutti (2002). DCs induce CD40-independent immunoglobulin class switching through BLyS and APRIL. Nat Immunol **3**(9): 822-829.

Liu T., P. Tannergard, P. Hackman, C. Rubio, U. Kressner, G. Lindmark, D. Hellgren, B. Lambert and A. Lindblom (1999). Missense mutations in hMLH1 associated with colorectal cancer. Hum Genet **105**(5): 437-441.

Longerich S., U. Basu, F. Alt and U. Storb (2006). AID in somatic hypermutation and class switch recombination. Curr Opin Immunol **18**(2): 164-174.

Lynch H. T. and A. de la Chapelle (2003). Hereditary colorectal cancer. N Engl J Med **348**(10): 919-932.

Mackay F., P. Schneider, P. Rennert and J. Browning (2003). BAFF AND APRIL: a tutorial on B cell survival. Annu Rev Immunol **21**: 231-264.

Mackay F., S. A. Woodcock, P. Lawton, C. Ambrose, M. Baetscher, P. Schneider, J. Tschopp and J. L. Browning (1999). Mice transgenic for BAFF develop lymphocytic disorders along with autoimmune manifestations. J Exp Med **190**(11): 1697-1710.

Maizels N. (2005). Immunoglobulin gene diversification. Annu Rev Genet **39**: 23-46.

Maki R., A. Traunecker, H. Sakano, W. Roeder and S. Tonegawa (1980). Exon shuffling generates an immunoglobulin heavy chain gene. Proc Natl Acad Sci U S A **77**(4): 2138-2142.

Mantchev G. T., C. S. Cortesao, M. Rebrovich, M. Cascalho and R. J. Bram (2007). TACI is required for efficient plasma cell differentiation in response to T-independent type 2 antigens. J Immunol **179**(4): 2282-2288.

Marsters S. A., M. Yan, R. M. Pitti, P. E. Haas, V. M. Dixit and A. Ashkenazi (2000). Interaction of the TNF homologues BLyS and APRIL with the TNF receptor homologues BCMA and TACI. Curr Biol **10**(13): 785-788.

Martin A., Z. Li, D. P. Lin, P. D. Bardwell, M. D. Iglesias-Ussel, W. Edelmann and M. D. Scharff (2003). Msh2 ATPase activity is essential for somatic hypermutation at a-T basepairs and for efficient class switch recombination. J Exp Med **198**(8): 1171-1178.

Martinez-Pomar N., I. Munoz-Saa, D. Heine-Suner, A. Martin, A. Smahi and N. Matamoros (2005). A new mutation in exon 7 of NEMO gene: late skewed X-chromosome inactivation in an incontinentia pigmenti female patient with immunodeficiency. Hum Genet **118**(3-4): 458-465.

Martomo S. A., W. W. Yang and P. J. Gearhart (2004). A role for Msh6 but not Msh3 in somatic hypermutation and class switch recombination. J Exp Med **200**(1): 61-68.

Mathis D. J., C. Benoist, V. E. Williams, 2nd, M. Kanter and H. O. McDevitt (1983). Several mechanisms can account for defective E alpha gene expression in different mouse haplotypes. Proc Natl Acad Sci U S A **80**(1): 273-277.

McNeill D. R. and D. M. Wilson, 3rd (2007). A dominant-negative form of the major human abasic endonuclease enhances cellular sensitivity to laboratory and clinical DNA-damaging agents. Mol Cancer Res **5**(1): 61-70.

Mellemkjaer L., L. Hammarström, V. Andersen, J. Yuen, C. Heilmann, T. Barington, J. Bjorkander and J. H. Olsen (2002). Cancer risk among patients with IgA deficiency or common variable immunodeficiency and their relatives: a combined Danish and Swedish study. Clin Exp Immunol **130**(3): 495-500.

Mensink E. J., A. Thompson, L. A. Sandkuyl, M. E. Kraakman, J. D. Schot, T. Espanol and R. K. Schuurman (1987). X-linked immunodeficiency with hyperimmunoglobulinemia M appears to be linked to the DXS42 restriction fragment length polymorphism locus. Hum Genet **76**(1): 96-99.

Miller J. P., J. E. Stadanlick and M. P. Cancro (2006). Space, selection, and surveillance: setting boundaries with BLyS. J Immunol **176**(11): 6405-6410.

Min I. M. and E. Selsing (2005). Antibody class switch recombination: roles for switch sequences and mismatch repair proteins. Adv Immunol **87**: 297-328.

Minegishi Y., A. Lavoie, C. Cunningham-Rundles, P. M. Bedard, J. Hebert, L. Cote, K. Dan, D. Sedlak, R. H. Buckley, A. Fischer, A. Durandy and M. E. Conley (2000). Mutations in activation-induced cytidine deaminase in patients with hyper IgM syndrome. Clin Immunol **97**(3): 203-210.

Mitchell M. K., P. K. Gregersen, S. Johnson, R. Parsons and D. Vlahov (2004). The New York Cancer Project: rationale, organization, design, and baseline characteristics. J Urban Health **81**(2): 301-310.

Morales M., J. W. Theunissen, C. F. Kim, R. Kitagawa, M. B. Kastan and J. H. Petrini (2005). The Rad50S allele promotes ATM-dependent DNA damage responses and suppresses ATM deficiency: implications for the Mre11 complex as a DNA damage sensor. Genes Dev **19**(24): 3043-3054.

Moshous D., I. Callebaut, R. de Chasseval, B. Corneo, M. Cavazzana-Calvo, F. Le Deist, I. Tezcan, O. Sanal, Y. Bertrand, N. Philippe, A. Fischer and J. P. de Villartay (2001). Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. Cell **105**(2): 177-186.

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

Muto T., M. Muramatsu, M. Taniwaki, K. Kinoshita and T. Honjo (2000). Isolation, tissue distribution, and chromosomal localization of the human activation-induced cytidine deaminase (AID) gene. Genomics **68**(1): 85-88.

Nakahara M., E. Sonoda, K. Nojima, J. E. Sale, K. Takenaka, K. Kikuchi, Y. Taniguchi, K. Nakamura, Y. Sumitomo, R. T. Bree, N. F. Lowndes and S. Takeda (2009). Genetic evidence for single-strand lesions initiating Nbs1-dependent homologous recombination in diversification of Ig v in chicken B lymphocytes. PLoS Genet **5**(1): e1000356.

Neuberger M. S. (2008). Antibody diversification by somatic mutation: from Burnet onwards. Immunol Cell Biol **86**(2): 124-132.
Nilsen H., M. Otterlei, T. Haug, K. Solum, T. A. Nagelhus, F. Skorpen and H. E. Krokan (1997). Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the UNG gene. Nucleic Acids Res **25**(4): 750-755.
Noelle R. J., M. Roy, D. M. Shepherd, I. Stamenkovic, J. A. Ledbetter and A. Aruffo (1992). A 39-kDa protein on activated helper T cells binds CD40 and transduces the signal for cognate activation of B cells. Proc Natl Acad Sci U S A **89**(14): 6550-6554.
O'Connor B. P., V. S. Raman, L. D. Erickson, W. J. Cook, L. K. Weaver, C. Ahonen, L. L. Lin, G. T. Mantchev, R. J. Bram and R. J. Noelle (2004). BCMA is essential for the survival of long-lived bone marrow plasma cells. J Exp Med **199**(1): 91-98.
Obmolova G., C. Ban, P. Hsieh and W. Yang (2000). Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. Nature **407**(6805): 703-710.
Oettinger M. A., D. G. Schatz, C. Gorka and D. Baltimore (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. Science **248**(4962): 1517-1523.
Offer S. M., Q. Pan-Hammarström, L. Hammarström and R. S. Harris Unique DNA Repair Gene Variations Associate with the Primary Antibody Deficiency Syndromes IgAD and CVID. Hum Mutat **in review**.
Olerup O., C. I. Smith and L. Hammarström (1990). Different amino acids at position 57 of the HLA-DQ beta chain associated with susceptibility and resistance to IgA deficiency. Nature **347**(6290): 289-290.
Orange J. S., O. Levy, S. R. Brodeur, K. Krzewski, R. M. Roy, J. E. Niemela, T. A. Fleisher, F. A. Bonilla and R. S. Geha (2004). Human nuclear factor kappa B essential modulator mutation can result in immunodeficiency without ectodermal dysplasia. J Allergy Clin Immunol **114**(3): 650-656.
Orstavik K. H., M. Kristiansen, G. P. Knudsen, K. Storhaug, A. Vege, K. Eiklid, T. G. Abrahamsen, A. Smahi and J. Steen-Johnsen (2006). Novel splicing mutation in the NEMO (IKK-gamma) gene with severe immunodeficiency and heterogeneity of X-chromosome inactivation. Am J Med Genet A **140**(1): 31-39.
Ostergaard J. R., L. Sunde and H. Okkels (2005). Neurofibromatosis von Recklinghausen type I phenotype and early onset of cancers in siblings compound heterozygous for mutations in MSH6. Am J Med Genet A **139A**(2): 96-105; discussion 196.
Otterlei M., T. Haug, T. A. Nagelhus, G. Slupphaug, T. Lindmo and H. E. Krokan (1998). Nuclear and mitochondrial splice forms of human uracil-DNA glycosylase contain a complex nuclear localisation signal and a strong classical mitochondrial localisation signal, respectively. Nucleic Acids Res **26**(20): 4611-4617.
Padayachee M., C. Feighery, A. Finn, C. McKeown, R. J. Levinsky, C. Kinnon and S. Malcolm (1992). Mapping of the X-linked form of hyper-IgM syndrome (HIGM1) to Xq26 by close linkage to HPRT. Genomics **14**(2): 551-553.
Padayachee M., R. J. Levinsky, C. Kinnon, A. Finn, C. McKeown, C. Feighery, L. D. Notarangelo, R. W. Hendriks, A. P. Read and S. Malcolm (1993). Mapping of the X linked form of hyper IgM syndrome (HIGM1). J Med Genet **30**(3): 202-205.
Palanduz S., A. Palanduz, I. Yalcin, A. Somer, U. Ones, D. Ustek, S. Ozturk, N. Salman, N. Guler and H. Bilge (1998). In vitro chromosomal radiosensitivity in common variable immune deficiency. Clin Immunol Immunopathol **86**(2): 180-182.
Pan Q., C. Petit-Frere, A. Lahdesmaki, H. Gregorek, K. H. Chrzanowska and L. Hammarström (2002). Alternative end joining during switch recombination in patients with ataxia-telangiectasia. Eur J Immunol **32**(5): 1300-1308.

Pan-Hammarström Q., S. Dai, Y. Zhao, I. F. van Dijk-Hard, R. A. Gatti, A. L. Borresen-Dale and L. Hammarström (2003). ATM is not required in somatic hypermutation of VH, but is involved in the introduction of mutations in the switch mu region. J Immunol **170**(7): 3707-3716.

Pan-Hammarström Q. and L. Hammarström (2008). Antibody deficiency diseases. Eur J Immunol **38**(2): 327-333.

Pan-Hammarström Q., A. M. Jones, A. Lahdesmaki, W. Zhou, R. A. Gatti, L. Hammarström, A. R. Gennery and M. R. Ehrenstein (2005). Impact of DNA ligase IV on nonhomologous end joining pathways during class switch recombination in human cells. J Exp Med **201**(2): 189-194.

Pan-Hammarström Q., U. Salzer, L. Du, J. Bjorkander, C. Cunningham-Rundles, D. L. Nelson, C. Bacchelli, H. B. Gaspar, S. Offer, T. W. Behrens, B. Grimbacher and L. Hammarström (2007). Reexamining the role of TACI coding variants in common variable immunodeficiency and selective IgA deficiency. Nat Genet **39**(4): 429-430.

Pape K. A., V. Kouskoff, D. Nemazee, H. L. Tang, J. G. Cyster, L. E. Tze, K. L. Hippen, T. W. Behrens and M. K. Jenkins (2003). Visualization of the genesis and fate of isotype-switched B cells during a primary immune response. J Exp Med **197**(12): 1677-1687.

Peron S., A. Metin, P. Gardes, M. A. Alyanakian, E. Sheridan, C. P. Kratz, A. Fischer and A. Durandy (2008). Human PMS2 deficiency is associated with impaired immunoglobulin class switch recombination. J Exp Med **205**(11): 2465-2472.

Petersen S., R. Casellas, B. Reina-San-Martin, H. T. Chen, M. J. Difilippantonio, P. C. Wilson, L. Hanitsch, A. Celeste, M. Muramatsu, D. R. Pilch, C. Redon, T. Ried, W. M. Bonner, T. Honjo, M. C. Nussenzweig and A. Nussenzweig (2001). AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. Nature **414**(6864): 660-665.

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

Phung Q. H., D. B. Winter, R. Alrefai and P. J. Gearhart (1999). Hypermutation in Ig V genes from mice deficient in the MLH1 mismatch repair protein. J Immunol **162**(6): 3121-3124.

Phung Q. H., D. B. Winter, A. Cranston, R. E. Tarone, V. A. Bohr, R. Fishel and P. J. Gearhart (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.

Pilia G., G. Porta, M. Padayachee, S. Malcolm, I. Zucchi, A. Villa, P. Macchi, P. Vezzoni and D. Schlessinger (1994). Human CD40L gene maps between DXS144E and DXS300 in Xq26. Genomics **22**(1): 249-251.

Puel A., J. Reichenbach, J. Bustamante, C. L. Ku, J. Feinberg, R. Doffinger, M. Bonnet, O. Filipe-Santos, L. de Beaucoudrey, A. Durandy, G. Horneff, F. Novelli, V. Wahn, A. Smahi, A. Israel, T. Niehues and J. L. Casanova (2006). The NEMO mutation creating the most-upstream premature stop codon is hypomorphic because of a reinitiation of translation. Am J Hum Genet **78**(4): 691-701.

Quartier P., J. Bustamante, O. Sanal, A. Plebani, M. Debre, A. Deville, J. Litzman, J. Levy, J. P. Fermand, P. Lane, G. Horneff, G. Aksu, I. Yalcin, G. Davies, I. Tezcan, F. Ersoy, N. Catalan, K. Imai, A. Fischer and A. Durandy (2004). Clinical, immunologic and genetic analysis of 29 patients with autosomal recessive hyper-IgM syndrome due to Activation-Induced Cytidine Deaminase deficiency. Clin Immunol **110**(1): 22-29.

Rada C., J. M. Di Noia and M. S. Neuberger (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.

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

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

Raevaara T. E., M. K. Korhonen, H. Lohi, H. Hampel, E. Lynch, K. E. Lonnqvist, E. Holinski-Feder, C. Sutter, W. McKinnon, S. Duraisamy, A. M. Gerdes, P. Peltomaki, M. Kohonen-Ccorish, E. Mangold, F. Macrae, M. Greenblatt, A. de la Chapelle and M. Nystrom (2005). Functional significance and clinical phenotype of nontruncating mismatch repair variants of MLH1. Gastroenterology **129**(2): 537-549.

Ramiro A. R., M. Jankovic, T. Eisenreich, S. Difilippantonio, S. Chen-Kiang, M. Muramatsu, T. Honjo, A. Nussenzweig and M. C. Nussenzweig (2004). AID is required for c-myc/IgH chromosome translocations in vivo. Cell **118**(4): 431-438.

Reina-San-Martin B., M. C. Nussenzweig, A. Nussenzweig and S. Difilippantonio (2005). Genomic instability, endoreduplication, and diminished Ig class-switch recombination in B cells lacking Nbs1. Proc Natl Acad Sci U S A **102**(5): 1590-1595.

Revy P., T. Muto, Y. Levy, F. Geissmann, A. Plebani, O. Sanal, N. Catalan, M. Forveille, R. Dufourcq-Labelouse, A. Gennery, I. Tezcan, F. Ersoy, H. Kayserili, A. G. Ugazio, N. Brousse, M. Muramatsu, L. D. Notarangelo, K. Kinoshita, T. Honjo, A. Fischer and A. Durandy (2000). Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). Cell **102**(5): 565-575.

Reynaud C. A., V. Anquez, A. Dahan and J. C. Weill (1985). A single rearrangement event generates most of the chicken immunoglobulin light chain diversity. Cell **40**(2): 283-291.

Reynaud C. A., V. Anquez, H. Grimal and J. C. Weill (1987). A hyperconversion mechanism generates the chicken light chain preimmune repertoire. Cell **48**(3): 379-388.

Reynaud C. A., B. Bertocci, A. Dahan and J. C. Weill (1994). Formation of the chicken B-cell repertoire: ontogenesis, regulation of Ig gene rearrangement, and diversification by gene conversion. Adv Immunol **57**: 353-378.

Ruis B. L., K. R. Fattah and E. A. Hendrickson (2008). The catalytic subunit of DNA-dependent protein kinase regulates proliferation, telomere length, and genomic stability in human somatic cells. Mol Cell Biol **28**(20): 6182-6195.

Sabouri Z., I. M. Okazaki, R. Shinkura, N. Begum, H. Nagaoka, D. Tsuchimoto, Y. Nakabeppu and T. Honjo (2009). Apex2 is required for efficient somatic hypermutation but not for class switch recombination of immunoglobulin genes. Int Immunol **21**(8): 947-955.

Sale J. E. (2004). Immunoglobulin diversification in DT40: a model for vertebrate DNA damage tolerance. DNA Repair (Amst) **3**(7): 693-702.

Salzer U., C. Bacchelli, S. Buckridge, Q. Pan-Hammarström, S. Jennings, V. Lougaris, A. Bergbreiter, T. Hagena, J. Birmelin, A. Plebani, A. D. Webster, H. H. Peter, D. Suez, H. Chapel, A. McLean-Tooke, G. P. Spickett, S. Anover-Sombke, H. D. Ochs, S. Urschel, B. H. Belohradsky, S. Ugrinovic, D. S. Kumararatne, T. C. Lawrence, A. M. Holm, J. L. Franco, I. Schulze, P. Schneider, E. M. Gertz, A. A. Schaffer, L. Hammarström, A. J. Thrasher, H. B. Gaspar and B. Grimbacher (2009). Relevance of biallelic versus monoallelic TNFRSF13B mutations in distinguishing disease-causing from risk-increasing TNFRSF13B variants in antibody deficiency syndromes. Blood **113**(9): 1967-1976.

Salzer U., H. M. Chapel, A. D. Webster, Q. Pan-Hammarström, A. Schmitt-Graeff, M. Schlesier, H. H. Peter, J. K. Rockstroh, P. Schneider, A. A. Schaffer, L. Hammarström and B. Grimbacher (2005). Mutations in TNFRSF13B encoding TACI are associated with common variable immunodeficiency in humans. Nat Genet **37**(8): 820-828.

Salzer U. and B. Grimbacher (2006). Common variable immunodeficiency: The power of co-stimulation. Semin Immunol **18**(6): 337-346.

Salzer U., A. Maul-Pavicic, C. Cunningham-Rundles, S. Urschel, B. H. Belohradsky, J. Litzman, A. Holm, J. L. Franco, A. Plebani, L. Hammarström, A. Skrabl, W. Schwinger and B. Grimbacher (2004). ICOS deficiency in patients with common variable immunodeficiency. Clin Immunol **113**(3): 234-240.

Saribasak H., N. N. Saribasak, F. M. Ipek, J. W. Ellwart, H. Arakawa and J. M. Buerstedde (2006). Uracil DNA glycosylase disruption blocks Ig gene conversion and induces transition mutations. J Immunol **176**(1): 365-371.

Sarvas H. and O. Makela (1970). Haptenated bacteriophage in the assay of antibody quantity and affinity: maturation of an immune response. Immunochemistry **7**(11): 933-943.

Savitsky K., A. Bar-Shira, S. Gilad, G. Rotman, Y. Ziv, L. Vanagaite, D. A. Tagle, S. Smith, T. Uziel, S. Sfez, M. Ashkenazi, I. Pecker, M. Frydman, R. Harnik, S. R. Patanjali, A. Simmons, G. A. Clines, A. Sartiel, R. A. Gatti, L. Chessa, O. Sanal, M. F. Lavin, N. G. Jaspers, A. M. Taylor, C. F. Arlett, T. Miki, S. M. Weissman, M. Lovett, F. S. Collins and Y. Shiloh (1995). A single ataxia telangiectasia gene with a product similar to PI-3 kinase. Science **268**(5218): 1749-1753.

Schaffer A. A., U. Salzer, L. Hammarström and B. Grimbacher (2007). Deconstructing common variable immunodeficiency by genetic analysis. Curr Opin Genet Dev **17**(3): 201-212.

Schaffer F. M., J. Palermos, Z. B. Zhu, B. O. Barger, M. D. Cooper and J. E. Volanakis (1989). Individuals with IgA deficiency and common variable immunodeficiency share polymorphisms of major histocompatibility complex class III genes. Proc Natl Acad Sci U S A **86**(20): 8015-8019.

Schiemann B., J. L. Gommerman, K. Vora, T. G. Cachero, S. Shulga-Morskaya, M. Dobles, E. Frew and M. L. Scott (2001). An essential role for BAFF in the normal development of B cells through a BCMA-independent pathway. Science **293**(5537): 2111-2114.

Schneider P., H. Takatsuka, A. Wilson, F. Mackay, A. Tardivel, S. Lens, T. G. Cachero, D. Finke, F. Beermann and J. Tschopp (2001). Maturation of marginal zone and follicular B cells requires B cell activating factor of the tumor necrosis factor family and is independent of B cell maturation antigen. J Exp Med **194**(11): 1691-1697.

Schrader C. E., W. Edelmann, R. Kucherlapati and J. Stavnezer (1999). Reduced isotype switching in splenic B cells from mice deficient in mismatch repair enzymes. J Exp Med **190**(3): 323-330.

Schrader C. E., E. K. Linehan, S. N. Mochegova, R. T. Woodland and J. Stavnezer (2005). Inducible DNA breaks in Ig S regions are dependent on AID and UNG. J Exp Med **202**(4): 561-568.

Schrader C. E., J. Vardo and J. Stavnezer (2002). Role for mismatch repair proteins Msh2, Mlh1, and Pms2 in immunoglobulin class switching shown by sequence analysis of recombination junctions. J Exp Med **195**(3): 367-373.

Scott R. H., S. Mansour, K. Pritchard-Jones, D. Kumar, F. MacSweeney and N. Rahman (2007). Medulloblastoma, acute myelocytic leukemia and colonic carcinomas in a child with biallelic MSH6 mutations. Nat Clin Pract Oncol **4**(2): 130-134.

Sekine H., R. C. Ferreira, Q. Pan-Hammarström, R. R. Graham, B. Ziemba, S. S. de Vries, J. Liu, K. Hippen, T. Koeuth, W. Ortmann, A. Iwahori, M. K. Elliott, S. Offer, C. Skon, L. Du, J. Novitzke, A. T. Lee, N. Zhao, J. D. Tompkins, D. Altshuler, P. K. Gregersen, C. Cunningham-Rundles, R. S. Harris, C. Her, D. L. Nelson, L. Hammarström, G. S. Gilkeson and T. W. Behrens (2007). Role for Msh5 in the regulation of Ig class switch recombination. Proc Natl Acad Sci U S A **104**(17): 7193-7198.

Sekine H., K. L. Graham, S. Zhao, M. K. Elliott, P. Ruiz, P. J. Utz and G. S. Gilkeson (2006). Role of MHC-linked genes in autoantigen selection and renal disease in a murine model of systemic lupus erythematosus. J Immunol **177**(10): 7423-7434.

Seshasayee D., P. Valdez, M. Yan, V. M. Dixit, D. Tumas and I. S. Grewal (2003). Loss of TACI causes fatal lymphoproliferation and autoimmunity, establishing TACI as an inhibitory BLyS receptor. Immunity **18**(2): 279-288.

Shen H. M., A. Tanaka, G. Bozek, D. Nicolae and U. Storb (2006). Somatic hypermutation and class switch recombination in Msh6(-/-)Ung(-/-) double-knockout mice. J Immunol **177**(8): 5386-5392.

Shulga-Morskaya S., M. Dobles, M. E. Walsh, L. G. Ng, F. MacKay, S. P. Rao, S. L. Kalled and M. L. Scott (2004). B cell-activating factor belonging to the TNF family acts through separate receptors to support B cell survival and T cell-independent antibody formation. J Immunol **173**(4): 2331-2341.

Smith E. L. and R. H. Shmerling (1999). The American College of Rheumatology criteria for the classification of systemic lupus erythematosus: strengths, weaknesses, and opportunities for improvement. Lupus **8**(8): 586-595.

Snowden T., S. Acharya, C. Butz, M. Berardini and R. Fishel (2004). hMSH4-hMSH5 recognizes Holliday Junctions and forms a meiosis-specific sliding clamp that embraces homologous chromosomes. Mol Cell **15**(3): 437-451.

Sonoda E., C. Morrison, Y. M. Yamashita, M. Takata and S. Takeda (2001). Reverse genetic studies of homologous DNA recombination using the chicken B-lymphocyte line, DT40. Philos Trans R Soc Lond B Biol Sci **356**(1405): 111-117.

Stavnezer J., J. E. Guikema and C. E. Schrader (2008). Mechanism and regulation of class switch recombination. Annu Rev Immunol **26**: 261-292.

Stewart G. S., R. S. Maser, T. Stankovic, D. A. Bressan, M. I. Kaplan, N. G. Jaspers, A. Raams, P. J. Byrd, J. H. Petrini and A. M. Taylor (1999). The DNA double-strand break repair gene hMRE11 is mutated in individuals with an ataxia-telangiectasia-like disorder. Cell **99**(6): 577-587.

Stohl W. (2005). BlySfulness does not equal blissfulness in systemic lupus erythematosus: a therapeutic role for BLyS antagonists. Curr Dir Autoimmun **8**: 289-304.

Svetlanov A. and P. E. Cohen (2004). Mismatch repair proteins, meiosis, and mice: understanding the complexities of mammalian meiosis. Exp Cell Res **296**(1): 71-79.

Ta V. T., H. Nagaoka, N. Catalan, A. Durandy, A. Fischer, K. Imai, S. Nonoyama, J. Tashiro, M. Ikegawa, S. Ito, K. Kinoshita, M. Muramatsu and T. Honjo (2003). AID mutant analyses indicate requirement for class-switch-specific cofactors. Nat Immunol **4**(9): 843-848.

Takeda S., E. L. Masteller, C. B. Thompson and J. M. Buerstedde (1992). RAG-2 expression is not essential for chicken immunoglobulin gene conversion. Proc Natl Acad Sci U S A **89**(9): 4023-4027.

Taube C., J. M. Thurman, K. Takeda, A. Joetham, N. Miyahara, M. C. Carroll, A. Dakhama, P. C. Giclas, V. M. Holers and E. W. Gelfand (2006). Factor B of the alternative complement pathway regulates development of airway hyperresponsiveness and inflammation. Proc Natl Acad Sci U S A **103**(21): 8084-8089.

Theofilopoulos A. N. and F. J. Dixon (1985). Murine models of systemic lupus erythematosus. Adv Immunol **37**: 269-390.

Thompson C. B. and P. E. Neiman (1987). Somatic diversification of the chicken immunoglobulin light chain gene is limited to the rearranged variable gene segment. Cell **48**(3): 369-378.

Thompson J. S., S. A. Bixler, F. Qian, K. Vora, M. L. Scott, T. G. Cachero, C. Hession, P. Schneider, I. D. Sizing, C. Mullen, K. Strauch, M. Zafari, C. D. Benjamin, J. Tschopp, J. L. Browning and C. Ambrose (2001). BAFF-R, a newly identified TNF receptor that specifically interacts with BAFF. Science **293**(5537): 2108-2111.

Thompson J. S., P. Schneider, S. L. Kalled, L. Wang, E. A. Lefevre, T. G. Cachero, F. MacKay, S. A. Bixler, M. Zafari, Z. Y. Liu, S. A. Woodcock, F. Qian, M. Batten, C. Madry, Y. Richard, C. D. Benjamin, J. L. Browning, A. Tsapis, J. Tschopp and C. Ambrose (2000). BAFF binds to the tumor necrosis factor receptor-like molecule B cell maturation antigen and is important for maintaining the peripheral B cell population. J Exp Med **192**(1): 129-135.

Tonegawa S. (1983). Somatic generation of antibody diversity. Nature **302**(5909): 575-581.

Valle A., C. E. Zuber, T. Defrance, O. Djossou, M. De Rie and J. Banchereau (1989). Activation of human B lymphocytes through CD40 and interleukin 4. Eur J Immunol **19**(8): 1463-1467.

van Zelm M. C., I. Reisli, M. van der Burg, D. Castano, C. J. van Noesel, M. J. van Tol, C. Woellner, B. Grimbacher, P. J. Patino, J. J. van Dongen and J. L. Franco (2006). An antibody-deficiency syndrome due to mutations in the CD19 gene. N Engl J Med **354**(18): 1901-1912.

Varon R., C. Vissinga, M. Platzer, K. M. Cerosaletti, K. H. Chrzanowska, K. Saar, G. Beckmann, E. Seemanova, P. R. Cooper, N. J. Nowak, M. Stumm, C. M. Weemaes, R. A. Gatti, R. K. Wilson, M. Digweed, A. Rosenthal, K. Sperling, P. Concannon and A. Reis (1998). Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. Cell **93**(3): 467-476.

von Bulow G. U., J. M. van Deursen and R. J. Bram (2001). Regulation of the T-independent humoral response by TACI. Immunity **14**(5): 573-582.

Vora K. A., K. M. Tumas-Brundage, V. M. Lentz, A. Cranston, R. Fishel and T. Manser (1999). Severe attenuation of the B cell immune response in Msh2-deficient mice. J Exp Med **189**(3): 471-482.

Vorechovsky I., M. Cullen, M. Carrington, L. Hammarström and A. D. Webster (2000). Fine mapping of IGAD1 in IgA deficiency and common variable immunodeficiency: identification and characterization of haplotypes shared by affected members of 101 multiple-case families. J Immunol **164**(8): 4408-4416.

Vorechovsky I., D. Scott, M. R. Haeney and D. A. Webster (1993). Chromosomal radiosensitivity in common variable immune deficiency. Mutat Res **290**(2): 255-264.

Vorechovsky I., A. D. Webster, A. Plebani and L. Hammarström (1999). Genetic linkage of IgA deficiency to the major histocompatibility complex: evidence for allele segregation distortion, parent-of-origin penetrance differences, and the role of anti-IgA antibodies in disease predisposition. Am J Hum Genet **64**(4): 1096-1109.

Vorechovsky I., H. Zetterquist, R. Paganelli, S. Koskinen, A. D. Webster, J. Bjorkander, C. I. Smith and L. Hammarström (1995). Family and linkage study of selective IgA deficiency and common variable immunodeficiency. Clin Immunol Immunopathol **77**(2): 185-192.

Wagner S. D. and M. S. Neuberger (1996). Somatic hypermutation of immunoglobulin genes. Annu Rev Immunol **14**: 441-457.

Waltes R., R. Kalb, M. Gatei, A. W. Kijas, M. Stumm, A. Sobeck, B. Wieland, R. Varon, Y. Lerenthal, M. F. Lavin, D. Schindler and T. Dork (2009). Human RAD50 deficiency in a Nijmegen breakage syndrome-like disorder. Am J Hum Genet **84**(5): 605-616.

Warnatz K., U. Salzer, M. Rizzi, B. Fischer, S. Gutenberger, J. Bohm, A. K. Kienzler, Q. Pan-Hammarström, L. Hammarström, M. Rakhmanov, M. Schlesier, B. Grimbacher, H. H. Peter and H. Eibel (2009). B-cell activating factor receptor deficiency is associated with an adult-onset antibody deficiency syndrome in humans. Proc Natl Acad Sci U S A.

Watson P. and H. T. Lynch (1994). The tumor spectrum in HNPCC. Anticancer Res **14**(4B): 1635-1639.

Weinstock D. M., B. Elliott and M. Jasin (2006). A model of oncogenic rearrangements: differences between chromosomal translocation mechanisms and simple double-strand break repair. Blood **107**(2): 777-780.

Whiteside D., R. McLeod, G. Graham, J. L. Steckley, K. Booth, M. J. Somerville and S. E. Andrew (2002). A homozygous germ-line mutation in the human MSH2 gene predisposes to hematological malignancy and multiple cafe-au-lait spots. Cancer Res **62**(2): 359-362.

Wiesendanger M., B. Kneitz, W. Edelmann and M. D. Scharff (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., A. Vaisman, S. A. Martomo, P. Sullivan, L. Lan, F. Hanaoka, A. Yasui, R. Woodgate and P. J. Gearhart (2005). MSH2-MSH6 stimulates DNA polymerase eta, suggesting a role for A:T mutations in antibody genes. J Exp Med **201**(4): 637-645.

Wu X., C. Y. Tsai, M. B. Patam, H. Zan, J. P. Chen, S. M. Lipkin and P. Casali (2006). A role for the MutL mismatch repair Mlh3 protein in immunoglobulin class switch DNA recombination and somatic hypermutation. J Immunol **176**(9): 5426-5437.

Wu Y., D. Bressette, J. A. Carrell, T. Kaufman, P. Feng, K. Taylor, Y. Gan, Y. H. Cho, A. D. Garcia, E. Gollatz, D. Dimke, D. LaFleur, T. S. Migone, B. Nardelli, P. Wei, S. M. Ruben, S. J. Ullrich, H. S. Olsen, P. Kanakaraj, P. A. Moore and K. P. Baker (2000). Tumor necrosis factor (TNF) receptor superfamily member TACI is a high affinity receptor for TNF family members APRIL and BLyS. J Biol Chem **275**(45): 35478-35485.

Xu S. and K. P. Lam (2001). B-cell maturation protein, which binds the tumor necrosis factor family members BAFF and APRIL, is dispensable for humoral immune responses. Mol Cell Biol **21**(12): 4067-4074.

Xu Y. (2006). DNA damage: a trigger of innate immunity but a requirement for adaptive immune homeostasis. Nat Rev Immunol **6**(4): 261-270.

Xue K., C. Rada and M. S. Neuberger (2006). The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in msh2-/- ung-/- mice. J Exp Med **203**(9): 2085-2094.

Yabuki M., M. M. Fujii and N. Maizels (2005). The MRE11-RAD50-NBS1 complex accelerates somatic hypermutation and gene conversion of immunoglobulin variable regions. Nat Immunol **6**(7): 730-736.

Yan M., J. R. Brady, B. Chan, W. P. Lee, B. Hsu, S. Harless, M. Cancro, I. S. Grewal and V. M. Dixit (2001). Identification of a novel receptor for B lymphocyte stimulator that is mutated in a mouse strain with severe B cell deficiency. Curr Biol **11**(19): 1547-1552.

Yan M., S. A. Marsters, I. S. Grewal, H. Wang, A. Ashkenazi and V. M. Dixit (2000). Identification of a receptor for BLyS demonstrates a crucial role in humoral immunity. Nat Immunol **1**(1): 37-41.

Yan M., H. Wang, B. Chan, M. Roose-Girma, S. Erickson, T. Baker, D. Tumas, I. S. Grewal and V. M. Dixit (2001). Activation and accumulation of B cells in TACI-deficient mice. Nat Immunol **2**(7): 638-643.

Yi W., X. Wu, T. H. Lee, N. A. Doggett and C. Her (2005). Two variants of MutS homolog hMSH5: prevalence in humans and effects on protein interaction. Biochem Biophys Res Commun **332**(2): 524-532.

Yong P. F., R. Chee and B. Grimbacher (2008). Hypogammaglobulinaemia. Immunol Allergy Clin North Am **28**(4): 691-713, vii.

Zenone T. and G. Souillet (1997). [Cancer and primary humoral immunodeficiency]. Bull Cancer **84**(8): 813-821.

Zhu Y., S. Nonoyama, T. Morio, M. Muramatsu, T. Honjo and S. Mizutani (2003). Type two hyper-IgM syndrome caused by mutation in activation-induced cytidine deaminase. J Med Dent Sci **50**(1): 41-46.

Zonana J., M. E. Elder, L. C. Schneider, S. J. Orlow, C. Moss, M. Golabi, S. K. Shapira, P. A. Farndon, D. W. Wara, S. A. Emmal and B. M. Ferguson (2000). A novel X-linked disorder of immune deficiency and hypohidrotic ectodermal dysplasia is allelic to incontinentia pigmenti and due to mutations in IKK-gamma (NEMO). Am J Hum Genet **67**(6): 1555-1562.

APPENDIX I: MSH5 VARIATIONS ASSOCIATE WITH CVID, POTENTIALLY IMPLICATING THE PROTEIN IN CLASS SWITCH RECOMBINATION

In this appendix, one report, *Role for Msh5 in the regulation of Ig class switch recombination* published in the journal *Proceedings of the National Academy of Sciences of the United States of America* is presented (Sekine et al. 2007). This manuscript combines two related but independent data sets. The first series of data revolves around characterizing the role of MSH5 in murine CSR. The remainder of this manuscript examines the role of MSH5 variants in CVID and IgAD – the portion of this project in which I was actively involved.

My major contribution to this manuscript was intellectual input, although some of the laboratory experiments were performed by me as well. The MSH4/5 homologous recombination complex was previously thought to only be involved in homologous recombination-mediated crossing over during meiosis, and not involved in DNA break generation or repair during CSR. I was intimately involved with the development of a model to explain the phenotype observed in patients (and collaborators were seeing in mice). Additionally, in the CVID subgroup meeting of the Behrens lab, I originally proposed the model by which MSH4/5 could bind to CSR intermediates and direct break joining to NHEJ-mediated repair (low microhomology), and in the absence of MSH4/5 binding, less efficient microhomology-mediated processes would moderate end-joining. This model explained both the reduction in CSR and increased dependence on microhomology at switch joints in MSH5-mutant patients.

I performed genotyping of human samples in conjunction with Thearith Koeuth. I designed and conducted quantitative RT-PCR assays to measure MSH5 mRNA expression in patient-derived peripheral blood B cells and Epstein bar virus (EBV)-immortalized patient B cell lines. Additionally, I assisted Keli Hippen with designing and performing quantitative RT-PCR assays to confirm MSH5 expression in mRNA derived from murine B cells.

Sekine H, Ferreira RC, Pan-Hammarstrom Q, Graham RR, Ziemba B, de Vries SS, Liu J, Hippen K, Koeuth T, Ortmann W, Iwahori A, Elliott MK, Offer S, Skon C, Du L, Novitzke J, Lee AT, Zhao N, Tompkins JD, Altshuler D, Gregersen PK, Cunningham-Rundles C, Harris RS, Her C, Nelson DL, Hammarstrom L, Gilkeson GS, Behrens TW. Role for Msh5 in the regulation of Ig class switch recombination. Proceedings of the National Academy of Sciences of the United States of America. 104(17):7193-7198. 2007.

Role for Msh5 in the regulation of Ig class switch recombination

Hideharu Sekine,[a] Ricardo C. Ferreira,[bc] Qiang Pan-Hammarström,[d] Robert R. Graham,[e]

Beth Ziemba,[b] Sandra S. de Vries,[f] Jiabin Liu,[b] Keli Hippen,[b] Thearith Koeuth,[b]

Ward Ortmann,[bc] Akiko Iwahori,[a] Margaret K. Elliott,[a] Steven Offer,[b] Cara Skon,[b]

Likun Du,[d] Jill Novitzke,[b] Annette T. Lee,[g] Nianxi Zhao,[h] Joshua D. Tompkins,[h]

David Altshuler,[e] Peter K. Gregersen,[g] Charlotte Cunningham-Rundles,[i]

Reuben S. Harris,[b] Chengtao Her,[h] David L. Nelson,[j] Lennart Hammarström,[d]

Gary S. Gilkeson,[a] and Timothy W. Behrens[bck]

a Medical University of South Carolina, Charleston, SC 29425;

b University of Minnesota Medical School, Minneapolis, MN 55455;

d Karolinska University Hospital, SE-141 86 Huddinge, Sweden;

e Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA 02142;

f The Netherlands Cancer Institute, 1066 CX, Amsterdam, The Netherlands;

g Feinstein Institute for Medical Research, Manhasset, NY 11030;

h Washington State University, Pullman, WA 99164;

i Mount Sinai School of Medicine, New York, NY 10029; and

j National Cancer Institute, Bethesda, MD 20892

k To whom correspondence should be addressed. E-mail:behrens.tim@gene.com

Communicated by Richard H. Scheller, Genentech, Inc., South San Francisco, CA, February 19, 2007.

[c] Present address: Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080.

ABSTRACT

Ig class switch recombination (CSR) and somatic hypermutation serve to diversify antibody responses and are orchestrated by the activity of activation-induced cytidine deaminase and many proteins involved in DNA repair and genome surveillance. Msh5, a gene encoded in the central MHC class III region, and its obligate heterodimerization partner Msh4 have a critical role in regulating meiotic homologous recombination and have not been implicated in CSR. Here, we show that MRL/lpr mice carrying a congenic H-2b/b MHC interval exhibit several abnormalities regarding CSR, including a profound deficiency of IgG3 in most mice and long microhomologies at Ig switch (S) joints. We found that Msh5 is expressed at low levels on the H-2b haplotype and, importantly, a similar long S joint microhomology phenotype was observed in both Msh5 and Msh4-null mice. We also present evidence that genetic variation in MSH5 is associated with IgA deficiency and common variable immune deficiency (CVID) in humans. One of the human MSH5 alleles identified contains two nonsynonymous polymorphisms, and the variant protein encoded by this allele shows impaired binding to MSH4. Similar to the mice, Ig S joints from CVID and IgA deficiency patients carrying disease-associated MSH5 alleles show increased donor/acceptor microhomology, involving pentameric DNA repeat sequences and lower mutation rates than controls. Our findings suggest that Msh4/5 heterodimers contribute to CSR and support a model whereby Msh4/5 promotes the resolution of DNA breaks with low or no terminal microhomology by a classical nonhomologous end-joining mechanism while possibly suppressing an alternative microhomology-mediated pathway.

Keywords: immunoglobulin subclass deficiency, mismatch repair, Msh4

Abbreviations

CSR class switch recombination

CVID common variable immune deficiency

IgAD IgA deficiency

KO knockout

INTRODUCTION

After appropriate stimulation, B cells undergo class switch recombination (CSR), whereby the functionally rearranged V(D)J DNA segment is recombined with a downstream Ig constant region segment. The biochemistry of CSR is complex and involves the B cell-specific gene activation-induced cytidine deaminase, which initiates both CSR and somatic hypermutation (Min and Selsing 2005). CSR also requires many ubiquitously expressed genes important for detecting DNA mismatches and breaks and regulating DNA repair (Xu 2006). CSR occurs at specific DNA segments called switch (S) regions, which lie upstream of each constant region and contain hotspots for activation-induced cytidine deaminase-mediated cytosine deamination. The ligation of the Sμ region with the downstream S regions is carried out by protein factors that comprise the nonhomologous end joining machinery for DNA repair (Min and Selsing 2005; Xu 2006).

Mismatch repair proteins play a critical role in safeguarding genetic stability. The key proteins for initiation of eukaryotic mismatch repair are homologues of bacterial MutS and MutL. In mammals, there are five MutS (Msh2, Msh3, Msh4, Msh5, and Msh6) and four MutL (Mlh1, Mlh3, Pms1, and Pms2) homologues. Each Mut homologue acts at the DNA repair or recombination site by forming heterodimers; Msh2-Msh6 (MutSα), Msh2-Msh3 (MutSβ), Msh4-Msh5 (MutSγ), Mlh1-Pms2 (MutLα), and Mlh1-Mlh3 (MutLγ) (Svetlanov and Cohen 2004). MutS heterodimers are thought to recruit MutL heterodimers. Experiments using Mut homologue gene-knockout (KO) mice revealed that Msh2-, Msh6-, Mlh1-, and Pms2-deficient animals had decreased efficiency of CSR and somatic hypermutation (Wu et al. 2006). Deficiencies of MutS and MutL genes often result in differences in microhomology lengths at S joints and show three phenotypes, decreased ($Msh2^{-/-}$ and $Mlh3^{-/-}$) (Schrader et al. 2002; Wu et al. 2006), no change ($Msh6^{-/-}$) (Li et al. 2004), or increased ($Mlh1^{-/-}$ and $Pms2^{-/-}$) (Ehrenstein et al. 2001; Schrader et al. 2002) microhomology. The differences in S joint phenotypes

between $Msh2^{-/-}$ mice and $Mlh1^{-/-}$ or $Pms2^{-/-}$ mice suggest the existence of other proteins that function in the same pathway of CSR as Mlh1 and Pms2.

Msh5 and Msh4 are involved in the resolution of DNA Holliday junctions, the four-stranded DNA structures that form during homologous recombination in meiosis (Snowden et al. 2004). Msh4 and Msh5 KO mice are sterile due to an inability to resolve these meiotic chromosomal crossovers (de Vries et al. 1999; Edelmann et al. 1999; Kneitz et al. 2000).

Based on these studies in mice, the Mut homologues are attractive candidate genes for human Ig deficiencies. Selective IgA deficiency (IgAD) (serum IgA <0.05 g/liter) is the most common primary immunodeficiency disorder in man, with a prevalence of ≈1/600 Caucasian individuals (Burrows and Cooper 1997). The selective nature of the CSR defect in IgAD is not understood. Common variable immune deficiency (CVID) is a more severe disease and affects ≈1/25,000 Caucasians. Patients show a marked reduction in serum levels of both IgG (usually <3 g/liter) and IgA (<0.05 g/liter), together with reductions of IgM in about half the cases (<0.3 g/liter). CVID patients have a high incidence of infectious complications and, paradoxically, are prone to autoimmune disorders (Cunningham-Rundles and Bodian 1999).

The available evidence suggests a common genetic basis for IgAD and CVID (Vorechovsky et al. 1995) and individuals with IgAD may transition into CVID. Haplotypes of the MHC show genetic association with IgAD, notably HLA (HLA) A1-B8-DR3 and B14-DR1 (Hammarström and Smith 1983; Olerup et al. 1990; Schaffer et al. 1989). Homozygosity for the A1-B8-DR3 haplotype is a particularly strong risk factor for IgAD in Caucasians, with an incidence reported as high as 13% (Alper et al. 2000). Whereas the association of IgAD and CVID with the MHC is clearly documented, the identity of the genetic effect(s) within the MHC remains controversial, with studies suggesting that class II molecules and/or genes in the centromeric class III region are involved (Schaffer et al. 1989; Vorechovsky et al. 2000; Vorechovsky et al. 1999). Other genes that contribute to CVID include rare mutations in the T

cell costimulatory molecule ICOS (Grimbacher et al. 2003) and TACI (TNFRSF13B) (Castigli et al. 2005; Salzer et al. 2005).

Here, we provide evidence that Msh5 contributes to dysregulated Ig CSR in mice and identify a possible role for MSH5 in human IgAD and CVID.

RESULTS

H-2b Congenic MRL/lpr Mice Show Defects in CSR.

We generated H-2b/b congenic MRL/lpr mice by introgressing the H-2b MHC haplotype from 129/Sv mice onto the MRL/lpr background. After nine generations of backcrossing, animals were genotyped for 136 polymorphic microsatellites, which confirmed that all markers outside the H-2 region were MRL/lpr derived. The congenic H-2b interval measured ≈13 Mb and included the entire MHC region (**Figure AI-1A**). H-2b/b MRL/lpr mice exhibited no differences in disease compared with wild-type animals (Sekine et al. 2006). Strikingly, however, 11/16 (68%) H-2b/b congenics had undetectable serum IgG3 antibodies (**Figure AI-1B**), diminished levels of serum IgA antibodies in older mice, together with elevated serum levels of IgM and IgG2a antibodies. Serum IgG1 and IgG2b levels were similar in H-2b/b and H-2k/k MRL/lpr mice (**Supplementary Figure AI-5**). The deficiency of IgG3 in the H-2b/b congenics was confirmed by ELISpot assays of splenic antibody secreting cells (**Supplementary Figure AI-6**). Importantly, the antibody phenotypes were similar in congenic H-2b/b MRL/lpr animals backcrossed nine generations, and those animals backcrossed >20 generations (data not shown), demonstrating that the genetic effect is stable, shows consistent incomplete penetrance, and is localized to the H-2 region.

Hypomorphic Allele of Msh5 on the H-2b Haplotype.

To identify the gene(s) from the H-2 region contributing to the IgG3 deficiency, we used gene expression microarrays to assay spleen RNA from 8-week-old congenic IgG3pos H-2b/b, IgG3neg H-2b/b, and H-2k/k MRL/lpr littermates. Essentially all of the significant differences in gene expression, with the exception of IgG3 mRNA, were genes encoded within the MHC congenic interval. IgG3 mRNA expression was significantly higher in H-2k/k MRL/lpr mice (average 36,385 affymetrix expression units) compared to IgG3neg H-2b/b mice (average 2,837 affymetrix expression units; $P = 1 \times 10{-}4$) (**Supplementary Table AI-2**). The H-2 Ea gene is deleted on the H-2b haplotype (Mathis et al. 1983) and showed low expression in the H-2b/b congenic spleens. Expression differences were also observed for other class I and II MHC genes, which likely reflect polymorphisms between the H-2b and H-2k haplotypes. Msh5, which is located in the MHC class III region (**Figure AI-1A**), was the only other differentially expressed gene within the MHC region and showed ≈6-fold lower expression in the H-2b congenic spleens (IgG3pos H-2b/b, 100 expression units; IgG3neg H-2b/b, 135 expression units) compared with the wild-type H-2k/k mice (668 expression units; $P = 9 \times 10{-}3$ vs. IgG3pos H-2b/b and $P = 8.2 \times 10{-}3$ vs. IgG3neg H-2b/b) (**Supplementary Table AI-2**). The microarray expression results for Msh5 were confirmed by using TaqMan real-time quantitative PCR (**Figure AI-1C**).

We next evaluated the influence of the low Msh5 levels on B cell class switching in vitro. Surprisingly, there were no differences in the efficiency of in vitro switching to IgG3 (LPS) or IgG1 (LPS + IL-4) between the IgG3neg H-2b/b congenic and control H-2k/k MRL/lpr B cells (**Figure AI-1D** and data not shown). The ability of H-2b congenic B cells to switch in vitro is reminiscent of human IgAD, where in vitro stimulation of B cells from IgAD patients with CD40 and IL-10 induces normal levels of IgA secretion (Briere et al. 1994). Interestingly, immunization of IgG3neg H-2b/b MRL/lpr mice with T-independent stimuli

TNP-LPS or TNP-Ficoll failed to elicit an IgG3 anti-TNP response, whereas IgM (data not shown) and IgG2b responses were intact (**Figure AI-1E**).

We then surveyed several mouse strains and identified two distinct groups: high Msh5 expressers [MRL/lpr (H-2k), AKR (H-2k) (data not shown) and BALB/c (H-2d)] and low Msh5 expressers [129/Sv (H-2b), C57BL/6 (H-2b) and FVB (H-2q)] (**Figure AI-1F**). The difference in B cell Msh5 mRNA expression levels between the two groups was ≈100-fold. The higher relative levels of Msh5 in the B cells of MRL/lpr H-2b congenics, compared with native H-2b strains, may reflect the activated state of B cells in MRL/lpr mice (Theofilopoulos and Dixon 1985). Consistent with this idea, Msh5 was inducible in B cells from C57BL/6 mice (**Supplementary Figure AI-7**), although induced Msh5 expression levels were at least 10-fold lower than high expressers at baseline. Msh5 expression in B cells from high-expressing strains was not inducible (data not shown). We conclude that H-2b/b congenic MRL/lpr mice express a hypomorphic allele of Msh5 and hypothesize that the low expression of Msh5 on the MRL background contributes to the observed antibody phenotype.

Increased Microhomology at IgG3 Switch Junctions in H-2b/b MRL/lpr B Cells.

We next asked whether the IgG3 deficiency observed in the congenic MRL/lpr animals was accompanied by phenotypic differences in IgG3 S joints. Sμ-Sγ3 joints, amplified from splenic B cells of IgG3neg MRL/lpr H-2b/b mice, showed significantly longer segments of microhomology than the S joints from IgG3pos congenic ($P = 0.0012$) or H-2k/k MRL/lpr B cells ($P = 0.0014$; **Figure AI-2**). The H-2b/b IgG3neg mice also had longer microhomology segments at Sμ-Sα joints (**Supplementary Figure AI-8A**). No other significant abnormalities were observed at the S junctions (**Supplementary Tables AI-3 and AI-4**).

To extend these findings, we evaluated B cells in Msh5 KO FVB mice (9). Serum antibody levels and in vitro switching showed no significant differences between $Msh5^{-/-}$ mice

and littermates (**Supplementary Figure AI-9A** and data not shown). Importantly, the Sμ-Sγ3 (**Figure AI-2**) and Sμ-Sα (**Supplementary Figure AI-8A**) joints from splenic B cells of $Msh5^{-/-}$ mice showed significantly increased microhomology compared with wild-type littermates. To address the concern that other genes in tight linkage disequilibrium with Msh5 might be contributing to the observed phenotype, in both the congenics and the Msh5 KOs, we took two approaches. First, we studied KO mice for complement Factor B (Bf), another MHC class III region gene located <500 Kb centromeric of Msh5. Gene targeting for Bf was performed on the 129/Sv background, and the mice studied were backcrossed seven generations onto the C57BL/6 genetic background (Taube et al. 2006). Importantly, Sμ-Sγ3 and Sμ-Sα joints from splenic B cells of C57BL/6 wild-type and $Bf^{-/-}$ C57BL/6 mice showed no significant differences in the lengths of microhomologies (**Supplementary Figure AI-8B and AI-8C**). Second, we evaluated mice deficient for Msh4, the heterodimeric partner of Msh5. Msh4 is located on mouse chromosome 3, a location distinct from Msh5. Similar to $Msh5^{-/-}$ mice, there were no significant differences in serum antibody levels (**SI Fig. 9B**) or in vitro class switching (**Supplementary Figure AI-10**). However, splenic Sμ-Sγ3 and Sμ-Sα joints from $Msh4^{-/-}$ mice showed significantly longer microhomologies compared with wild-type littermates (**Figure AI-2** and **Supplementary Figure AI-8A**). Taken together, these data strongly support a role for Msh5 and Msh4 in the regulation of microhomology at Ig S joints in mice.

Genetic Variants of MSH5 Are Associated with Human CVID and IgAD.

The selective antibody isotype deficiency observed in the H-2b/b congenic MRL/lpr mice prompted us to investigate MSH5 as a candidate gene for human IgAD and CVID. MSH5 mRNA is expressed in human tonsillar B lymphocytes and is present at levels that are higher in CD77+ germinal center B cells than in naïve or memory B cells (Klein et al. 2003)

(**Supplementary Figure AI-11**). Using quantitative PCR, we confirmed the constitutive expression of MSH5 mRNA in both purified peripheral blood human B cells and Epstein–Barr virus transformed B cells (**Supplementary Table AI-5**).

To identify genetic variation in MSH5, we sequenced the 25 coding and noncoding exons of MSH5 together with the promoter region in 96 IgAD and CVID cases and identified five nonsynonymous polymorphisms and a number of SNPs in noncoding regions of the gene (**Supplementary Figure AI-12** and data not shown). To determine whether the identified variation in MSH5 contributed genetic susceptibility to IgAD or CVID, we genotyped relevant SNPs in 207 Swedish IgAD cases, 83 Swedish CVID cases, and 198 Swedish control cases and compared allele frequencies (**Table AI-1**).

We identified two rare nonsynonymous SNPs: a G/T SNP, Q292H, in exon 11, which was present in one Swedish CVID case and absent in all controls tested; and a G/T SNP, C580G, in exon 19, which was present in 2 of 212 IgAD patients and not found in either controls or CVID cases (**Supplementary Table AI-6**). We also identified SNPs in exon 3 (L85F, rs28381349) and exon 24 (P786S, rs28399984). Interestingly, the L85F SNP was always found together with the P786S SNP ($D' = 1$), indicating they are located on the same chromosomal segment. By oligotyping HLA-B and DR alleles in the Swedish cases, we determined that the L85F/P786S allele is present on the ancestral HLA B14-DR1 haplotype. Fourteen of 16 L85F/P786S cases (88%) were DR1 and/or B14 positive (**Supplementary Figure AI-12**). Importantly, B14-DR1 is one of the MHC haplotypes that has shown strong genetic association with IgAD and CVID (Olerup et al. 1990). The MSH5 L85F/P786S allele was present in the Swedish controls at a frequency of 1.8% and was enriched ≈2-fold in IgAD cases (3.6%) and to a lesser extent in CVID (2.4%). These differences did not reach statistical significance (**Table AI-1**). Another nonsynonymous SNP, P29S, in exon 2 (rs2075789) was a

frequent polymorphism in the control population (12.1%) and was not enriched in patients (**Supplementary Table AI-6**).

SNP rs3131378, located in intron 12 of MSH5, is present on the extended A1-B8-DR3 MHC haplotype. Of 124 haplotypes carrying SNP rs3131378, 111 were positive for DR3 (90%), and 107/127 were positive for B8 (84%). rs3131378 was strongly associated with IgAD (31.4% allele frequency compared with 11.9% in controls, $P = 2.1 \times 10^{-11}$) (**Table AI-1**). CVID patients showed only a modestly increased allele frequency of rs3131378 (15.7%).

We next typed these SNPs in an independent cohort of 102 United States Caucasian CVID cases and 488 U.S. controls. Three U.S. IgAD cases were also typed. Although not reaching statistical significance, the L85F/P786S double missense allele of MSH5 was more frequent in the CVID cohort (5.4%) than in controls (3.2%). All three U.S. IgAD cases typed were heterozygous for the L85F/P786S allele. The MSH5 allele that was present on the extended DR3 haplotype was also modestly increased in U.S. CVID cases (13.2% vs. 9.7% in controls) (**Table AI-1**).

In a combined analysis of the Swedish and U.S. cohorts, the L85F/P786S allele showed significant association with IgAD ($P = 5.8 \times 10^{-3}$), borderline significance in CVID ($P = 0.058$), and evidence for association in the combined IgAD+CVID analysis ($P = 1.8 \times 10^{-3}$). rs3131378 showed strong association in the combined IgAD analysis ($P = 7.9 \times 10^{-11}$) and significant evidence in the pooled CVID analysis ($P = 0.026$) (**Table AI-1**).

We next performed yeast two-hybrid assays to measure the interaction of the L85F/P786S variant MSH5 protein with MSH4, because both L85F and P786S are located within identified MSH4-interacting domains of MSH5 (**Supplementary Figure AI-12**) (Yi et al. 2005). Using sensitive liquid β-galactosidase assays, we found that the L85F/P786S MSH5 protein variant showed a diminished ability to bind to MSH4 as compared with the wild-type MSH5 (**Figure AI-3**).

Given the background frequency of the MSH5 L85F/P786S allele in the population, it was important to determine whether controls carrying this allele were IgA deficient. We measured IgA levels in the plasma of 11 controls heterozygous for the L85F/P786S allele. All showed normal IgA levels (ranging from 1.0 to 4.4 g/liter), suggesting there is incomplete penetrance for the Ig deficiency phenotype associated with the L85F/P786S allele, consistent with the hypothesis that IgAD is a complex multigenic disease.

Increased Microhomology and Lower Mutation Rate at Sμ-Sα1 Joints of CVID Patients Carrying Disease Associated MSH5 Alleles.

We next investigated whether patients with IgAD and CVID carrying the identified MSH5 alleles showed differences in Ig S joint phenotypes. Sμ-Sα1 joints were amplified from peripheral blood DNA in three groups of controls: healthy donors lacking MSH5 or TACI polymorphisms (Control 1), healthy donors homozygous for rs3131378 on the B8-DR3 extended haplotype (Control 2, DR3++), and healthy donors heterozygous for the MSH5 L85F/P786S allele (Control 3, L85F/P786S). These sequences were compared with those amplified from three patient groups: CVID patients carrying TACI mutations and lacking MSH5 nonsynonymous or DR3 alleles (Patient 1, TACI*), DR3++ patients (Patient 2, DR3++), and patients carrying nonsynonymous alleles of MSH5 (Patient 3, MSH5*). The CVID group carrying TACI mutations tested the hypothesis that S joint phenotypes would differ between patients carrying TACI and MSH5 alleles.

Strikingly, CVID patients carrying MSH5 nonsynonymous polymorphisms displayed significantly longer stretches of Sμ-Sα1 microhomology than the various control groups (median 9 bp vs. 2 bp in controls; $P = 1.9 \times 10^{-7}$) (**Figure AI-4A**). DR3++ CVID patients demonstrated a similar long microhomology phenotype (median 8 bp, $P = 3 \times 10^{-4}$). Importantly, S junction microhomology in CVID patients carrying alleles of the TACI gene and

lacking any of the disease associated MSH5 alleles showed levels of microhomology similar to controls (**Figue AI-4A** and **Supplementary Table AI-7**). Thus, the long microhomology phenotype was specific to CVID cases carrying disease-associated MSH5 alleles.

We also found differences in S joint mutation rates between the groups. In controls lacking disease-associated MSH5 alleles, the mutation rate across the entire S joints averaged five mutations per 1,000 bp (**Supplementary Table AI-8**). In contrast, there were far fewer mutations in joints from patients carrying MSH5 nonsynonymous polymorphisms (1.3 mutations per 1,000 bp; $P = 2 \times 10^{-12}$) or DR3++ (0.3 mutations per 1,000 bp; $P = 1.3 \times 10^{-7}$). Interestingly, mutation rates were also significantly lower in the DR3++ controls (3.2 mutations per 1,000 bp, $P = 0.01$) compared with other control groups. Across all of the groups studied, ≈90% of the S region mutations were targeted to dG:dC base pairs, but no differences were observed for the rate of transitions at dG:dC base pairs (**Supplementary Figure AI-13** and **Supplementary Table AI-8**). We also analyzed Sμ-Sα joints from IgAD patients carrying the associated MSH5 alleles, and the results mirrored those observed in CVID (**Supplementary Table AI-9**).

To investigate whether mutations in MSH5 could also be contributing to alterations in the CSR process to other Ig isotypes, we characterized Sμ-Sγ3 junctions from a group of CVID patients carrying MSH5 disease-associated alleles. No significant differences in the length of microhomology were observed among the different groups (**Supplementary Table AI-10**). Of interest, and similar to the findings in Sμ-Sα1 joints, the mutation rate across Sμ-Sγ3 joints was lower for both MSH5 L85F/P786S and DR3+/+ patients compared with controls (**Supplementary Table AI-11**).

Finally, we examined the targeting of breakpoints to either pentamers or activation-induced cytidine deaminase hotspots and the pattern of alignment of pentamer repeats at the S junctions. In controls, ≈50% of the Sμ and Sα1 breakpoints occurred within pentamer repeats.

In contrast, breakpoints from MSH5 L85F/P786S patients were significantly targeted to pentamer motifs at Sμ (82%, $P = 5.4 \times 10^{-3}$) (**Supplementary Table AI-7**) but not at Sα1 (**Supplementary Table AI-11**). Controls carrying the MSH5 L85F/P786S allele showed a similar preferential targeting of Sμ breakpoints to pentamers (75%, P = 0.013) as observed in L85F/P786S patients. More significantly, we found that the vast majority of Sμ-Sα1 junctions from MSH5 L85F/P786S (95%) and DR3++ (100%) patients showed an "in-phase" alignment of pentamer motifs, whereas in control junctions, only ≈50% of pentamer motifs were aligned (**Figure AI-4B** and **Supplementary Figure AI-14**).

DISCUSSION

We found that ≈70% of H-2b/b congenic MRL/lpr mice carrying a hypomorphic Msh5 allele were deficient in serum IgG3, and nucleotide sequence analysis of Ig S junctions revealed increased donor/acceptor microhomology. Similar long microhomologies at S junctions were observed in KO mice for Msh5 and Msh4. Furthermore, we identified several alleles of MSH5 in humans that show genetic association with CVID and IgAD, including one (L85F/P786S) where the encoded mutant protein showed reduced binding affinity to its heterodimerization partner MSH4. Similar to the phenotype observed in the mice, Sμ-Sα1 joints from patients with associated MSH5 alleles showed increased microhomology, together with a reduced mutation rate, an increased in-phase alignment of pentamer repeats at the junctions, and targeting of Sμ breaks to pentamers (**Supplementary Figure AI-15**). Given these findings, what might be the mechanism by which Msh5 participates in the complex biochemistry of CSR?

Msh5 has a well-characterized role in resolving Holliday junctions that form between homologous DNA strands during meiosis (8–10). Heterodimers of Msh5 and Msh4 are postulated to form a "sliding clamp" on DNA and serve as scaffolding for the recombination machinery including the DNA repair proteins Mlh1 and Pms2 (8). We envision a similar

function for Msh4/5 during the early stages of intra-chromosomal "synapsis" of Sμ to Sγ3 or Sα, thereby facilitating the recruitment of proteins required for nonhomologous end joining.

An important observation relevant to the current data are the higher level of sequence homology between Sμ and Sγ3 regions than between Sμ and any of the other S regions in mice (1). Similarly, in humans the Sμ region shows high levels of overall homology with the Sα region (≈70%) and much less homology (≈20%) with the various IgG S regions (Dunnick et al. 1993). We speculate that, given its potential contribution to IgAD, MSH5 in humans may have a specific role in facilitating CSR between Sμ and Sα. There is more IgA produced in the body than any other Ig isotype and this function of MSH5 may have evolved to ensure high-level IgA production for mucosal defenses.

Another important question is the relationship between Msh5 and the long S joint microhomologies. S junctions with extended microhomology are rarely found in the B cells of healthy individuals and may reflect the activity of an alternative microhomology mediated end-joining (MMEJ) pathway, which uses homology searching and exonuclease activity to ligate homologous DNA strands with 3′ or 5′ overhangs. Increased S joint microhomology is also found in humans with ATM and DNA Ligase IV missense mutations (Pan et al. 2002; Pan-Hammarström et al. 2005) and in $Pms2^{-/-}$ (Ehrenstein et al. 2001) and $Mlh1^{-/-}$ (Schrader et al. 2002) mice, suggesting that these proteins may form a complementation group for CSR. If this group of genes is important for recruitment of the nonhomologous end joining machinery, reduced function of these proteins may result in a net loss in efficiency of CSR, as observed in IgAD and CVID, and an increased dependence on microhomology-directed mechanisms for alignment and ligation of S joints. The low mutation rate noted in long microhomology S joints may reflect exonuclease activity that is required for MMEJ, which, we postulate, could "erase" the footprints of activation-induced cytidine deaminase activity at the initial double-strand break, such that the ligated ends of the resulting S regions are well upstream or downstream

from the initial site of double-strand break and in an area of reduced mutation frequency. We are also intrigued by the possibility that Msh5 may have anti-recombinational activity and secondarily function to suppress MMEJ of S joints (**Supplementary Figure AI-16**). Many oncogenic chromosome translocations in the Ig S region contain short stretches of microhomology between the donor and acceptor DNA strands (Ramiro et al. 2004; Weinstock et al. 2006). Thus, active suppression of the MMEJ pathway may reduce the number of chromosomal translocations resulting from CSR. Interestingly, the frequency of malignant B cell lymphomas is increased in patients with CVID (Cunningham-Rundles and Bodian 1999).

Antibody deficiencies were observed in congenic H-2b/b MRL/lpr mice, but not in $Msh4^{-/-}$ or $Msh5^{-/-}$ mice, whereas long S joint microhomologies were found in each strain. MRL/lpr mice have strong spontaneous self-antigen driven B cell responses in vivo, with antibody levels up to 10-fold higher than non-autoimmune animals (Theofilopoulos and Dixon 1985). We speculate that the genetically based activated B cell phenotype in MRL/lpr mice accentuates the defects in CSR caused by the hypomorphic allele of Msh5 in the congenics. It seems likely that the IgG3 deficiency in MRL/lpr H-2b/b mice reflects specific interactions between the hypomorphic H-2b Msh5 allele and other genes in the MRL/lpr background.

As for the human studies, the nonsynonymous alleles of MSH5 identified in the current study are rare (Q292H and C580G) or uncommon (L85S/P786S), and thus our power to definitively conclude that these alleles contribute to immune deficiency based on genetic data is limited. We believe it is important to interpret the human genetic data under the hypothesis that IgAD and CVID are complex genetic diseases. As shown here, Ig deficiencies were not observed in controls heterozygous for MSH5 nonsynonymous alleles. However, there were subtle, yet significant, changes in S joint phenotypes in controls carrying the various MSH5 alleles: decreased switch joint mutation rates in DR3++ controls and increased targeting of Sμ breakpoints to pentamers in MSH5 L85F/P786S controls (**Supplementary Figure AI-15** and

Supplementary Table AI-8). Furthermore, B8-DR3 is a common MHC haplotype in the Caucasian population (≈10% allele frequency), yet the vast majority of DR3++ individuals are not immune deficient. We have not yet identified functional effects of the MSH5 allele carried on the A1-B8-DR3 extended haplotype. Although there are seven unique SNPs within the MSH5 gene found only on this extended haplotype, there are no nonsynonymous polymorphisms, and further work will be required to determine whether this allele is associated with altered splicing, expression, folding or inducibility of MSH5 during CSR. Because of the high level of linkage disequilibrium on the B8-DR3 haplotype, it is currently not possible to rule out the potential role of additional genes on the haplotype contributing to CSR.

In summary, these data provide evidence that the contribution of Msh5 to Ig CSR is complex and likely regulatory. Antibody deficiency was only observed in the congenic MRL/lpr mice, and inbred strains carrying hypomorphic alleles of Msh5 (e.g., C57BL/6, 129/Sv) did not show Ig deficiencies, possibly due to balancing selection of the MHC region (Hughes and Yeager 1998). In humans, the various mutant MSH5 alleles identified may not be sufficient by themselves to cause clinically significant antibody deficiencies and may be compensated by other genes that confer disease resistance. Thus, the disease state in IgAD and CVID is likely a result of complex interactions with other susceptibility genes and possibly environmental factors, similar to other multigenic complex diseases in humans.

MATERIALS AND METHODS

Mice

129/Sv (H-2b), C57BL/6 (H-2b), BALB/c (H-2d), FVB (H-2q) and MRL/lpr (H-2k) mice were purchased from The Jackson Laboratory (Bar Harbor, ME). H-2b/b MRL/lpr mice were generated as described in (Sekine et al. 2006). Msh5-deficient mice were described in (de Vries et al. 1999) and were bred onto the FVB (H-2q) background for >15 generations. Bf-gene

KO mice were provided by J. Thurman (University of Colorado, Boulder, CO) (Taube et al. 2006), and Msh4-gene KO mice, backcrossed to C57BL/6, were provided by W. Edelmann (Albert Einstein College of Medicine, Bronx, NY) (Kneitz et al. 2000).

Human DNA Samples

DNA from 207 IgAD and 83 CVID Swedish cases and 198 controls were collected at the Karolinska Institute (Stockholm, Sweden). Three IgAD and 102 CVID U.S. cases were collected at the Mt. Sinai Hospital (New York, NY) and at the NCI/NIH (Bethesda, MD). 488 healthy U.S. controls were selected from the New York Health Project (Mitchell et al. 2004). All U.S. patients and controls are of self-reported European-Caucasian ancestry. Informed consent was obtained from all subjects, and the studies were approved by human subjects research institutional review boards.

MSH5 Sequencing and Genotyping.

The 25 exons and 1 Kb of the promoter of MSH5 were sequenced in 63 CVID cases from the U.S. or Sweden, and 33 IgAD cases from Sweden at the Broad Institute. Automated sequence analysis software (SNP_COMPARE) and manual examination was used to screen the sequencing files. Twenty-seven "high-quality" SNPs were identified, of which, 17 were already described in dbSNP (v124), and 10 SNPs were new. Additional sequencing was performed to resolve discrepancies or to fill-in missing data. Genotyping assays for the identified SNPs were performed by using both Taqman and Sequenom platforms. Primer and probe sequences are shown in **Supplementary Table AI-12**. Additional details are provided in SI Materials and Methods.

Mouse and human switch junction sequence alignments are also provided in **SI Appendices 1 and 2** (available online at http://www.pnas.org/content/104/17/7193/suppl/DC1).

ACKNOWLEDGEMENTS

We thank the many patients and physicians for their contributions. These studies were supported by Fundação para a Ciência e Tecnologia, Portugal Fellowship SFRH/BD/16281/2004 (to R.C.F.), National Institutes of Health Grants U19 AI067152 and AR043274, and the Swedish Research Council.

FOOTNOTES

The authors declare no conflict of interest.

This article contains supporting information online at www.pnas.org/cgi/content/full/0700815104/DC1.

FIGURES

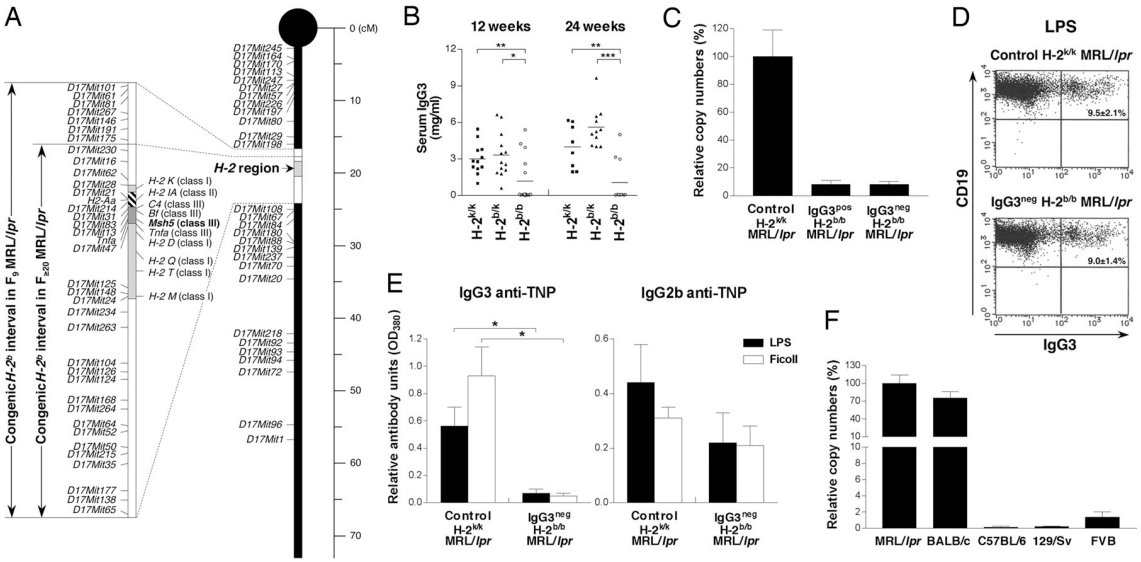


Figure AI-1. Serum IgG3 deficiency, Msh5 gene expression, and CSR in H-2b/b congenic MRL/lpr mice. (A) Map of the 129/Sv congenic interval in F9 and F≥20 congenic H-2b/b MRL/lpr mice. The microsatellite markers and gene polymorphisms used to characterize the introgressed region are shown. (B) Serum IgG3 levels in the F9 H-2k/k, H-2b/k, and H-2b/b MRL/lpr mice. n = 12–16 mice in each group at 12 weeks of age. The number of mice in each group decreased with aging because of mortality. Bars indicate mean values. (C) Msh5 mRNA expression levels were measured in cDNA from splenic B cells of H-2k/k MRL/lpr mice and IgGpos and IgGneg H-2b/b MRLlpr congenic mice (n = 3 each) (D) CSR of splenic B cells was induced in vitro with LPS for class switch induction to IgG3. Representative FACS plots show the percentage of CD19+ IgG3 positive cells from IgG3pos H-2k/k and IgG3neg H-2b/b MRL/lpr mice. Numbers shown are average percentage ± SEM switched cells for three mice in each group. (E) IgG3pos H-2k/k and IgG3neg H-2b/b MRL/lpr mice were immunized with TNP-LPS or TNP-Ficoll, and IgG2b (Right) and IgG3 (Left) anti-TNP responses were measured at 2 weeks. Serum OD380 values are represented on the y axis. Data shown represent the mean ± SEM; n = 10 in each group. (F) Msh5 expression profile in BALB/c (H-2d) (n = 4), 129/Sv (H-2b) (n = 2), C57BL/6 (H-2b) (n = 3), and FVB (H-2q) (n = 3) mice, using quantitative PCR (mean ± SEM). (C and F) Data represent relative Msh5 mRNA copy numbers when compared with resting B cells from H-2k/k MRL/lpr mice (H-2k/k MRL/lpr = 100%; mean ± SEM). [low asterisk], $P < 0.05$; [low asterisk][low asterisk], $P < 0.01$; [low asterisk][low asterisk][low asterisk], $P < 0.001$. P values were calculated by using two-tailed Student's t tests.

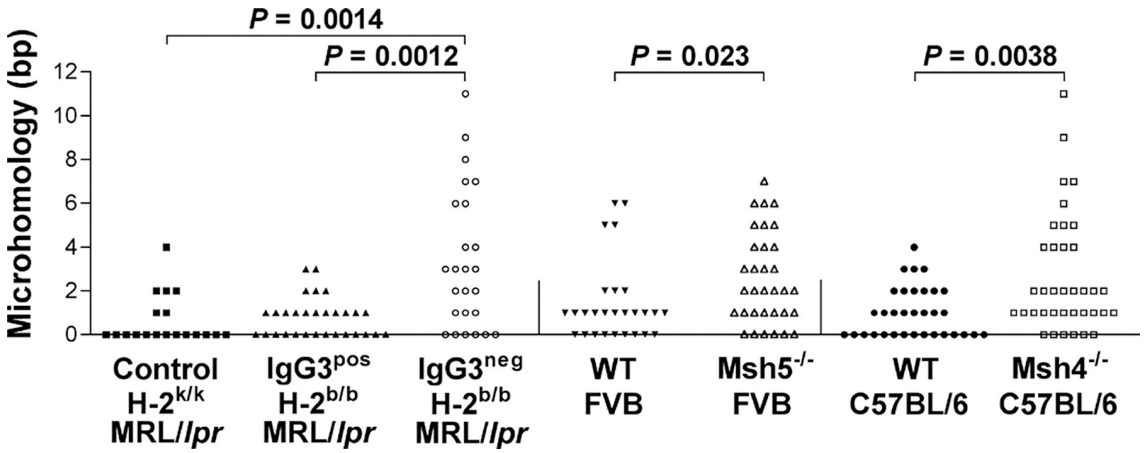


Figure AI-2. Increased microhomology at Sμ-Sγ3 junctions in IgG3neg H-2b/b congenic MRL/lpr, Msh5−/− FVB, and Msh4−/− C57BL/6 mice. S joints were amplified from three 6- to 8-week-old mice in the H-2k/k MRL/lpr, IgG3pos H-2b/b MRL/lpr, wild-type FVB, Msh5−/− FVB, and wild-type C57BL/6 groups, and four 6- to 8-week-old mice in the IgG3neg H-2b/b MRL/lpr and Msh4−/− C57BL/6 groups. Each dot represents the number of nucleotides of donor/acceptor identity at the junction for an individual S joint. P values were calculated by using two-tailed Mann–Whitney tests.

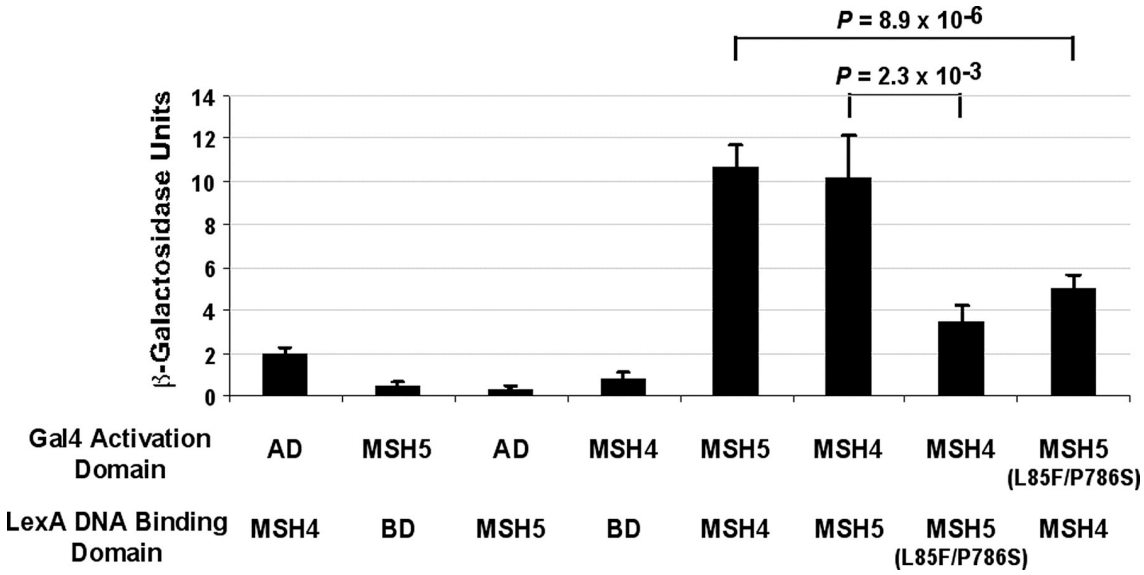


Figure AI-3. MSH5 L85F/P786S variant has reduced binding affinity to MSH4. Yeast two-hybrid assays were performed to assess the ability of the wild-type MSH5 and MSH5 L85F/P786S variant proteins, fused with either a LexA DNA binding domain (BD) or a Gal4 activation domain (AD) to interact with wild-type MSH4. The strength of interaction was measured by using a liquid β-galactosidase assay. Data represent mean ± SE of nine replicates from three independent experiments for the AD-MSH5 L85F/P786S interaction with BD-MSH4, and 12 replicates for all other conditions in four independent experiments. Western blots of whole yeast lysates confirmed equivalent expression of the wild-type and MSH5 L85F/P786S proteins (**SI Fig. 17**).

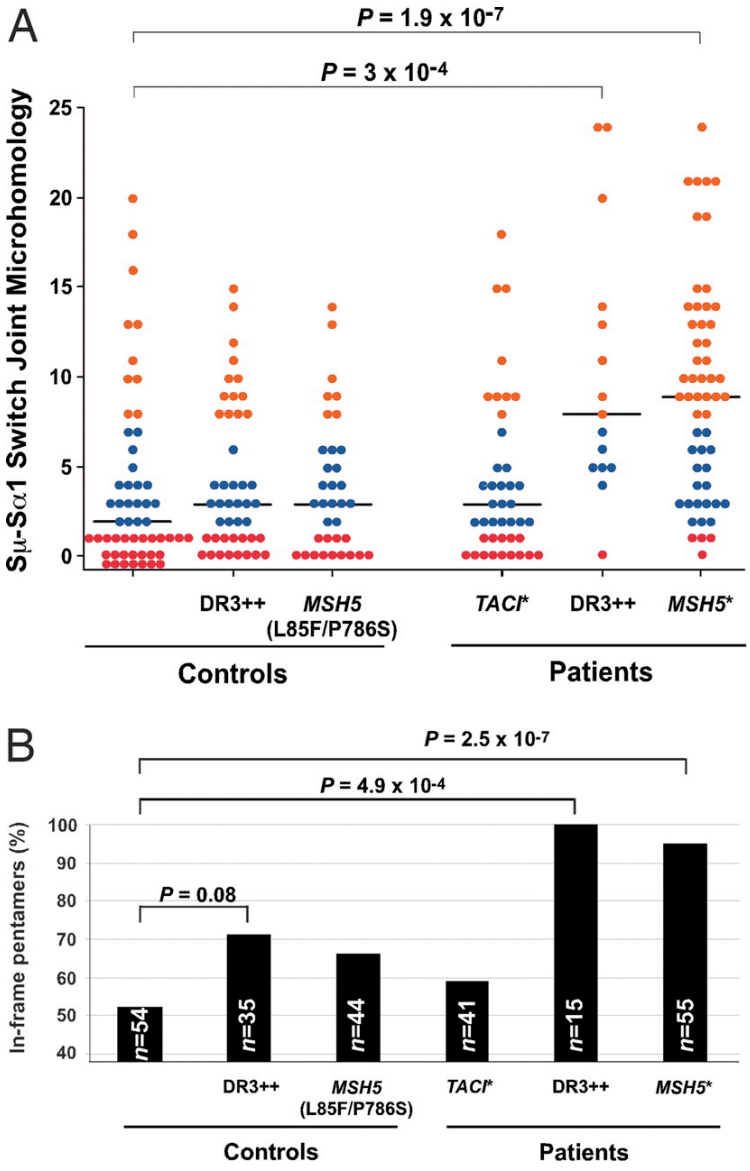
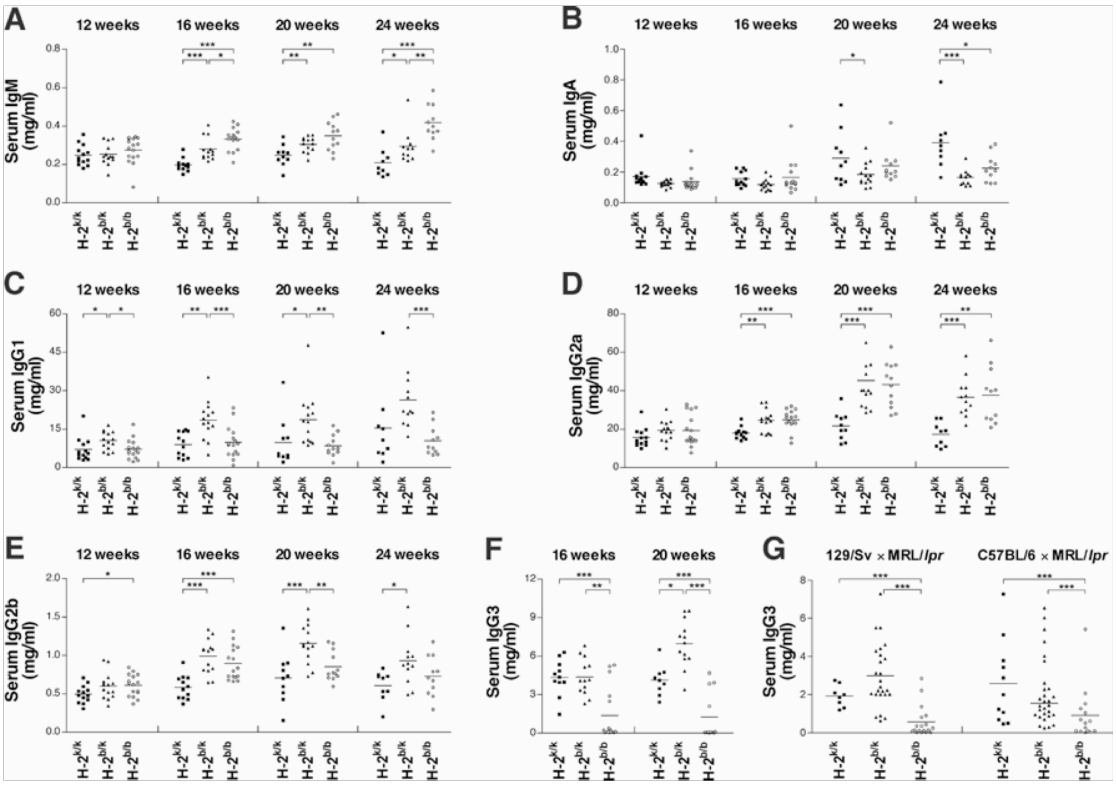
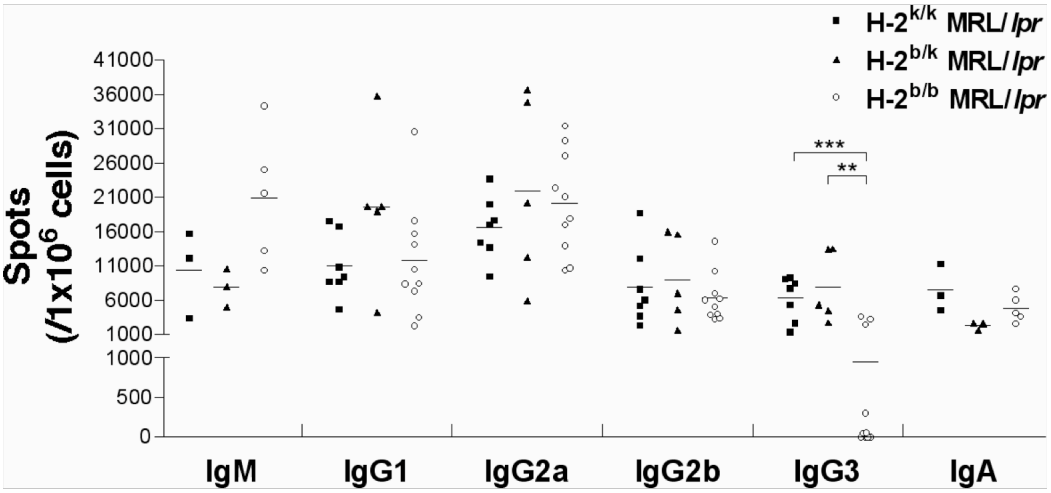


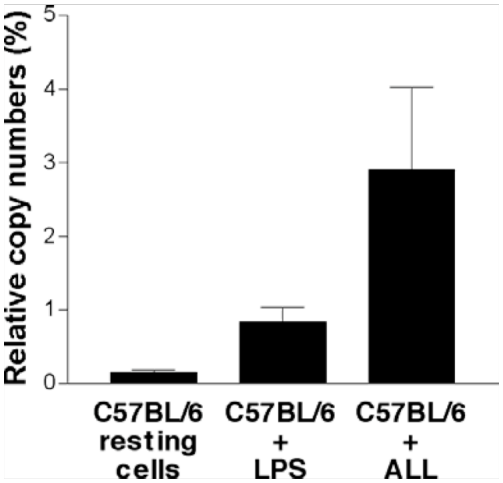
Figure AI-4. Extended microhomology at B cell Ig switch joints of CVID patients carrying associated alleles of MSH5. (A) Distribution of microhomology length in Sμ-Sα1 junctions from CVID patients and controls. Each dot represents the length of microhomology of an independent S joint. The unlabeled group of controls lack any MSH5 nonsynonymous or DR3 alleles. MSH5 (L85F/P786S), heterozygote for the L85F/P786S allele. DR3++, homozygous for the rs3131378 SNP on the extended B8-DR3 MHC haplotype. TACI*, carrying one or more TACI missense mutations. Red, 0–1 bp microhomology; blue, 2–7 bp microhomology; orange, ≥ 8 bp microhomology. P values were calculated by using Mann–Whitney tests. (B) The percentage of S joints where there was "in-frame" alignment of pentamer repeat units between germline Sμ and Sα1 is represented. n = number of joints sequenced and examined. Statistical significance was determined by using two-tailed Fisher's exact tests.



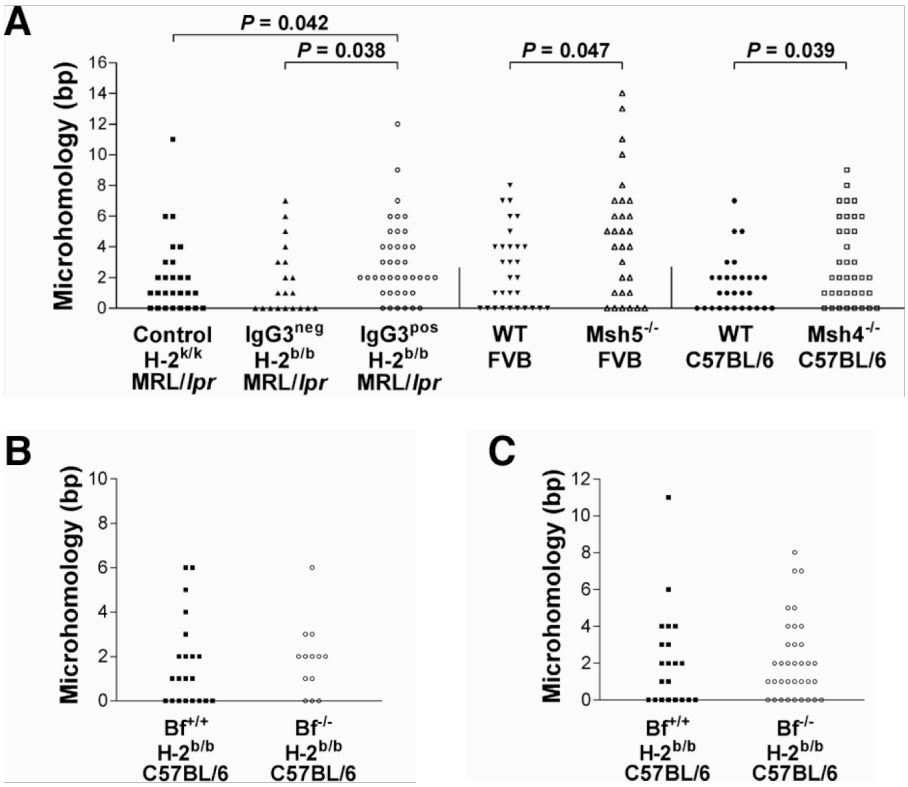
Supplementary Figure AI-5. Serum antibodies profile in H-2b/b congenic MRL/lpr mice. Serum IgM (A), IgA (B), IgG1 (C), IgG2a (D), IgG2b (E), and IgG3 (F) levels in the 129/Sv strain-derived F9 H-2k/k, H-2b/k, and H-2b/b MRL/lpr mice were determined by ELISA, using standards of known concentration. n = 12-16 mice in each group at 12 weeks of age. The number of the mice in each group decreased with aging due to mortality. (G) Serum IgG3 levels in 12-week-old 129/Sv and C57BL/6 strain derived F3 H-2k/k, H-2b/k and H-2b/b MRL/lpr mice were determined by ELISA. n = 8-30 mice in each group. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. P values were calculated using two-tailed Student's t-tests. Bars indicate mean values.



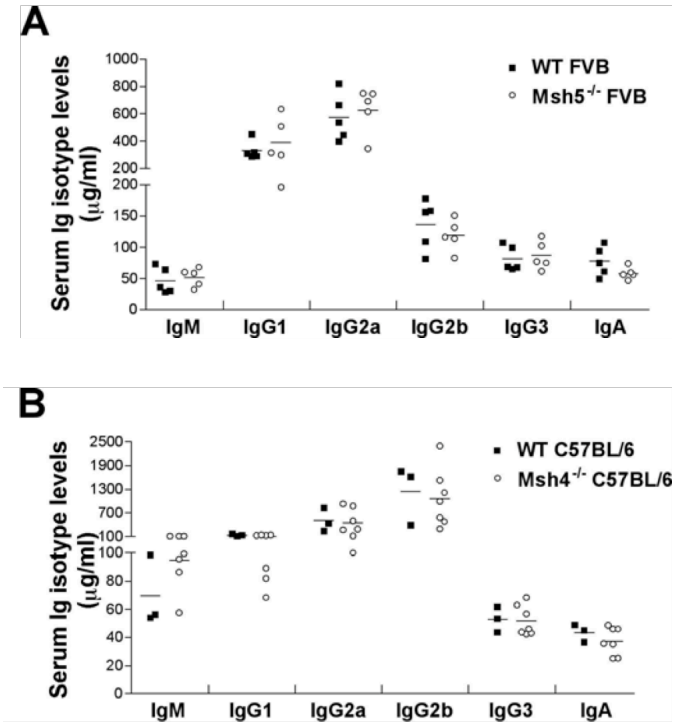
Supplementary Figure AI-6. Populations of antibody isotype secreting B cells in the spleens of H-2b/b congenic MRL/lpr mice. Enzyme-linked immunospot (ELISpot) assays were performed in whole spleen cell preparations of 32-week-old F9 H-2k/k, H-2b/k and H-2b/b MRL/lpr mice to detect antibody isotype secreting B cells. For the IgG ELISpot assays: H-2k/k, n = 7; H-2b/k, n = 5; and H-2b/b, n = 10. For the IgM and IgA ELISpot assays: H-2k/k, n = 3; H-2b/k, n = 3; and H-2b/b, n = 5.



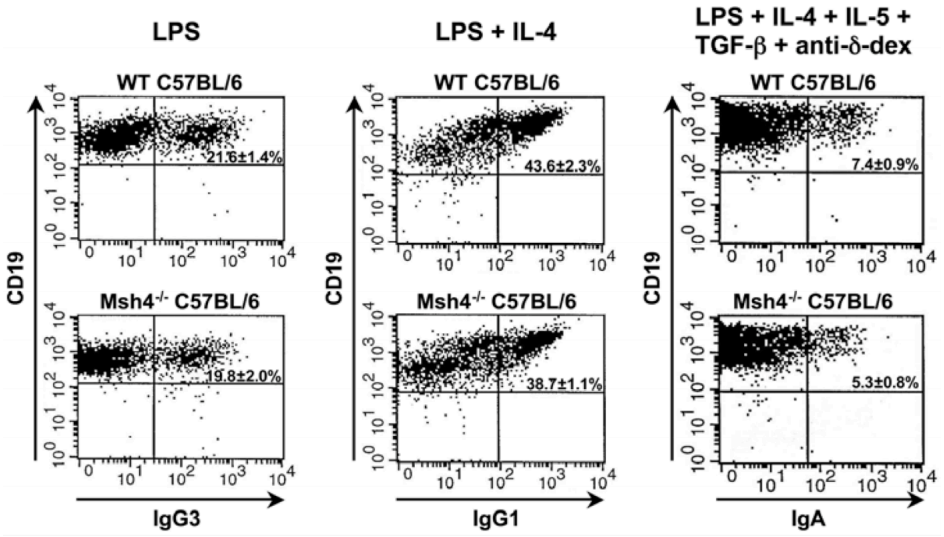
Supplementary Figure AI-7. Msh5 gene expression following stimulation in C57BL/6 B cells. "ALL" refers to induction for class switch to IgA with LPS, IL-4, IL-5, TGF-b, and anti-d-dextran. Data represent relative Msh5 mRNA copy numbers when compared with resting B cells from H-2k/k MRL/lpr mice (H-2k/k MRL/lpr = 100%; mean ± SEM). n = 5 in each group.



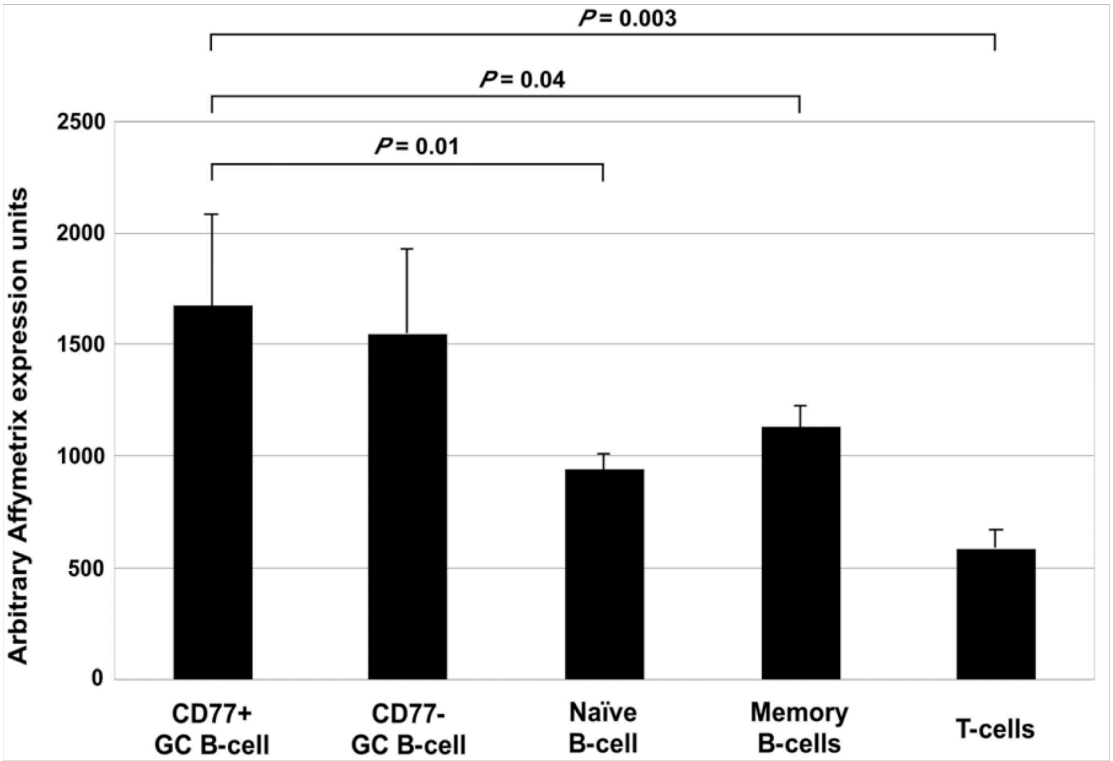
Supplementary Figure AI-8. Long switch junction microhomologies in IgG3neg H-2b/b congenic MRL/lpr, Msh5-/- FVB, and Msh4-/- C57BL/6 mice. (A) Increased microhomology at Sm-Sa junctions in the IgG3neg H-2b/b MRL/lpr, Msh5-/- FVB and Msh4-/- C57BL/6 mice. Switch joints were amplified from three mice in the 6-8 week old H-2k/k MRL/lpr, IgG3pos H-2b/b MRL/lpr, wild-type FVB, Msh5-/- FVB, and wild-type C57BL/6 groups and four mice in the 6-8 week old IgG3neg H-2b/b MRL/lpr and Msh4-/- C57BL/6 group. (B and C) No differences were detected in the length of microhomology at Sm-Sg3 (B) and Sm-Sa (C) junctions between C57BL/6 Factor B (Bf) knockout mice and wild-type C57BL/6 mice. Switch joints were amplified from four mice in each group. Each dot represents the number of nucleotides of donor/acceptor identity at the junction for an individual switch joint. P values were calculated using two-tailed Mann-Whitney tests.



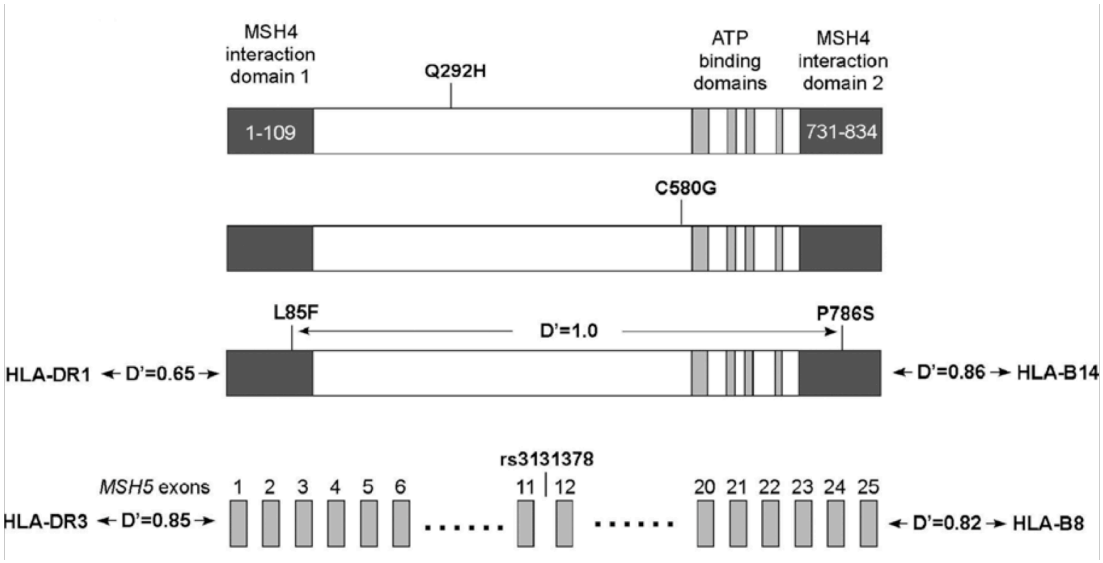
Supplementary Figure AI-9. Serum Ig isotype levels in Msh5-/- FVB, Msh4-/- C57BL/6, and wild-type mice. (A) Serum Ig isotype levels in 6-8 week old wild-type and Msh5-/- FVB mice determined by ELISA, using standards of known concentration (n = 5 in each group). (B) Serum Ig isotype levels in 6-8 week old wild-type (n = 3) and Msh4-/- C57BL/6 (n = 5) mice determined by ELISA using standards of known concentration.



Supplementary Figure AI-10. In vitro Ig CSR efficiency in B cells from 8-week-old wild-type and Msh4-/- C57BL/6 mice. CSR was induced in vitro by incubating purified B cells with LPS for class switch induction to IgG3 (Left), the combination of LPS and rIL-4 for class switch to IgG1 (Center), or the combination of LPS, IL-4, IL-5, TGF-b, and anti-d-dextran to induce class switching to IgA (Right). The average percentage of CD19+ cells that are sIgG3, sIgG1, or sIgA is shown in each panel based on analysis of groups. Data shown represent the mean ± SEM. For sIgG3 and sIgG1 analysis, n = 6 in each group. For sIgA analysis, n = 4 in wild-type, n = 3 in Msh4-/- C57BL/6 mice.



Supplementary Figure AI-11. MSH5 mRNA expression in purified tonsillar human B and T cells. Two-tailed t tests were used to compare MSH5 mRNA levels in CD77+ GC B-cells with the MSH5 expression levels in the other groups. Data are from a publicly available Affymetrix microarray dataset (Sekine et al. 2006).



Supplementary Figure AI-12. MSH5 alleles contribute to CVID and IgAD. Shown are the four MSH5 alleles that are associated with CVID and IgAD and their location in the MSH5 cDNA (Q292H, C580G and L85F/P786S) or genomic sequence (rs3131378/DR3). The Q292H and C580G nonsynonymous alleles are rare and were found in Swedish CVID and IgAD patients but not in controls (see Table 1 for allele frequencies). The L85F and P786S polymorphisms show perfect linkage disequilibrium (D' = 1.0) and thus are carried on the same ancestral allele (L85F/P786S). Both polymorphisms localize to previously defined MSH4 interaction domains. The L85F/P786S allele shows strong linkage disequilibrium to the HLA-DR1 and -B14 alleles, which have previously shown association with IgAD and CVID. The rs3131378 SNP is localized to MSH5 intron 12, and marks an allele of MSH5 that is strongly linked to both DR3 and HLA-B8. D' is a measure of linkage disequilibrium, and D' = 1.0 indicates no historical recombination between two alleles.

Control 1
97 mutations

	To A	C	G	T	
A	-	0	4	1	5.2%
C	6	-	18	20	45.4%
G	24	11	-	8	44.3%
T	1	1	3	-	5.2%

Patient 1 (TACI*)
65 mutations

	To A	C	G	T	
A	-	1	1	0	3.1%
C	3	-	13	13	44.6%
G	18	3	-	10	47.7%
T	1	2	0	-	4.6%

Control 2 (DR3++)
49 mutations

	To A	C	G	T	
A	-	1	2	0	6.1%
C	2	-	6	13	42.9%
G	15	5	-	3	46.9%
T	0	1	1	-	4.1%

Patient 2 (DR3++)
2 mutations

	To A	C	G	T	
A	-	0	0	0	0.0%
C	0	-	0	1	50.0%
G	1	0	-	0	50.0%
T	0	0	0	-	0.0%

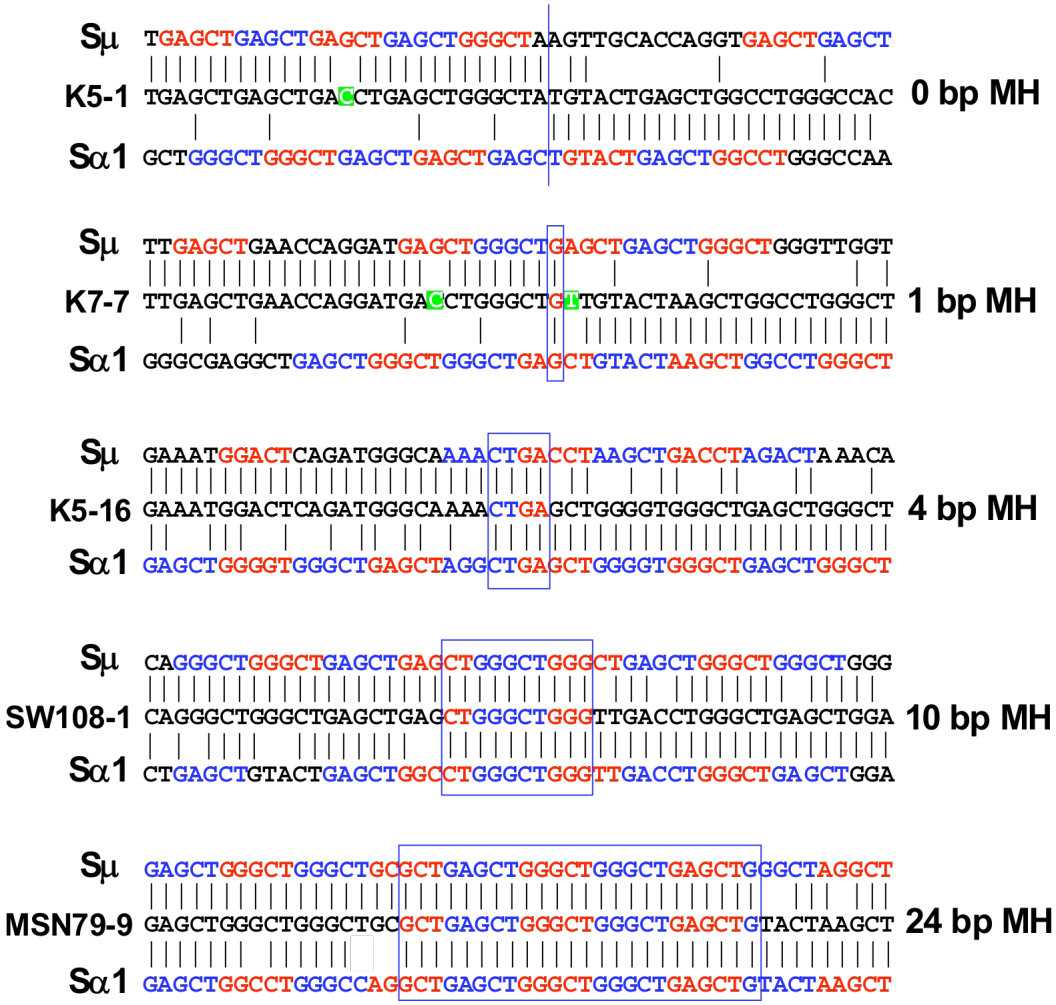
Control 3 (MSH5 L85F/P786S)
104 mutations

	To A	C	G	T	
A	-	0	4	1	4.8%
C	8	-	15	19	40.4%
G	22	20	-	9	49.0%
T	0	3	3	-	5.8%

Patient 3 (MSH5*)
32 mutations

	To A	C	G	T	
A	-	0	3	2	15.6%
C	1	-	2	6	28.1%
G	9	6	-	2	53.1%
T	1	0	0	-	3.1%

Supplementary Figure AI-13. Spectrum of somatic mutations in Sm-Sa1 joints.

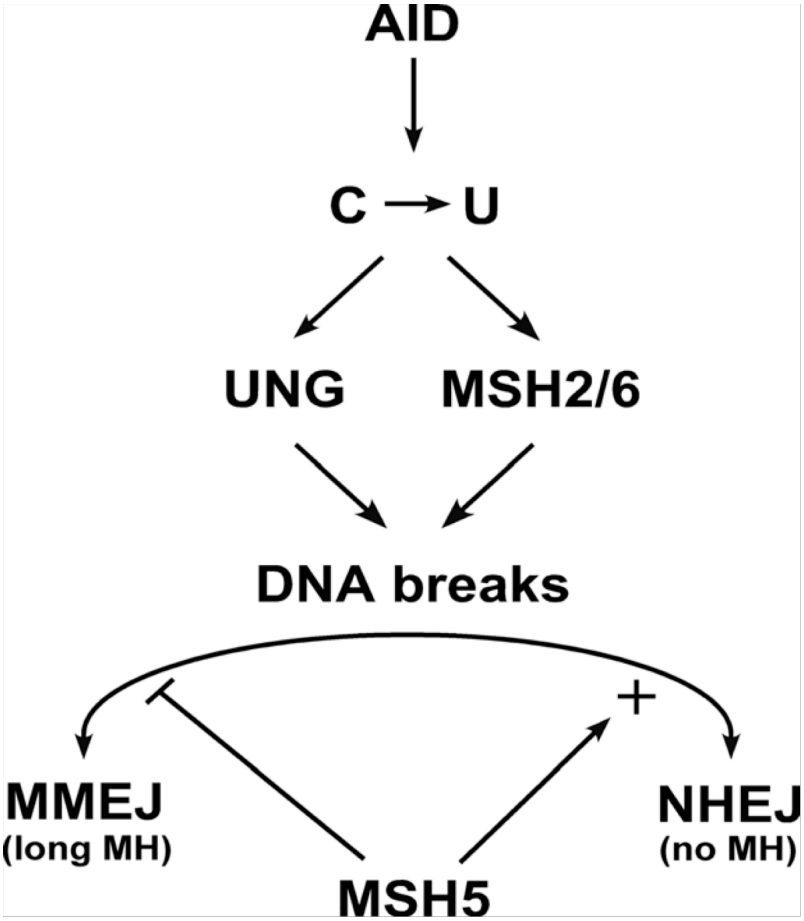


Supplementary Figure AI-14. Increased "in-frame" alignment of pentamer repeats at Ig Sm-Sa1 joints of CVID patients carrying associated alleles of MSH5. Shown are representative Sm-Sa1 joints amplified from peripheral blood DNA obtained from healthy controls (K) and CVID patients (MSN, SW). Each sequence is compared to germline Sm(above) and Sa1 (below) sequences. Pentamer repeats (XXGCT and XXXCT) are highlighted in alternating blue and red font. Microhomology (MH) is boxed. Mutations from germline sequence are highlighted in green, and bolded in the reference sequence. Note the "in-frame" nature of the pentamer repeats in the long microhomology joints.

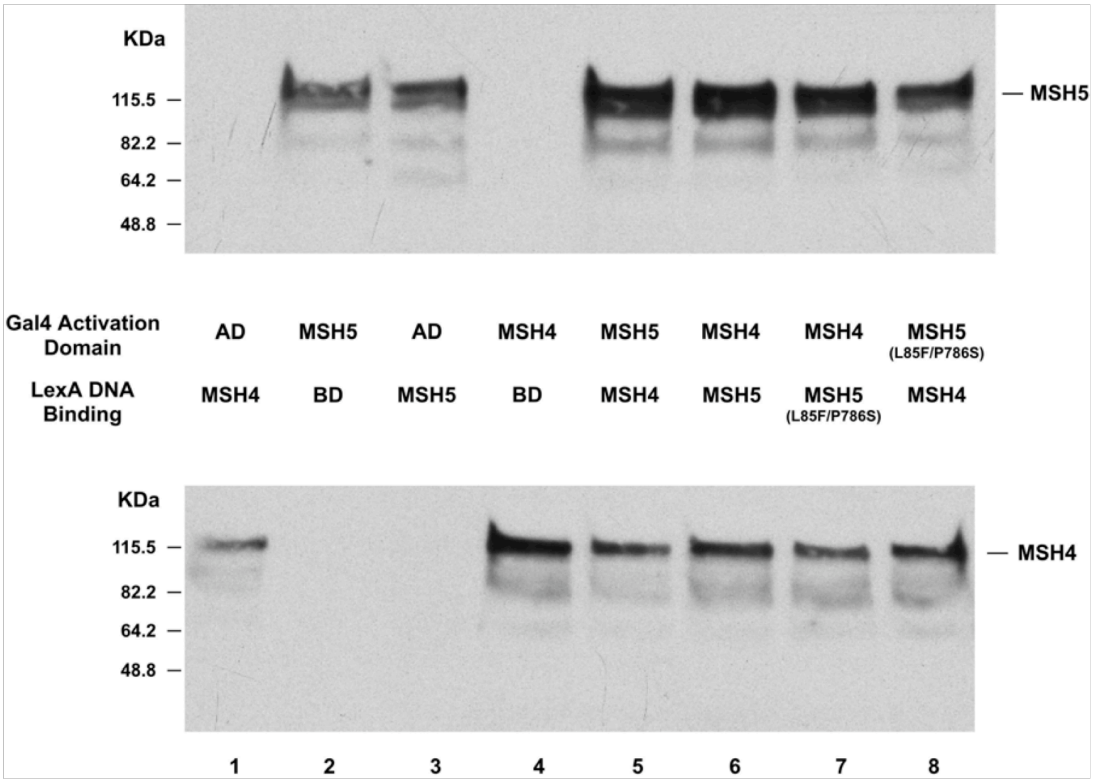
Groups	Sμ-Sα1 joints				Sμ-Sγ3 joints	
	Increased microhomology	Reduced mutation rate	Increased in-frame pentamers	Sμ breaks within pentamers	Increased microhomology	Reduced mutation rate
Control 1 (no associated *MSH5* alleles)	NO	NO	NO	NO	NO	NO
Control 2 (DR3++)	NO	YES	YES*	NO	-	-
Control 3 (*MSH5* L85F/P786S)	NO	NO	NO	YES	-	-
Patient 1 (*TACI**)	NO	NO	NO	NO	-	-
Patient 2 (DR3++)	YES	YES	YES	NO	NO	YES
Patient 3 (*MSH5**)	YES	YES	YES	YES	NO	NO

* *P* value trended towards, but did not reach statistical significance

Supplementary Figure AI-15. Summary of switch joint phenotypes in controls and CVID cases.



Supplementary Figure AI-16. Model for the potential role of MSH5 in class switch recombination. AID, activation induced cytidine deaminase; C, cytidine; U, uracil; UNG, Uracil-DNA glycosylase; MSH2/6, heterodimer of MSH2 and MSH6; MMEJ, microhomology mediated end-joining pathway; NHEJ, nonhomologous end joining DNA repair pathway; MH, microhomology. See Text for details.



Supplementary Figure AI-17. Western blot of MSH4 and MSH5 in transformed yeast cells. See SI Materials and Methods for details.

TABLES

Table AI-1.

Table 1. Association of MSH5 alleles with CVID and IgAD in Sweden and the U.S.

Cohort	*MSH5* alleles: L85F/P786S (exons 3, 24)	*MSH5* alleles: rs3131378 (intron 12)
	Allele frequency (*n*)*	
Sweden		
Controls ($N = 396$)†	1.8% (7)	11.9% (47)
IgAD ($N = 414$)	3.6% (15)	31.4% (128)
	$P = 0.104$‡	$P = 2.1 \times 10^{-1}$
CVID ($N = 166$)	2.4% (4)	15.7% (26)
		$P = 0.22$
U.S.		
Controls ($N = 976$)	3.2% (31)	9.7% (95)
IgAD ($N = 6$)	50% (3)	0 (0)
CVID ($N = 204$)	5.4% (11)	13.2% (27)
	$P = 0.12$	$P = 0.135$
	Pooled odds ratio (confidence interval, 95%)§	
Combined		
All IgAD	2.85 (1.24–6.51)	3.28 (2.28–4.72)
($N = 420$ cases, 1,372 controls)	$P = 5.8 \times 10^{-3}$	$P = 7.9 \times 10^{-11}$
All CVID	1.63 (0.88–3.02)	1.40 (1.00–1.98)
($N = 370$ cases, 1,372 controls)	$P = 0.058$	$P = 0.026$
All IgAD and CVID	2.04 (1.29–3.30)	2.15 (1.69–2.73)
($N = 790$ cases, 1,372 controls)	$P = 1.8 \times 10^{-3}$	$P = 2.6 \times 10^{-10}$

*N = total number of chromosomes genotyped in each group.
†n = number of positive alleles.
‡P values determined by using χ^2 tests.
§Pooled odds ratios and P values determined by using Mantel-Haenszel tests.

Supplementary Table AI-2.

Table S1. Differential expression of *Msh5*, *H-2* genes, and IgG3 between MRL/*lpr* and H-2^b congenic spleens

Probeset	Identifier	Gene	Average Affymetrix expression units: Control H-2$^{k/k}$ MRL/lpr	H-2$^{b/b}$ Congenic MRL/*lpr* IgG3^{pos}	H-2$^{b/b}$ Congenic MRL/*lpr* IgG3^{neg}	IgG3^{pos} Congenic/Ctrl FD	IgG3^{pos} Congenic/Ctrl *P**	IgG3^{neg} Congenic/Ctrl FD	IgG3^{neg} Congenic/Ctrl *P*
103719_at	AF109905	Msh5 (MutS homolog 5)	668	100	135	-6.68	0.009	-4.95	0.0082
102904_at	V00833	H-2, class II MHC, E alpha	1601	88	63	-18.19	0.0048	-25.41	0.0059
97174_r_at	M27134	H-2, class I MHC, K region locus 2	297	1380	1327	4.65	0.0171	4.47	0.0045
97048_at	X05315	H-2, class II MHC, E beta	78	396	418	5.08	0.0027	5.36	0.0017
99378_f_at	M18837	H-2, class I MHC, Q region	6825	42171	45193	6.18	0.0067	6.62	0.0005
102721_at	D14625	IgG3 mRNA	36385	20742	2837	-1.75	NS	-12.83	0.0001
102722_g_at	D14625	IgG3 mRNA	35166	20894	3099	-1.68	NS	-11.35	0.0004

FD, fold Δ from control H-2$^{k/k}$ MRL/*lpr*. NS, not significant.

*P values were determined using two-tailed Student's t tests comparing the various groups to control H-2$^{k/k}$ MRL/*lpr*.

Supplementary Table AI-3.

Table S2. Targeting of breakpoints to pentamer repeats[a] in the Sμ-Sγ3 and Sμ-Sα junctions from control H-2$^{k/k}$ MRL/*lpr*, IgG3^{pos} H-2$^{b/b}$ MRL/*lpr*, IgG3^{neg} H-2$^{b/b}$ MRL/*lpr*, WT FVB, Msh5$^{-/-}$ FVB, WT C57BL/6, and Msh4$^{-/-}$ C57BL/6 mice

	Sμ			Sγ3			Sμ			Sα		
	Yes[b] *n* (%)	No *n* (%)	*P*[c]	Yes *n* (%)	No *n* (%)	*P*	Yes *n* (%)	No *n* (%)	*P*	Yes *n* (%)	No *n* (%)	*P*
Control H-2$^{k/k}$ MRL/*lpr*	14 (74)	5 (26)		4 (21)	15 (79)		13 (48)	14 (52)		12 (44)	15 (56)	
IgG3^{pos} H-2$^{b/b}$ MRL/*lpr*	**12 (40)**	**18 (60)**	***0.039***	3 (10)	27 (90)	*0.407*	8 (40)	12 (60)	*0.767*	**3 (15)**	**17 (85)**	***0.056***
IgG3^{neg} H-2$^{b/b}$ MRL/*lpr*	18 (72)	7 (28)	*1.000*	2 (8)	23 (92)	*0.378*	17 (43)	23 (57)	*0.803*	14 (35)	26 (65)	*0.456*
WT FVB	20 (74)	7 (26)		4 (15)	23 (85)		12 (40)	18 (60)		10 (33)	20 (67)	
Msh5$^{-/-}$ FVB	21 (64)	12 (36)	*0.419*	6 (18)	27 (82)	*1.000*	17 (57)	13 (43)	*0.301*	12 (40)	18 (60)	*0.789*
WT C57BL/6	18 (55)	15 (45)		7 (21)	26 (79)		18 (60)	12 (40)		8 (27)	22 (73)	
Msh4$^{-/-}$ C57BL/6	26 (70)	11 (30)	*0.219*	12 (32)	25 (68)	*0.420*	24 (65)	13 (35)	*0.801*	11 (30)	26 (70)	*1.000*

[a] Pentamer repeats - GAGCT, GGGCT, and GGGGT.

[b] "Yes" is defined as a breakpoint occurring within a pentamer repeat.

[c] *P* values were determined using a two-tailed Fisher's exact test.

Supplementary Table AI-4.

Table S3. Targeting of breakpoints to AID hotspots[a] in the Sμ-Sγ3 and Sμ-Sα junctions from control H-2$^{k/k}$ MRL/*lpr*, IgG3^{pos} H-2$^{b/b}$ MRL/*lpr*, IgG3^{neg} H-2$^{b/b}$ MRL/*lpr*, WT FVB, Msh5$^{-/-}$ FVB, WT C57BL/6, and Msh4$^{-/-}$ C57BL/6 mice

	Sμ			Sγ3			Sμ			Sα		
	Yes[b] *n* (%)	No *n* (%)	*P*[c]	Yes *n* (%)	No *n* (%)	*P*	Yes *n* (%)	No *n* (%)	*P*	Yes *n* (%)	No *n* (%)	*P*
Control H-2$^{k/k}$ MRL/*lpr*	11 (58)	8 (42)		4 (21)	15 (79)		12 (44)	15 (56)		18 (67)	9 (33)	
IgG3^{pos} H-2$^{b/b}$ MRL/*lpr*	14 (47)	16 (53)	*0.561*	8 (27)	22 (73)	*0.743*	9 (45)	11 (55)	*1.000*	15 (75)	5 (25)	*0.748*
IgG3^{neg} H-2$^{b/b}$ MRL/*lpr*	13 (52)	12 (48)	*0.766*	7 (28)	18 (72)	*0.731*	20 (50)	20 (50)	*0.804*	25 (63)	15 (37)	*0.799*
WT FVB	17 (63)	10 (37)		15 (56)	12 (44)		15 (50)	15 (50)		19 (63)	11 (37)	
Msh5$^{-/-}$ FVB	20 (61)	13 (39)	*1.000*	16 (48)	17 (52)	*0.614*	18 (60)	12 (40)	*0.604*	19 (63)	11 (37)	*1.000*
WT C57BL/6	17 (52)	16 (48)		16 (48)	17 (52)		23 (77)	7 (23)		19 (63)	11 (37)	
Msh4$^{-/-}$ C57BL/6	27 (73)	10 (27)	*0.084*	24 (65)	13 (35)	*0.227*	25 (68)	12 (32)	*0.432*	24 (65)	13 (35)	*1.000*

[a] AID hotspots - Defined as RGYW/WRCY motifs (R=A/G; W=A/T; Y=C/T).

[b] "Yes" is defined as a breakpoint occurring within a pentamer repeat.

[c] *P* values were determined using a two-tailed Fisher's exact test.

Supplementary Table AI-5.

Table S4. Expression of *MSH5* in human B cells by quantitative Taqman PCR

	N	Mean GAPDH Ct ± SD	Mean MSH5 Ct ± SD	Δ Ct ± SD
CD19+ peripheral blood B cells	**3**	**21.6 ± 1.2**	**32.2 ± 1.5**	**10.6 ± 1.4**
Epstein-Barr transformed B cell lines	**6**	**17.3 ± 0.3**	**25.9 ± 0.4**	**8.6 ± 0.3**

Supplementary Table AI-6.

Table S5. Association of *MSH5* alleles with CVID and IgAD in Sweden and the U

Cohort	*MSH5* Alleles		
	Q292H (exon 11)	C580G (exon 19)	P29S (exon 2)
	Allele Frequency (n)*		
Sweden			
Controls (*N* = 396) †	0 (0)	0 (0)	12.1% (47)
IgAD (*N* = 414)	0 (0)	0.5% (2)	9.6% (34)
CVID (*N* = 166)	0.6% (1)	0 (0)	13% (20)
U.S.			
Controls (*N* = 976)	0 (0)	0.1% (1)	9.2% (33)
IgAD (*N* = 6)	0 (0)	0 (0)	16.7% (1)
CVID (*N* = 204)	0 (0)	0 (0)	6.9% (14)

* n = number of positive alleles.

† *N* = total number of chromosomes genotyped in each group.

‡ *P* values determined using $\chi 2$ tests.

Supplementary Table AI-7.

Table S6. Sμ-Sα1 switch junction microhomology in CVID patients and controls

		Switch Junction Donor/Acceptor Microhomology									
	Number of junctions	Average length (bp)	*P* [a, b]	Median length (bp)	*P* [c]	0-1 bp N (%)	*P* [d]	2-7 bp N (%)	*P* [d]	≥8 bp N (%)	*P* [d]
Control 1 [e]	**54**	**3.8**		**2**		**26 (48%)**		**18 (33%)**		**10 (18%)**	
Control 2 (DR3++)	**35**	**3.8**	**-**	**3**	**-**	**13 (37%)**	**-**	**15 (43%)**	**-**	**7 (20%)**	**-**
Control 3 (*MSH5* L85F/P786S)	**44**	**4.3**	**-**	**3**	**-**	**15 (34%)**	**-**	**16 (36%)**	**-**	**13 (30%)**	**-**
Patient 1 (*TACI)**	**41**	**4.1**	**-**	**3**	**-**	**14 (34%)**	**-**	**18 (44%)**	**-**	**9 (22%)**	**-**
Patient 2 (DR3++)	**15**	**10.3**	***4.7 x 10^{-3}***	**8**	***3 x 10^{-4}***	**1 (7%)**	***5.6 x 10^{-3}***	**6 (40%)**	**-**	**8 (53%)**	***0.02***
Patient 3 (*MSH5)**	**55**	**9.3**	***9.3 x 10^{-7}***	**9**	***1.9 x 10^{-7}***	**4 (7%)**	***1.3 x 10^{-6}***	**18 (33%)**	**-**	**33 (60%)**	***2.6 x 10^{-5}***

[a] *P* values were calculated comparing the various groups to Control group 1. For all comparisons nonsignificant *P* values (*P* >0.05) are designated by "-".

[b] *P* values were calculated using two-tailed t-tests.

[c] *P* values were calculated using two-tailed Mann-Whitney tests.

[d] *P* values were calculated using two-tailed Fisher's exact tests.

[e] Control 1 - no associated *MSH5* alleles or *TACI* mutations (54 joints from 8 donors); Control 2 - homozygous for DR3 extended haplotype (35 joints from 6 donors); Control 3 - heterozygous for the L85F/P786S allele (44 joints from 5 donors); Patient 1 - *TACI* mutations (of the 41 total joints, 14 joints from a A181E heterozygote; 11 joints from a C104R/204insA compound heterozygote, 7 joints from a C104R/S194X compound heterozygote, and 9 joints from a C104R/P251L compound heterozygote); Patient 2 - homozygous for DR3 extended haplotype (15 joints from 2 cases); Patient 3 - carrying associated *MSH5* alleles (of the 55 total joints, 46 joints from L85F/P786S heterozygotes, 7 from L85F/P786S / DR3+ compound heterozygotes, and 2 from a Q292H heterozygote)

Supplementary Table AI-8.

Table S7. Mutation and breakpoint analysis of Sμ-Sα1 switch junctions in CVID patients and controls

	Junctions, no	Total mutations/ total bp sequenced, no	Overall mutation frequency per 1,000 bp	*P*[†]	Mutations at junctions, N[‡]	Junctional mutation frequency per 1,000 bp	*P*	Mutations targeted to dG/dC, %[§]	Transitions in dG:dC , %[§]	Sμ breaks within pentamer motifs[¶], %	*P*
Control 1 [‖]	54	97 / 19,564	5.0		32	19.8		90	51	50	
Control 2 (DR3++)	35	49 / 15,158	3.2	0.01	14	13.3	-	90	64	69	-
Control 3 (*MSH5* L85F/P786S)	44	104 / 19,268	5.4	-	26	19.7	-	89	44	75	0.013
Patient 1 (*TACI**)	41	65 / 17,867	3.6	-	20	16.3	-	92	52	56	-
Patient 2 (DR3++)	15	2 / 6,492	0.3	1.3×10^{-7}	0	0	2.7×10^{-3}	100	100	67	-
Patient 3 (*MSH5**)	55	32 / 24,436	1.3	2.0×10^{-12}	9	5.5	2.4×10^{-4}	81	58	82	5.4×10^{-3}

Dashes indicate nonsignificant *P* values.

[†] *P* values calculated by χ^2 tests comparing the various groups to Control group 1.

[‡] Number of mutations occurring in the 15 bp region spanning either side of the breakpoint.

[§] No statistical differences between the groups using two-tailed Fisher's exact tests.

[¶] Pentamer motifs - GAGCT, GGGCT, and GGGGT.

[‖] See SI Table 6 for group definitions.

Supplementary Table AI-9.

Table S8. Sμ-Sα switch junction microhomology in Swedish IgAD patients and controls

		Switch Junction Donor/Acceptor Microhomology									
	Number of junctions	Average length (bp)	*P* [a, b]	Median length (bp)	*P* [c]	0-1 bp N (%)	*P* [b]	2-7 bp N (%)	*P* [b]	≥8 bp N (%)	*P* [b]
Control 1 [d]	**154**	**1.8**		**1**		**103 (67%)**		**41 (27%)**		**10 (7%)**	
Patient 2 (DR3++)	**11**	**5.2**	***1.6 x 10^{-3}***	**5**	***4 x 10^{-6}***	**0 (0%)**	***1.2 x 10^{-5}***	**10 (91%)**	***3.4 x 10^{-5}***	**1 (9%)**	**-**
Patient 3 (*MSH5)**	**13**	**1.4**	**-**	**3**	***0.01***	**4 (31%)**	***0.01***	**6 (46.2%)**	**-**	**3 (23%)**	**-**

[a] For all comparisons nonsignificant *P* values (*P* >0.05) are designated by "-".

[b] *P* values were calculated using two-tailed Fisher's exact tests.

[c] *P* values were calculated using two-tailed Mann-Whitney tests.

[d] See SI Table 6 for group definitions

Supplementary Table AI-10.

Table S9. Sμ-Sγ3 junction microhomology in CVID/IgAD patients and controls

		Switch Junction Donor/Acceptor Microhomology									
	Number of junctions	Average length (bp)	*P* [a, b]	Median length (bp)	*P* [c]	0-1 bp N (%)	*P* [b]	2-7 bp N (%)	*P* [b]	≥8 bp N (%)	*P* [b]
Control 1 [d]	**24**	**1.9**		**1.5**		**12 (50%)**		**50 (50%)**		**0 (0%)**	
Patient 2 (DR3++)	**19**	**1.5**	**-**	**1**	**-**	**10 (53%)**	**-**	**9 (47%)**	**-**	**0 (0%)**	**-**
Patient 3 (*MSH5)**	**18**	**1.4**	**-**	**1**	**-**	**12 (67%)**	**-**	**6 (33%)**	**-**	**0 (0%)**	**-**

[a] For all comparisons nonsignificant *P* values (*P* >0.05) are designated by "-".

[b] *P* values were calculated using two-tailed Fisher's exact tests.

[c] *P* values were calculated using two-tailed Mann-Whitney tests.

[d] Control 1 - no associated MSH5 alleles or TACI mutations (24 joints from 2 donors); Patient 2 - homozygous for DR3 extended haplotype (19 joints from 1 CVID and 3 IgAD patients); Patient 3 - carrying associated MSH5 alleles (of the 18 total joints, 17 joints from L85F/P786S heterozygotes, and 1 from L85F/P786S / DR3+ compound heterozygote).

Supplementary Table AI-11.

Table S10. Mutation and breakpoint analysis of Sμ-Sγ3 switch junctions in CVID/IgAD patients and controls

	Number of junctions	Total mutations / Total number of bp sequenced	Overall mutation frequency (per 1,000 bp)	P [a]	Mutations at junctions [b] N	Junctional mutation frequency (per 1,000 bp)	P	Mutations targeted to dG/dC [c] %	Transitions in dG:dC %	P	Sμ breaks within pentamer motifs [d] %	P
Control 1 [e]	24	108 / 13,996	7.7		19	26.4		80	39		58	
Patient 2 (DR3++)	19	48 / 10,666	4.6	1.6×10^{-3}	8	14	-	98	65	2.5×10^{-3}	47	-
Patient 3 (*MSH5**)	18	57 / 8440	6.8	-	10	18.5	-	81	42	-	33	-

[a] *P* values calculated by χ^2 square tests comparing the various groups to Control group 1

[b] Number of mutations occurring in the 15 bp region spanning either side of the breakpoint

[c] No statistical differences between the groups using two-tailed Fisher's exact tests

[d] Pentamer motifs - GAGCT, GGGCT, and GGGGT

[e] See SI Table 6 for group definitions

Supplementary Table AI-12.

Table S11. Targeting of CVID and control switch junctions to pentamer motifs and AID hotspots

	Number of breaks	Sα1 [a, c]		Sμ [b,c]		Sα1 [c]	
		Breaks within pentamer motifs N (%)	Breaks outside pentamer motifs N (%)	Breaks within AID hotspots N (%)	Breaks outside AID hotspots N (%)	Breaks within AID hotspots N (%)	Breaks outside AID hotspots N (%)
Control 1 [d]	54	31 (57%)	23 (43%)	36 (67%)	18 (33%)	37 (68%)	17 (31%)
Control 2 (DR3++)	35	22 (63%)	13 (37%)	27 (77%)	8 (23%)	25 (71%)	10 (29%)
Control 3 (*MSH5* L85F/P786S)	44	25 (57%)	19 (43%)	36 (82%)	8 (18%)	34 (77%)	10 (23%)
Patient 1 (*TACI**)	41	22 (54%)	19 (46%)	30 (73%)	11 (27%)	27 (66%)	14 (34%)
Patient 2 (DR3++)	15	10 (67%)	5 (33%)	12 (80%)	3 (20%)	13 (87%)	2 (13%)
Patient 3 (*MSH5**)	55	26 (47%)	29 (56%)	43 (78%)	12 (22%)	30 (55%)	25 (45%)

[a] Pentamer motifs - GAGCT, GGGCT, and GGGGT

[b] AID hotspots - Defined as RGYW/WRCY motifs (R=A/G; W=A/T; Y=C/T)

[c] No statistically significant differences between the various groups

[d] See SI Table 6 for group definitions

Supplementary Table AI-13.

Table S12. Summary of primers and probes

MSH5 sequencing	Primer ID		Sequence
	GE97096	for	5'-ATGTTCACCGCTTTGAGTCC-3'
		rev	5'-CCAGCCTAGAGATCCGACAG-3'
	GE97097	for	5'-GGCTGAGACAGAAGTCCTGC-3'
		rev	5'-TTACCCTTTGGAAGGAAGGG-3'
	GE97098	for	5'-GAAGGCAAATAGGCCAATCA-3'
		rev	5'-CGCAGAATGCAAAGCACTAA-3'
	GE97099	for	5'-CGTGGAGTTTCCCACAATCT-3'
		rev	5'-AGCTCTTCCCAACCCTGTTT-3'
	GE97100	for	5'-AAAACAGGGTTGGGAAGAGC-3'
		rev	5'-TGGTGAAACACCATCTGTGC-3'
	GE97101	for	5'-CTCACCATCCCACCACTCTT-3'
		rev	5'-GACACCGGAACAGGGAGTTA-3'
	GE97102	for	5'-GAGGAGGGCTATGGGTTTTC-3'
		rev	5'-GTGCCTTTTGGTTTGGTGAT-3'
	GE97103	for	5'-TCTCAGGCTGTGCTTTCAGA-3'
		rev	5'-CCACATTACTTTGGTTGGGG-3'
	GE97104	for	5'-TACAAGACCGTTCCCTTTGC-3'
		rev	5'-AGCATGCCTCCACCTCTTTA-3'
	GE97107	for	5'-CCTGTGAGTGTCCATCCCTT-3'
		rev	5'-TGCTGCTTGTATGTGTGCAA-3'
	GE97108	for	5'-CCTCAGAGTGAGCTGCAGTG-3'
		rev	5'-AATACAGGCTGAGGGCCTTT-3'
	GE97110	for	5'-TGATCTACCCACCTCAGCCT-3'
		rev	5'-CTGTGAAGCGATTTTCAGCA-3'
	GE97111	for	5'-GCAGTCCATCCAGCTCTTTC-3'
		rev	5'-CTCCCCCAAAAGACATCTGA-3'
	GE97112	for	5'-TAAAGAATGGGATGGTGGGA-3'
		rev	5'-AGGCGGGGAATAGAAAGAAG-3'
	GE97113	for	5'-CAGATGTCTTTTGGGGGAGA-3'
		rev	5'-GGTACAGGGGTTCAAAGCTG-3'
	GE97114	for	5'-CACTGCCCAATATGGGATCT-3'
		rev	5'-TGGTACATCAGCAGCGTCTC-3'
	GE97115	for	5'-TTCCTGCTCTCTGTCTCGCT-3'
		rev	5'-CCTGATGAGTTGGGTCCAGT-3'
	GE97116	for	5'-GAGACGCTGCTGATGTACCA-3'
		rev	5'-ACTCAGAGGCAGGTGAAGGA-3'
	GE97117	for	5'-CACAGGGCTATGGTCAGGAT-3'
		rev	5'-CCTGCTGGGTATAAGGTGGA-3'
	GE97118	for	5'-AGTGGGGAGCACTAAGGTCA-3'
		rev	5'-ATAGAGGAGCCAATCCCCAC-3'
	GE97119	for	5'-TTTGGAAAGGGAACCAACAC-3'
		rev	5'-AAGATCGTTGCCATCCTCAC-3'
	GE97120	for	5'-TGGTGCAGTATTTGGTGAGG-3'
		rev	5'-TTTCTGAAGGGCAAAGTGCT-3'
	GE97121	for	5'-ATGGAGACCTGTGAGGATGG-3'
		rev	5'-GGACTCTCAGAGGATGCTGG-3'
	GE97122	for	5'-TTCTCCCCTTTTCAGGGACT-3'
		rev	5'-AGGCTGGGGTGTTCCTACTT-3'
	Promoter	for	5'-TGTGTAGAAGTGCGAAGGGGTAT-3'
		rev	5'-GGAAGCCCGCGAGAAGGTC-3'
	Exon 1	for	5'-ATCCGGGCCTCTCCAATCAACAGC-3'
		rev	5'-AAAAAGTGAGGGCGGTTCGGAAGC-3'
	Exon 7-8	for	5'-GCCCCCAGGAGATTTAAGATTTAC-3'
		rev	5'-ACTAGAGCCTCATTTAGGATTATTG-3'
	Exon 11-12	for	5'-CCGTTGAAACTGTGATATGTAGAC-3'
		rev	5'-GAGGGTAGAATAAAATTGGACAGC-3'
	Exon 13	for	5'-GAGATCTTCCCTCTTTGTTACTGT-3'
		rev	5'-AACTCTCACCCTCCAGATACTTCC-3'

Taqman *MSH5* SNP Genotyping	Primers/Probes		Sequence
	P29S		ABI Assays-On-Demand Cat# C_15863099_10
	L85F	for	5'-CCACTATCCACTTCATGCCAGATG-3'
		rev	5'-CATCCCCACCTCTCTGGAGAA-3'
		VIC probe (T allele)	5'-ACGAGAGC**T**TCAAGC-3'
		FAM probe (C allele)	5'-ACGAGAGC**C**TCAAGC-3'
	DR3 (rs3131378)	for	5'-CCGATGAGGTTTTTTTGTCATTGTTC-3'
		rev	5'-TGTACAGAGTCGGGAGAAAGAGAAA-3'
		VIC probe (T allele)	5'-TCATGGCCTTTT**A**TAGCTA-3'
		FAM probe (C allele)	5'--CATGGCCTTTT**G**TAGCTA-3'
	P786S		ABI Assays-On-Demand Cat# C_7482280_10
Site-directed Mutagenesis	L85F		5'-CCTCTCTGGAGAAGCTTGAAGCTCTCGTGGTCTGGGGC-3'
	P786S		5'-GCAAATCCTTGACAGACTTGATGGGTTTTCCACTGCGGATC-3
Sμ-Sα1 nested PCR	Sμ-1		5'-GACCATGGGGACCTGCTCATTTTTATC-3'
	Sα-common		5'-ACGTCGACGCCCTCAGAACCCCTAAGAA-3'
	Sμ-2		5'-ACGCATGCGGCAATGAGATGGCTTTAG-3'
	Sα1-specific		5'-ACGTCGACCAGTCCAGCCCAAGTCATC-3'

SUPPLEMENTAL MATERIALS AND METHODS

Ig ELISA and ELISpot assays

Serum Ig isotype levels were measured by ELISA as described in (Sekine et al. 2006). Briefly, 96-well polystyrene ELISA plates (Dynatech Laboratories, Chantilly, VA) were coated overnight at 4°C with goat anti-mouse Ig (M+G+A) (Southern Biotechnology, Birmingham, AL) or goat anti-mouse IgG (Southern Biotechnology). Serum dilutions were then added and bound antibodies were detected with HRP-conjugated anti-mouse IgM, IgG1, IgG2a, IgG2b, IgG3, or IgA (Southern Biotechnology). Purified mouse Ig isotypes (Southern Biotechnology) were used as standards. Absorption at OD_{380} was determined. Standard curves and serum antibody concentrations were calculated using DeltaSoft 3 (v2.2; BioMetallics, Princeton, NJ), and statistical analysis was performed with Prism software (v3.0; GraphPad Software). The number of IgM, IgG1, IgG2a, IgG2b, IgG3, or IgA secreting cells in the spleen was assessed by ELISpot assays (Sekine et al. 2004). 96-well multiscreen plates (Millipore, Billerica, MA) were coated overnight with goat anti-mouse Ig (M+G+A). The plates were incubated in triplicate with spleen cells in serial dilutions starting at 1 x 10^6 cells per well. Then, the plates were incubated with 100 ml per well of biotinylated goat anti-mouse IgM, IgG1, IgG2a, IgG2b, IgG3, or IgA (Southern Biotechnology) followed by 50 ml per well of streptavidin-HRP (Jackson ImmunoResearch Laboratories, West Grove, PA). After washing, the plates were developed with 50 ml per well of a substrate solution containing 0.3 mg/ml of 3-amino-9-ethylcarbazole (Sigma-Aldrich, St. Louis, MO) in 0.05 M sodium acetate, pH 5.0, 0.015% H_2O_2. The number of spots was counted manually by using a dissecting microscope.

In vitro and in vivo class switch assays

For the in vitro experiments, splenic B cells from 8-12-wk-old MRL/lpr, $Msh5^{-/-}$ FVB and $Msh4^{-/-}$ C57BL/6 mice were purified (>95% CD19+) by negative selection using a magnetic

bead cell separation system (Miltenyi Biotec, Auburn, CA). B cells were cultured at 1 x 10^6 cells/ml in RPMI 1640 supplemented with 15% FCS, 5 x 10^{-5} M 2-ME and 50 mg/ml of LPS (Salmonella typhimurium; Sigma) with or without 50 ng/ml rIL-4 (R&D Systems, Minneapolis, MN), for IgG3 (LPS alone) or IgG1 (LPS + IL-4) class switch induction. For IgA class switch induction, B cells were cultured at the same condition in addition with 20 ng/ml rIL-5 (R&D Systems), 2 ng/ml human rTGF-b1 (R&D Systems), and 3 ng/ml anti-d-dextran (gift from C. Snapper, Uniformed Services University of Health Science, Bethesda, MD). After 4 days of culture, flow cytometric analysis of surface IgG3, IgG1, or IgA expression was performed by using FITC-conjugated anti-mouse IgG3, IgG1, or IgA and PE-conjugated anti-mouse CD19 (all from BD PharMingen). For in vivo experiments, 8- to 12-wk-old MRL/lpr mice were immunized with 0.1 mg of TNP conjugated with LPS (Sigma) or Ficoll (BTI). Two weeks after immunization, mice were bled and tested for titers of serum IgM, IgG2b, and IgG3 anti-TNP antibodies by ELISA.

Affymetrix Gene Arrays and Real Time PCR

RNA was isolated from whole spleen of groups of 3 MRL/lpr H-2b/b and H-2k/k congenic mice (backcross nine generations) at 8 weeks of age. Gene array expression analysis was performed as previously reported (Liu et al. 2006). One additional transcript encoded on the X chromosome, Xlr3b, also showed »5-fold reduced expression in the F9 H-2b/b congenic spleens (data not shown). Xlr3b is maternally imprinted, and contributes to cognitive function in mice (Davies et al. 2005); thus, it was not further pursued as a candidate for the antibody phenotype. Taqman quantitative real-time PCR for Msh5 was performed using Assay-On-Demand kits from Applied Biosystems, Inc. cDNA was generated from B cell RNA using a kit from Invitrogen and real-time RT-PCR was performed as previously reported (Liu et al. 2006). MSH5 mRNA expression was examined in human B cells using commercially available

Taqman assays (ABI, Assay on Demand). RNA was extracted from CD19+ B cells purified from peripheral blood of 3 healthy human donors using magnetic beads (Miltenyi Biotec), or EBV transformed cell lines, and converted to cDNA (Invitrogen, Carlsbad, CA).

Yeast two-hybrid analysis

The L85F/P786S double nonsynonymous mutant was generated from the full-length human MSH5 cDNA using standard site-directed mutagenesis (Quickchange; Strategene, La Jolla, CA), and was confirmed by sequencing. cDNA fragments encoding full-length MSH4 and MSH5 as well as MSH5 L85F/P786S were cloned into pBTMd (or pBTM116) and pACT2 vectors. Relevant pairs of two-hybrid constructs were transformed into the L40 reporter strain and protein interactions were ascertained using a liquid b-galactosidase (b-gal) assay, as described in ref. 5. Briefly, yeast transformation was performed using a standard polyethylene glycol-lithium acetate method, and transformants were cultured in SD/-Leu-Trp medium overnight and refreshed in YPD medium at 30°C until the OD_{600} reached 0.5-0.8. An aliquot of 1.5 ml of each culture was then centrifuged at 12,000 x g for 30 s, washed once with buffer, and resuspended in 0.3 ml of buffer. To measure the b-gal activity, the cells derived from 0.1 ml of the suspension were disrupted with 0.1 g of glass beads using a vortex mixer, added to 0.7 ml of buffer containing 1.89 ml of 2-mercaptoethanol, combined with 160 ml of 4 mg/ml O-nitrophenyl-1-thio-b-D-galactopyranoside in buffer, followed by incubation at 30° C. Reactions were stopped with the addition of 400 ml of 1 M Na_2CO_3. The reaction tubes were centrifuged at 12,000 x g for 5 min, and the supernatants were collected to obtain OD_{420} readings. The b-gal activity was calculated using the following formula: b-gal units = 1,000 ´ A420 nm/(t ´ V ´ A600 nm); where t = time of reaction (min), and V = volume of culture used in the assay (ml).

For Western blot analysis, yeast double transformants were cultured individually overnight at 30°C with shaking (230 rpm) in 7.5 ml of SD/-Leu-Trp medium. Following a brief vortexing, cells were transferred into 50 ml YPD medium for continued culture until the OD_{600} reached 0.4-0.6. Cells were then lysed by freezing in liquid nitrogen, rapidly warmed in cracking buffer (8 M urea/5% SDS/40 mM Tris×HCl, pH6.8/0.1 mM EDTA/0.4 mg/ml bromophenol blue/1% b-mercaptoethanol) containing a protease inhibitor mixture, and vortexed vigorously by using glass beads (Sigma). Protein-containing supernatants were separated on 4-20% SDS/PAGE, and proteins were transferred to nitrocellulose membranes (ISC BioExpress, Kaysville, UT) for immunoblot analysis performed with affinity-purified a-hMSH4 and a-hMSH5 antibodies (Yi et al. 2005). Western blot signals were visualized with the ECL Western blotting system (Pierce, Rockford, IL).

Analysis of Sm-Sg3 or Sm-Sa recombination junctions in mice

Genomic DNA was prepared from splenic cells after proteinase K digestion. The amplification of either Sm-Sg3 or Sm-Sa fragments from in vivo switched cells was performed according to the previously described method with minor modifications by using specific primers and Expand Long Template Taq and Pfu polymerase mix (Roche, Basel, Switzerland) for the nested PCR (Pan-Hammarström et al. 2005). The sequences of the primers for the first (Sm1 and either Sg3.1 or Sa1) and second round (Sm2 and either Sg3.2 or Sa2) of the nested PCR have been described in (Ehrenstein and Neuberger 1999). The number of Sm-Sg3 or Sm-Sa fragments was determined from sixteen PCR reactions run in parallel using DNA (30 ng per reaction) from one individual and represents random amplification of in vivo switched clones. One or two Sm-Sg3 or Sm-Sa fragments were normally amplified in one lane, when 30 ng of DNA was added in one reaction. Because of the small number of the Sm-Sg3 switched clones in IgG3neg H-2b/b MRL/lpr mice, additional PCR reactions were run. The PCR-amplified Sm-

Sg3 or Sm-Sa fragments were gel purified and directly sequenced by using primers either Sm2, Sg3.2 or Sa2. Switch junctions were identified by using the BLAST algorithm (NCBI) with the low-complexity filter disabled. Donor acceptor microhomology was calculated as the number of nucleotides of perfect identity between donor and acceptor sequences at the switch junction.

PCR amplification and characterization of human Sm-Sa1 junctions

Amplification of human Sm-Sa1 junctions from peripheral blood DNA was performed as described (Pan et al. 2001; Pan et al. 1997). Briefly, two pairs of Sm and Sa1 specific primers (Sm1/Sa-common and Sm2/Sa1-specific) were used in a nested PCR design. In the first round of PCR, 30-50 ng of genomic DNA were added to a 25-ml reaction and amplified for 35 cycles (consisting of 1 min at 94°C, 1 min at 60°C and 1min 30 sec at 72°C), following an initial denaturation step of 3 min at 94°C. One microliter of the first round amplification product was used as template for the second round of PCR, and further amplified for 35 cycles (1 min at 94°C, 1 min at 65°C and 1 min at 72°C). After the PCR reaction, samples were loaded in a 1% agarose gel and DNA bands corresponding to individual Sm-Sa1 joints were gel purified (Qiagen, Valencia, CA). Purified samples were then directly sequenced using an automated fluorescence sequencer. Starting DNA concentration was optimized to increase the probability of single switch joints amplification in each PCR reaction. Sequencing traces were then individually inspected to rule out the possibility of background noise resulting from contamination with multiple switch junction sequences. A similar nested PCR assay was used to amplify the Sm-Sg 3 fragments, as described in (Pan et al. 1997).

Sequence analysis was performed using the Blast application (NCBI) with the low complexity filter disabled. Breakpoint location was determined by aligning the switch joint sequences with Sm (X54713) and Sa1 (L19121) or Sm and Sg3 (U39935) germline sequences, and microhomology was defined as successive nucleotides shared by both the Sm and Sa1 or

Sm and Sg3 sequences at the switch region. Insertions were defined as nucleotides at the break that were not identical to either of the switch regions. We identified several common genomic polymorphisms in the Sm and Sa1 sequences (not found in the X54713 and L19121 reference sequences), and these were ignored for the analysis. We also examined the switch joints for mutation frequency, nature of the mutations (transitions, transversions), targeting of the breaks to pentamers and other AID hotspots, and for alignment of pentamer motifs at the breaks.

Statistical analyses

Differences between groups were determined using either Student's t test for parametric measures or the Mann-Whitney test for nonparametric measures. For categorized microhomology, mutation and breakpoint analyses, we used a two-tailed Fisher's exact probability test. For the overall mutation rate analysis, where the numbers in each group were higher, the Pearson c2 test was used. A Mantel-Haenszel test was performed to estimate genetic association of the typed markers in the combined Swedish/U.S. analyses. For all statistical tests, $P < 0.05$ was considered significant.

REFERENCES

Alper C. A., D. Marcus-Bagley, Z. Awdeh, M. S. Kruskall, G. S. Eisenbarth, S. J. Brink, A. J. Katz, R. Stein, D. H. Bing, E. J. Yunis and P. H. Schur (2000). Prospective analysis suggests susceptibility genes for deficiencies of IgA and several other immunoglobulins on the [HLA-B8, SC01, DR3] conserved extended haplotype. Tissue Antigens **56**(3): 207-216.

Briere F., J. M. Bridon, D. Chevet, G. Souillet, F. Bienvenu, C. Guret, H. Martinez-Valdez and J. Banchereau (1994). Interleukin 10 induces B lymphocytes from IgA-deficient patients to secrete IgA. J Clin Invest **94**(1): 97-104.

Burrows P. D. and M. D. Cooper (1997). IgA deficiency. Adv Immunol **65**: 245-276.

Castigli E., S. A. Wilson, L. Garibyan, R. Rachid, F. Bonilla, L. Schneider and R. S. Geha (2005). TACI is mutant in common variable immunodeficiency and IgA deficiency. Nat Genet **37**(8): 829-834.

Cunningham-Rundles C. and C. Bodian (1999). Common variable immunodeficiency: clinical and immunological features of 248 patients. Clin Immunol **92**(1): 34-48.

Davies W., A. Isles, R. Smith, D. Karunadasa, D. Burrmann, T. Humby, O. Ojarikre, C. Biggin, D. Skuse, P. Burgoyne and L. Wilkinson (2005). Xlr3b is a new imprinted candidate for X-linked parent-of-origin effects on cognitive function in mice. Nat Genet **37**(6): 625-629.

de Vries S. S., E. B. Baart, M. Dekker, A. Siezen, D. G. de Rooij, P. de Boer and H. te Riele (1999). Mouse MutS-like protein Msh5 is required for proper chromosome synapsis in male and female meiosis. Genes Dev **13**(5): 523-531.

Dunnick W., G. Z. Hertz, L. Scappino and C. Gritzmacher (1993). DNA sequences at immunoglobulin switch region recombination sites. Nucleic Acids Res **21**(3): 365-372.

Edelmann W., P. E. Cohen, B. Kneitz, N. Winand, M. Lia, J. Heyer, R. Kolodner, J. W. Pollard and R. Kucherlapati (1999). Mammalian MutS homologue 5 is required for chromosome pairing in meiosis. Nat Genet **21**(1): 123-127.

Ehrenstein M. R. and M. S. Neuberger (1999). Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation. EMBO J **18**(12): 3484-3490.

Ehrenstein M. R., C. Rada, A. M. Jones, C. Milstein and M. S. Neuberger (2001). Switch junction sequences in PMS2-deficient mice reveal a microhomology-mediated mechanism of Ig class switch recombination. Proc Natl Acad Sci U S A **98**(25): 14553-14558.

Grimbacher B., A. Hutloff, M. Schlesier, E. Glocker, K. Warnatz, R. Drager, H. Eibel, B. Fischer, A. A. Schaffer, H. W. Mages, R. A. Kroczek and H. H. Peter (2003). Homozygous loss of ICOS is associated with adult-onset common variable immunodeficiency. Nat Immunol **4**(3): 261-268.

Hammarström L. and C. I. Smith (1983). HLA-A, B, C and DR antigens in immunoglobulin A deficiency. Tissue Antigens **21**(1): 75-79.

Hughes A. L. and M. Yeager (1998). Natural selection at major histocompatibility complex loci of vertebrates. Annu Rev Genet **32**: 415-435.

Klein U., Y. Tu, G. A. Stolovitzky, J. L. Keller, J. Haddad, Jr., V. Miljkovic, G. Cattoretti, A. Califano and R. Dalla-Favera (2003). Transcriptional analysis of the B cell germinal center reaction. Proc Natl Acad Sci U S A **100**(5): 2639-2644.

Kneitz B., P. E. Cohen, E. Avdievich, L. Zhu, M. F. Kane, H. Hou, Jr., R. D. Kolodner, R. Kucherlapati, J. W. Pollard and W. Edelmann (2000). MutS homolog 4 localization to

meiotic chromosomes is required for chromosome pairing during meiosis in male and female mice. Genes Dev **14**(9): 1085-1097.

Li Z., S. J. Scherer, D. Ronai, M. D. Iglesias-Ussel, J. U. Peled, P. D. Bardwell, M. Zhuang, K. Lee, A. Martin, W. Edelmann and M. D. Scharff (2004). Examination of Msh6- and Msh3-deficient mice in class switching reveals overlapping and distinct roles of MutS homologues in antibody diversification. J Exp Med **200**(1): 47-59.

Liu J., G. Karypis, K. L. Hippen, A. L. Vegoe, P. Ruiz, G. S. Gilkeson and T. W. Behrens (2006). Genomic view of systemic autoimmunity in MRLlpr mice. Genes Immun **7**(2): 156-168.

Mathis D. J., C. Benoist, V. E. Williams, 2nd, M. Kanter and H. O. McDevitt (1983). Several mechanisms can account for defective E alpha gene expression in different mouse haplotypes. Proc Natl Acad Sci U S A **80**(1): 273-277.

Min I. M. and E. Selsing (2005). Antibody class switch recombination: roles for switch sequences and mismatch repair proteins. Adv Immunol **87**: 297-328.

Mitchell M. K., P. K. Gregersen, S. Johnson, R. Parsons and D. Vlahov (2004). The New York Cancer Project: rationale, organization, design, and baseline characteristics. J Urban Health **81**(2): 301-310.

Olerup O., C. I. Smith and L. Hammarström (1990). Different amino acids at position 57 of the HLA-DQ beta chain associated with susceptibility and resistance to IgA deficiency. Nature **347**(6290): 289-290.

Pan Q., C. Petit-Frere, S. Dai, P. Huang, H. C. Morton, P. Brandtzaeg and L. Hammarström (2001). Regulation of switching and production of IgA in human B cells in donors with duplicated alpha1 genes. Eur J Immunol **31**(12): 3622-3630.

Pan Q., C. Petit-Frere, A. Lahdesmaki, H. Gregorek, K. H. Chrzanowska and L. Hammarström (2002). Alternative end joining during switch recombination in patients with ataxia-telangiectasia. Eur J Immunol **32**(5): 1300-1308.

Pan Q., H. Rabbani, F. C. Mills, E. Severinson and L. Hammarström (1997). Allotype-associated variation in the human gamma3 switch region as a basis for differences in IgG3 production. J Immunol **158**(12): 5849-5859.

Pan-Hammarström Q., A. M. Jones, A. Lahdesmaki, W. Zhou, R. A. Gatti, L. Hammarström, A. R. Gennery and M. R. Ehrenstein (2005). Impact of DNA ligase IV on nonhomologous end joining pathways during class switch recombination in human cells. J Exp Med **201**(2): 189-194.

Ramiro A. R., M. Jankovic, T. Eisenreich, S. Difilippantonio, S. Chen-Kiang, M. Muramatsu, T. Honjo, A. Nussenzweig and M. C. Nussenzweig (2004). AID is required for c-myc/IgH chromosome translocations in vivo. Cell **118**(4): 431-438.

Salzer U., H. M. Chapel, A. D. Webster, Q. Pan-Hammarström, A. Schmitt-Graeff, M. Schlesier, H. H. Peter, J. K. Rockstroh, P. Schneider, A. A. Schaffer, L. Hammarström and B. Grimbacher (2005). Mutations in TNFRSF13B encoding TACI are associated with common variable immunodeficiency in humans. Nat Genet **37**(8): 820-828.

Schaffer F. M., J. Palermos, Z. B. Zhu, B. O. Barger, M. D. Cooper and J. E. Volanakis (1989). Individuals with IgA deficiency and common variable immunodeficiency share polymorphisms of major histocompatibility complex class III genes. Proc Natl Acad Sci U S A **86**(20): 8015-8019.

Schrader C. E., J. Vardo and J. Stavnezer (2002). Role for mismatch repair proteins Msh2, Mlh1, and Pms2 in immunoglobulin class switching shown by sequence analysis of recombination junctions. J Exp Med **195**(3): 367-373.

Sekine H., K. L. Graham, S. Zhao, M. K. Elliott, P. Ruiz, P. J. Utz and G. S. Gilkeson (2006). Role of MHC-linked genes in autoantigen selection and renal disease in a murine model of systemic lupus erythematosus. J Immunol **177**(10): 7423-7434.

Sekine H., H. Watanabe and G. S. Gilkeson (2004). Enrichment of anti-glomerular antigen antibody-producing cells in the kidneys of MRL/MpJ-Fas(lpr) mice. J Immunol **172**(6): 3913-3921.

Snowden T., S. Acharya, C. Butz, M. Berardini and R. Fishel (2004). hMSH4-hMSH5 recognizes Holliday Junctions and forms a meiosis-specific sliding clamp that embraces homologous chromosomes. Mol Cell **15**(3): 437-451.

Svetlanov A. and P. E. Cohen (2004). Mismatch repair proteins, meiosis, and mice: understanding the complexities of mammalian meiosis. Exp Cell Res **296**(1): 71-79.

Taube C., J. M. Thurman, K. Takeda, A. Joetham, N. Miyahara, M. C. Carroll, A. Dakhama, P. C. Giclas, V. M. Holers and E. W. Gelfand (2006). Factor B of the alternative complement pathway regulates development of airway hyperresponsiveness and inflammation. Proc Natl Acad Sci U S A **103**(21): 8084-8089.

Theofilopoulos A. N. and F. J. Dixon (1985). Murine models of systemic lupus erythematosus. Adv Immunol **37**: 269-390.

Vorechovsky I., M. Cullen, M. Carrington, L. Hammarström and A. D. Webster (2000). Fine mapping of IGAD1 in IgA deficiency and common variable immunodeficiency: identification and characterization of haplotypes shared by affected members of 101 multiple-case families. J Immunol **164**(8): 4408-4416.

Vorechovsky I., A. D. Webster, A. Plebani and L. Hammarström (1999). Genetic linkage of IgA deficiency to the major histocompatibility complex: evidence for allele segregation distortion, parent-of-origin penetrance differences, and the role of anti-IgA antibodies in disease predisposition. Am J Hum Genet **64**(4): 1096-1109.

Vorechovsky I., H. Zetterquist, R. Paganelli, S. Koskinen, A. D. Webster, J. Bjorkander, C. I. Smith and L. Hammarström (1995). Family and linkage study of selective IgA deficiency and common variable immunodeficiency. Clin Immunol Immunopathol **77**(2): 185-192.

Weinstock D. M., B. Elliott and M. Jasin (2006). A model of oncogenic rearrangements: differences between chromosomal translocation mechanisms and simple double-strand break repair. Blood **107**(2): 777-780.

Wu X., C. Y. Tsai, M. B. Patam, H. Zan, J. P. Chen, S. M. Lipkin and P. Casali (2006). A role for the MutL mismatch repair Mlh3 protein in immunoglobulin class switch DNA recombination and somatic hypermutation. J Immunol **176**(9): 5426-5437.

Xu Y. (2006). DNA damage: a trigger of innate immunity but a requirement for adaptive immune homeostasis. Nat Rev Immunol **6**(4): 261-270.

Yi W., X. Wu, T. H. Lee, N. A. Doggett and C. Her (2005). Two variants of MutS homolog hMSH5: prevalence in humans and effects on protein interaction. Biochem Biophys Res Commun **332**(2): 524-532.

APPENDIX II: IS APEX1 INVOLVED IN IG GENE CONVERSION?

In this appendix I report unpublished data from experiments to test what role, if any, the base excision repair protein APEX1 plays during antibody diversification. William Brown is credited for making the human and chicken wildtype APEX1 expression plasmids. Experiments reported in **Figure AII-1** and **Figure AII-2** were also conducted by William Brown. I performed all other experiments and data collection, unless noted otherwise in figure legends.

INTRODUCTION

In the DNA deamination model of antibody diversification, uracils introduced by activation induced deaminase (AID) are recognized and removed by the base excision repair (BER) pathway (Petersen-Mahrt et al. 2002). Consistent with this model, disruption or inhibition of the BER protein, uracil nucleotide glycosylase (UNG2), results in severely reduced levels of somatic hypermutation (SHM) (Di Noia and Neuberger 2002; Rada et al. 2004; Rada et al. 2002; Saribasak et al. 2006), immunoglobulin gene conversion (IGC) (Di Noia and Neuberger 2004; Saribasak et al. 2006), and class switch recombination (CSR) (Rada et al. 2004; Rada et al. 2002). The major vertebrate endonuclease that is responsible for recognizing and repairing abasic sites, such as those generated by UNG2, apurinic endonuclease (APEX1), has been proposed to act downstream of UNG2 (Petersen-Mahrt et al. 2002). To begin testing the hypothesis that APEX1 is involved in uracil excision during antibody diversification we first asked if APEX1 was limiting in IGC.

MATERIALS AND METHODS

Expression constructs

Human and chicken APEX1 were cloned into pIRES-Hyg by William Brown to make pRH_501 and pRH_502, respectively. E96Q was introduced into pRH_501 by site directed mutagenesis using primers 5'-CCA GAT ATA CTG TGC CTT CAA CAG ACC AAA TGT TCA GAG-3' and 5'-CTC TGA ACA TTT GGT CTG TTG AAG GCA CAG TAT ATC TGG-3'. D210N was introduced into pRH_501 using 5'-CCC TTG TGC TGT GTG GAA ACC TCA ATG TGG CA-3' and 5'-TGC CAC ATT GAG GTT TCC ACA CAG CAC AAG GG-3'. E79Q was introduced into pRH_502 using 5'-CGT TGT TTG CCT TCA GCA GAC CAA ATG TGG GGC-3' and 5'-GCC CCA CAT TTG GTC TGC TGA AGG CAA ACA ACG-3'. D193N was introduced using 5'-CGT CCT CTG CGG GAA CCT CAA CGT GGC-3' and 5'-

GCC ACG TTG AGG TTC CCG CAG AGG ACG-3'. The UGI expression plasmid was obtained from Michael Neuberger. The AID expression plasmid was generated by Donna MacDuff. A list of plasmids used in these studies can be found in **Table AII-1**.

DT40 cell culture and transfection

DT40 cultures were maintained in RPMI medium supplemented with 7% fetal bovine serum, 3% chicken serum, 50 μM β-mercapoethanol, penicillin, and streptomycin. Cultures were incubated at 37°C in 5% CO_2.

Stable cell lines were constructed by transfecting 1 x 10^7 DT40 cells in PBS with 30 μg linearized plasmid (250 V, 950 μF, BioRad GenePulser II). Transfected cells were selected using 1.5 mg/ml Hygromycin (Cellgro) 24 hours post-transfection. A list of stable cell lines generated can be found in **Table AII-2**.

Western Blotting

10^6 cells were lysed in 200 μl RIPA buffer containing protease inhibitors (Complete Protease Inhibitor Tablets, Roche), sonicated for 10 seconds and centrifuged at 16,000 x g for 20 minutes at 4°C to remove insoluble material. Protein concentration was determined by Bradford assay and samples normalized to a uniform concentration. 30 μl of lysate was mixed with 6 μl of 6x reducing protein loading buffer, heated at 98°C for 10 minutes, and loaded onto a 7.5% polyacrylamide gel. Gels were run at 160 volts until the dye reached the bottom. Separated protein was transferred to PVDF (80 volts for 90 minutes) at 4°C.

Membranes were blocked in 5% nonfat dry milk in PBS + 0.1% Tween (PBST) for at least one hour at room temperature. Blots were incubated with primary antibody (1:500 dilution of α-APEX1, Santa Cruz, in PBST containing 1% nonfat milk), washed 8x with PBST, incubated with HRP-conjugated α-goat secondary antibody (1:1000 dilution in PBST + 1%

nonfat milk), and washed at least 6 x 10 minutes with PBST. HRP substrate detection was performed using Roche Lumi-Lite according to manufacturer's instructions. Blots were stripped using pH 2.0 SDS stripping buffer and reprobed with α-tubulin (1:30,000 primary, 1:3000 α-mouse secondary).

Oligo cleavage assay

10^6 cells were lysed in 200 µl HED buffer containing protease inhibitors (Complete Protease Inhibitor Tabs, Roche) and centrifuged at 16,000 x g for 20 minutes to remove insoluble material. Protein concentration was determined by Bradford assay and samples normalized to a uniform concentration.

For each reaction, 1 pM of oligo (annealed by boiling equimolar amounts of 5'-ATT ATT ATT ATT CCG UGG ATT TAT TTA TTT ATT TAT TTA TTT-fluorescein-3' and 5'-AAA TAA ATA AAT AAA TAA ATA AAT CCG CGG AAT AAT AAT AAT-3' and allowing to cool slowly to room temperature in a beaker of 500 mls of water), 1 µl UDG Buffer (New England Biolabs), and 8 µl HED lysate were incubated at 37°C for 10 minutes and transferred to ice. 10 µl of 2x formamide sample buffer was added immediately to completely halt the reaction. All 20 µl were loaded onto a 10% TBE-Urea gel and electrophoresed at 150 volts until the dye reached the bottom of the gel. Gel products were imaged using a Fuji 7000 imager.

Quantitative RT-PCR

Total RNA was isolated from 10^6 DT40 cells using the RNeasy kit (Qiagen). RNA samples were DNAse according to manufacturer's directions (TurboDNAse, Ambion), quantified by spectrometer reading at A260, and normalized to 250 ng RNA/µl. 10 µl RNA was

reverse transcribed using Transcriptor reverse transcriptase (Roche) and oligo-d(T) primers according to manufacturer's directions.

PCR reactions were performed using 100 ng of reversed transcribed RNA per reaction. Endpoint RT-PCR screening for chicken APEX1 expression, shown in Figure AI-4, was performed using primers 5'- CAC GCT CTA CCT GGC CCT AT-3' and 5'- AAC ACA CAC CGG CCT TAT TC-3' (desired band size of 325 nucleotides). Quantitative PCR for chicken APE was done using primers 5'- GCA CAT GAA AGG GCG AAT-3' and 5'- GGC GGA ATT CCT CTA GTG C-3' and universal probe library (UPL) probe #47 (Roche). GAPDH mRNA was quantified using primers 5'- GGA GTC CAC TGG TGT CTT CAC-3' and 5'- GCT TAG CAC CAC CCT TCA GA-3' with UPL #133. Chicken AID expression was measured using primers 5'- CCA TCT GTC CAG GAA ACT CC-3' and 5'- AGT TTT AAA GGC ATC TCG TAA ATC A-3' with UPL #29. All quantitative PCR reactions were performed on a LightCycler 480 II (Roche).

Immunoglobulin gene conversion assay

On ice, 2 x 10^5 DT40 cells were incubated with 100 µl of R-PE conjugated mouse-anti chicken IgM antibody (Southern Biotechnology, Inc.) that had been diluted 1:400 in PBS. Cells were analyzed by flow cytometry using either a FACScalibur (Beckton Dickinson) or Quanta SC MPL (Beckman Coultier). Live cells were gated by forward scatter / side scatter (FACScalibur) or electronic volume / side scatter (Quanta SC MPL). Gates for IgM+ and IgM- populations were established using unstained and stained control cells.

RESULTS AND DISCUSSION

The DT40 cell line (clone c18/c10) contains a frameshift mutation in the variable region that introduces a premature stop and prevents the expression of IgM on the cell surface. DT40 cells constitutively undergo Ig gene conversion (IGC) whereby variable region DNA undergoes templated replacement with sequence from an upstream pseudo-variable region. If an IGC event repairs the frameshift mutation, surface IgM expression will be restored. Frequency of IGC can therefore be monitored by subcloning the surface IgM- DT40 clone c18/c10 and measuring the fraction of cells that express IgM after clonal expansion.

To determine if APEX was limiting during IGC, a human APEX1 expression construct was stably transfected into DT40 clone c18/c10. Clones were selected by resistance to Hygromycin and APEX expression confirmed by Western blot (**Figure AII-1A**). Cultures were subcloned by limiting dilution and allowed to expand for 28 days, at which time IGC was assayed by counting the fraction of cells expressing surface IgM by flow cytometry. Unexpectedly, expression of human APEX1 did not increase the rate of IGC in DT40 compared to cells transfected with empty vector (**Figure AII-1B**). In a repeat experiment in which two human APEX1 expressing clones were assayed, an approximate 40% reduction in IGC was noted (**Figure AII-2**). **Figures AII-1** and **AII-2** were prepared by William Brown, who also performed the experiments.

To confirm the results obtained by William Brown, I stably transfected human APEX1 into DT40 clone c18/c10 and screened for expressing clones (**Figure AII-3A**). Four expressing clones were selected and subcloned, along with four clones transfected with empty vector. After 28 days of culture, subclones were analyzed for surface IgM expression (**Figure AII-3B**). Again, human APEX1 expression failed to increase the rate of IGC, and in fact, a slight decrease was noted (**Figure AII-3C**).

As APEX1 may be playing a role in SHM that is independent from that in IGC, I also generated stable DT40 clone c18/c4 (surface IgM+) cell lines (**Figure AII-3A**). SHM may occasionally introduce mutations that prevent the expression of surface IgM. Loss of surface IgM after 28 days of culture was assayed as a measure of SHM in four human APEX1 expressing subclones and four vector only controls (**Figure AII-3D**). Again, no difference in surface IgM loss was noted between APEX1 expressing cell lines and vector controls.

We next decided to test if overexpression of chicken APEX1 (cloned by William Brown) had an effect on IGC. Chicken APEX1 overexpressing DT40 clone c18/c10 cell lines were made by stable transfection, selected by Hygromycin resistance, and expression of APEX1 mRNA from the transfected construct confirmed by RT-PCR (**Figure AII-4A**). Three chicken APEX1 overexpressing clones, three human APEX1 expressing clones, and three empty vector control DT40 cell lines were subcloned and expanded for 28 days, at which time surface IgM expressing fractions were measured (**Figure AII-4B**). A DT40 clone c18/c10 line expressing the UNG inhibitor UGI was included as a control (obtained from Michael Neuberger). Overexpression of chicken APEX1 did not appreciably alter IGC compared to vector controls in this experiment (**Figure AII-4C**). Human APEX1 expressing clones again trended towards slightly reduced levels of IGC.

During the course of these studies, it was reported that mutations in the two catalytically active sites created a dominant negative APEX1 protein with approximately 13-fold higher affinity for abasic sites compared to wildtype, and an estimated >56,000,000-fold reduction in nuclease capacity (McNeill and Wilson 2007). These two mutations (E96Q and D210N) were introduced into APEX1 expression plasmids by site directed mutagenesis. These constructs were stably transfected into DT40 clone c18/c10 in parallel with wildtype human APEX and control vectors expressing UGI, chicken AID, and empty vector. Expression of wildtype and mutant human APEX1 in selected clones was confirmed by Western blotting

(**Figure AII-5A**). Protein lysates from UGI expressing controls were assayed for ability to cleave a synthetic U:G mismatch containing oligo and an indicator of UNG2 inhibition (**Figure AII-5B**). Expression of cAID in transfected clones was determined by quantitative real-time PCR and was compared to clones transfected with empty vector (**Figure AII-5C**). Selected clones were again subcloned and cultured for 28 days before measuring surface IgM expressing fractions by flow cytometry (**Figure AII-5D**). Human APEX1 expressing clones, as well as the single-site catalytic mutants did not show an appreciable difference in IGC rate compared to empty-vector controls (**Figure AII-5E**). IGC was slightly reduced in clones expressing the double-mutant E96Q/D210N form of APEX1, however this reduction is not greater than what was observed for wildtype human APEX1 in previous experiments (**Figures AII-1 AII-4**). Unfortunately, the IGC was not effectively ablated in the UGI-expressing control cell line as was observed with the cell line obtained from Michael Neuberger (**Figure AII-5** and **AII-4**).

To test if chicken APEX1 containing equivalent mutations in the active site regions exerted an effect on IGC, an expression construct containing the mutations E79Q (equivalent to E96Q in human) and D193N (equivalent to D210N in human) were created by site directed mutagenesis and stably transfected into DT40 clone c18/c10. Expression of transfected constructs was confirmed prior to subcloning (**Figure AII-6**) and IGC was assayed by surface IgM expression after 28 days of culture (**Figure AII-7**). Overexpression of chicken APEX1 resulted in decreased IGC (**Figure AII-7B**). IGC in the catalytic mutant was slightly higher, although not as high as the vector control. Expression of both wildtype and mutant human APEX1 resulted in decreased IGC. Although not to the extent expected, AID overexpression caused a slight increase in IGC.

Overall, overexpression of APEX1, whether human or chicken, did not increase the level of IGC in DT40 cells. In multiple experiments, IGC levels were similar, or slightly reduced, compared to empty-vector controls. This suggests that it is not limiting during IGC,

and may perhaps be directing AID/UNG2 generated abasic sites toward repair rather than diversification. Overexpression of catalytic mutants of APEX1, which are reported to exert a dominant negative function, also did not have a drastic effect on IGC. Studies of APEX1 depletion or functional inhibition are required to more thoroughly address the role of APEX1 in IGC.

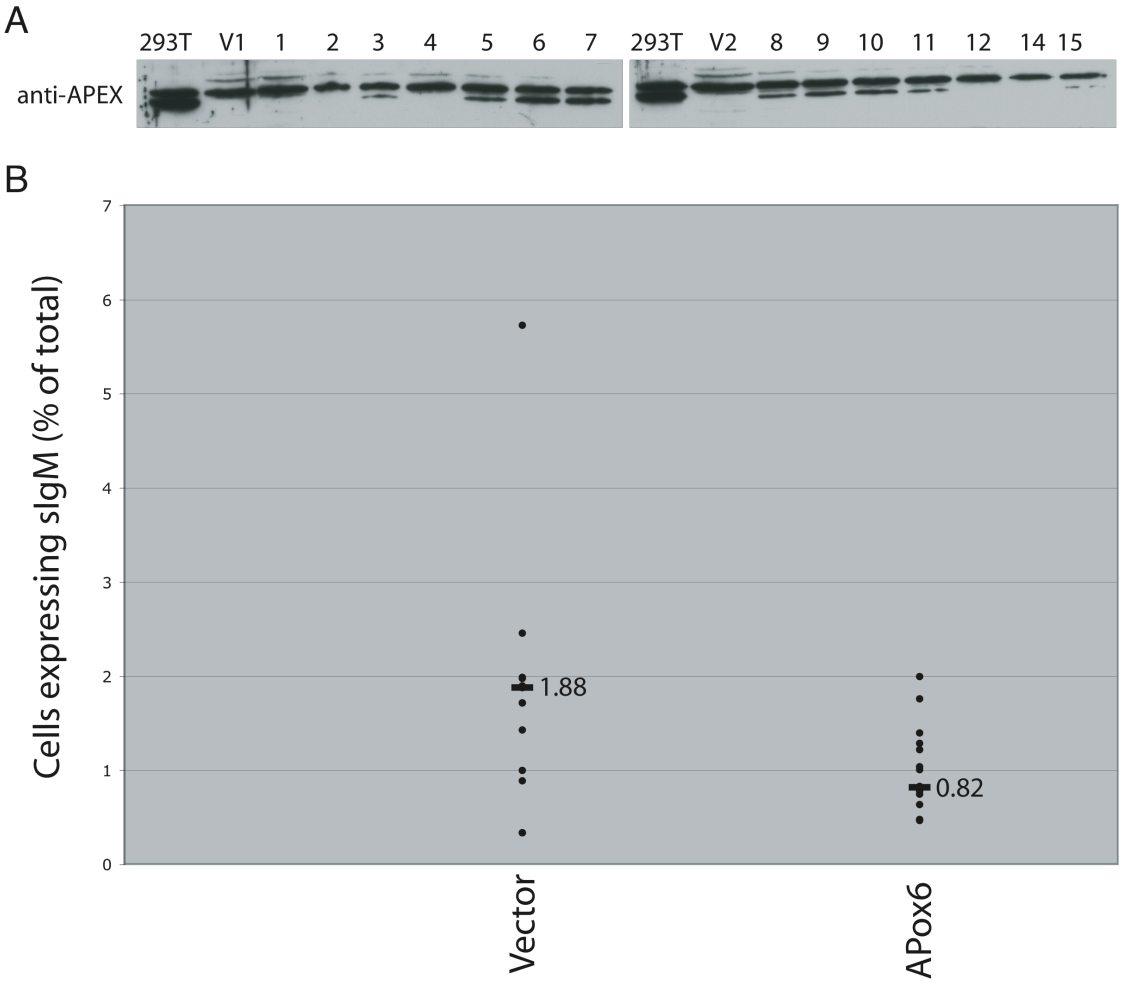


Figure AII-1. Human APEX1 expression in DT40 – Experiment done by William Brown. Human APEX1 and empty vector were linearized and stably transfected into DT40 c18/c10 (surface IgM-). (**A**) Clones were selected and screened for hAPEX1 expression by Western blot (α-APEX1, Novus, "Lot F" antibody). 293T cells express APEX1, and were used as a positive control. (**B**) Vector clone V1 and hAPEX1 expressing clone APOX6 were subcloned. When single colonies were visible, subclones were picked, expanded to larger culture volumes, and fed as needed. After 28 days of culture (from day of subcloning), cultures were stained with RPE conjugated anti-IgM and analyzed by flow cytometry for percentage of cells expressing suface Ig as an indicator of IGC. % of culture expressing surface IgM is reported for each subclone (dots). Median values are reported as bars.

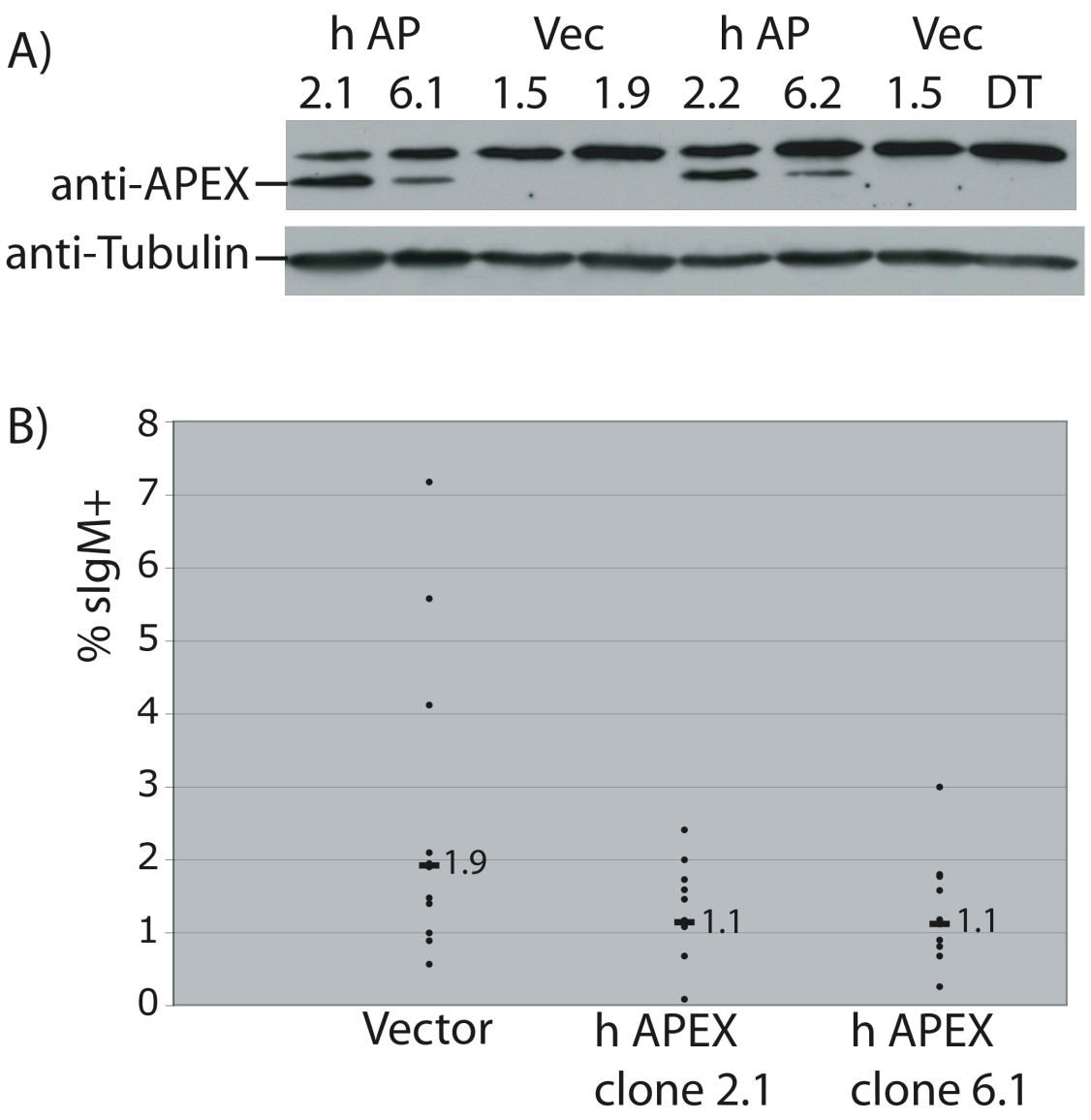


Figure AII-2. Human APEX1 overexpression in DT40 – Experiment done by William Brown. Transfected DT40 clones screened for hAPEX1 expression in **Figure AI-1** were sublconed. (**A**) Western blot to confirm expression of hAPEX1. (**B**) Cultures were subcloned and multiple subclones were analyzed for surface IgM expression after 28 days of culture. Percentage IgM positive cells in each subclonal culture are presented as dots. Median values are presented as bars.

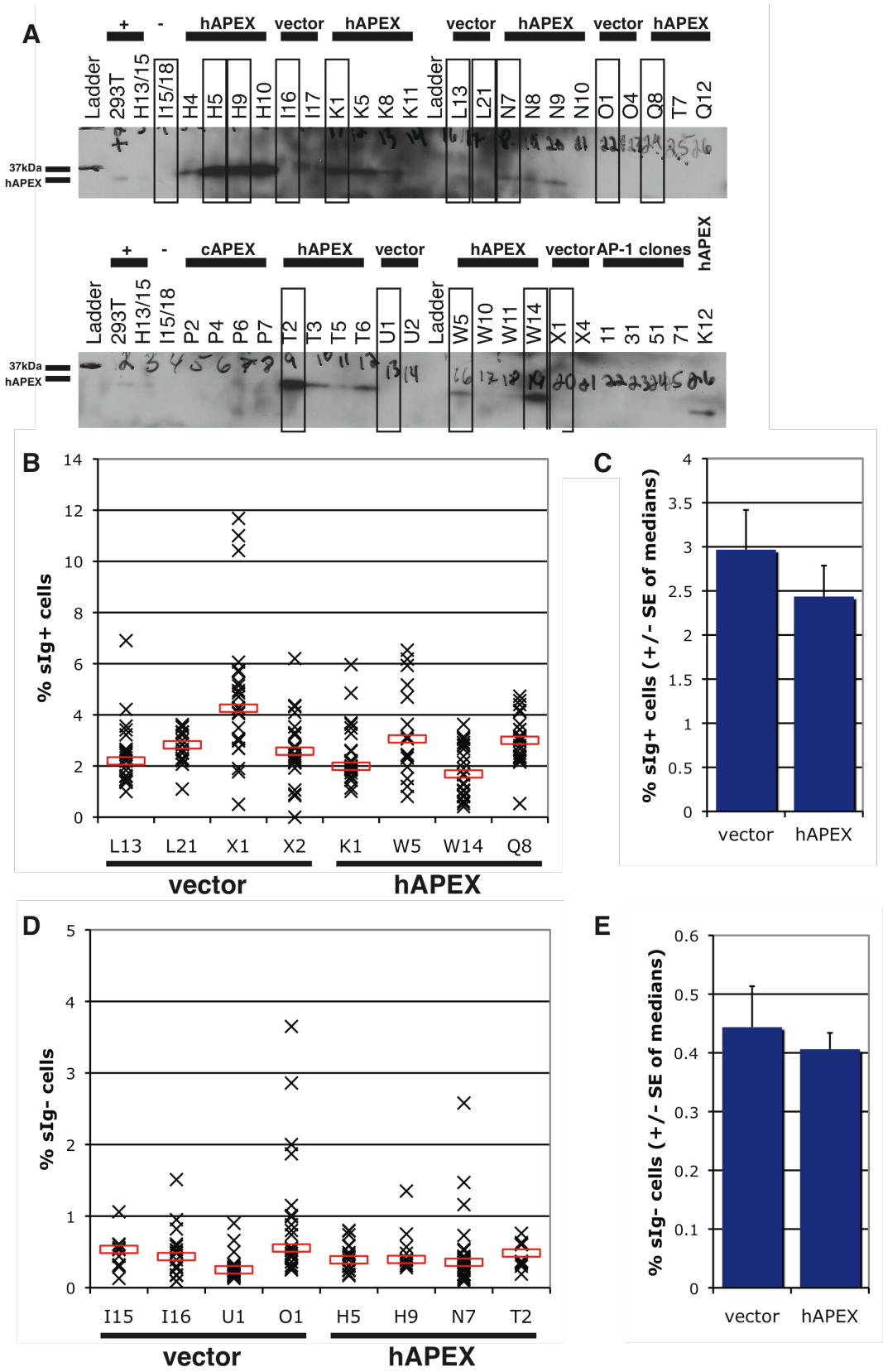

A
+
-
hAPEX
vector
hAPEX
vector
hAPEX
vector
hAPEX
Ladder
293T
H13/15
I15/18
H4
H5
H9
H10
I16
I17
K1
K5
K8
K11
Ladder
L13
L21
N7
N8
N9
N10
O1
O4
Q8
T7
Q12
37kDa
hAPEX
+
-
cAPEX
hAPEX
vector
hAPEX
vector
AP-1 clones
hAPEX
Ladder
293T
H13/15
I15/18
P2
P4
P6
P7
T2
T3
T5
T6
U1
U2
Ladder
W5
W10
W11
W14
X1
X4
11
31
51
71
K12
37kDa
hAPEX
B
% sIg+ cells
14
12
10
8
6
4
2
0
L13
L21
X1
X2
K1
W5
W14
Q8
vector
hAPEX
C
% sIg+ cells (+/- SE of medians)
4
3.5
3
2.5
2
1.5
1
0.5
0
vector
hAPEX
D
% sIg- cells
5
4
3
2
1
0
I15
I16
U1
O1
H5
H9
N7
T2
vector
hAPEX
E
% sIg- cells (+/- SE of medians)
0.6
0.5
0.4
0.3
0.2
0.1
0
vector
hAPEX

Figure AII-3. Confirm reduction in IGC efficiency observed by William Brown in Figures AII-1 and AII-2. (**A**) DT40 clones were generated by stable transfection of hAPEX1 and empty vector. Clones were screened for hAPEX expression by Western blotting using α-APEX1 (Santa Cruz). The boxed clones were used in IGC fluctuation testing. (**B**) Four hAPEX1 expressing DT40 (c18/c10) clones, and four vector controls (c18/c10) were subcloned and expanded for 28 days, at which time they were assayed for surface IgM expression by flow cytometry. The percentage of surface IgM expressing cells is presented as an "x" for each individual subclone. The median value for each clone is denoted as a red bar. (**C**) Calculated mean of median values +/- standard error for data presented in (B). (**D**) Four hAPEX1 expressing DT40 (c18/c4, sIgM+) clones, and four vector controls (c18/c4) were subcloned and expanded for 28 days. Proportion of cells reverting to IgM- was determined by flow cytometry as an indicator of SHM. (**E**) Mean +/- standard error of median values reported in (D).

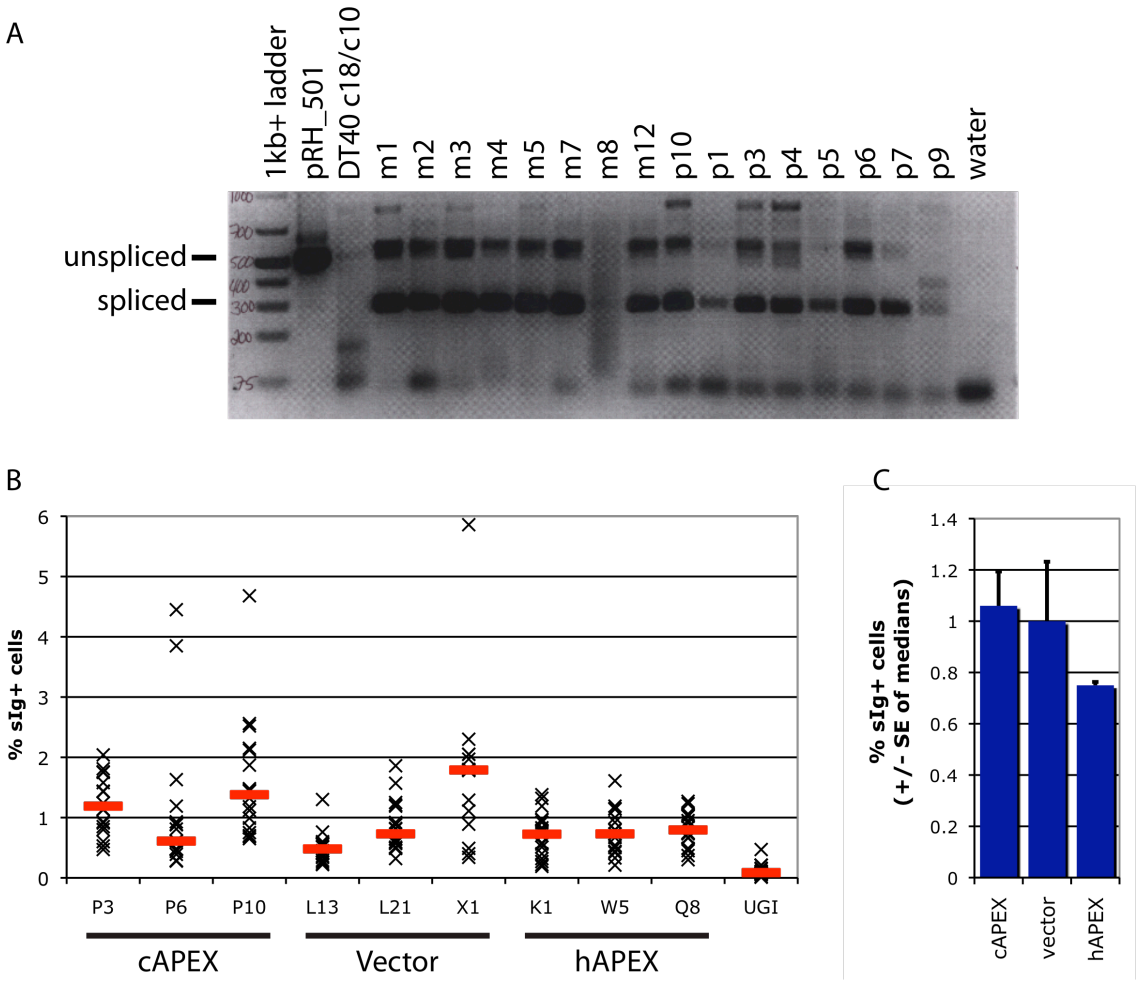


Figure AII-4. Overexpress chicken APEX1 in DT40. Chicken APEX1, hAPEX1, and empty vector were stably expressed in DT40. (**A**) As none of the antibodies tested by our lab significantly cross-reacted with chicken APEX1, a RT-PCR strategy was designed to screen transgenic clones for cAPEX1 expression. The mRNA expressed from the inserted transgene should contain both APEX1 and the Hyg resistance gene, separated by an internal ribosomal entry site and synthetic intron of approximately 225 bp. PCR primers were designed to amplify across the synthetic intron. Amplification of spliced transgenic reverse transcribed mRNA should yield a PCR product approximately 325 bp in length. Un-spliced mRNA, genomic DNA containing the integration, and plasmid should yield a band of ~550 bp. (**B**) IGC fluctuation of DT40 subclones. Individual subclones are represented as and "x." Medians are presented as red bars. Screening of hAPEX1 expressing lines is reported in **Figure AII-5**. An ugi-expression DT40 clone (Obtained from Michael Neuberger) was included as a control. (**C**) Mean of medians for each clone +/- standard error.

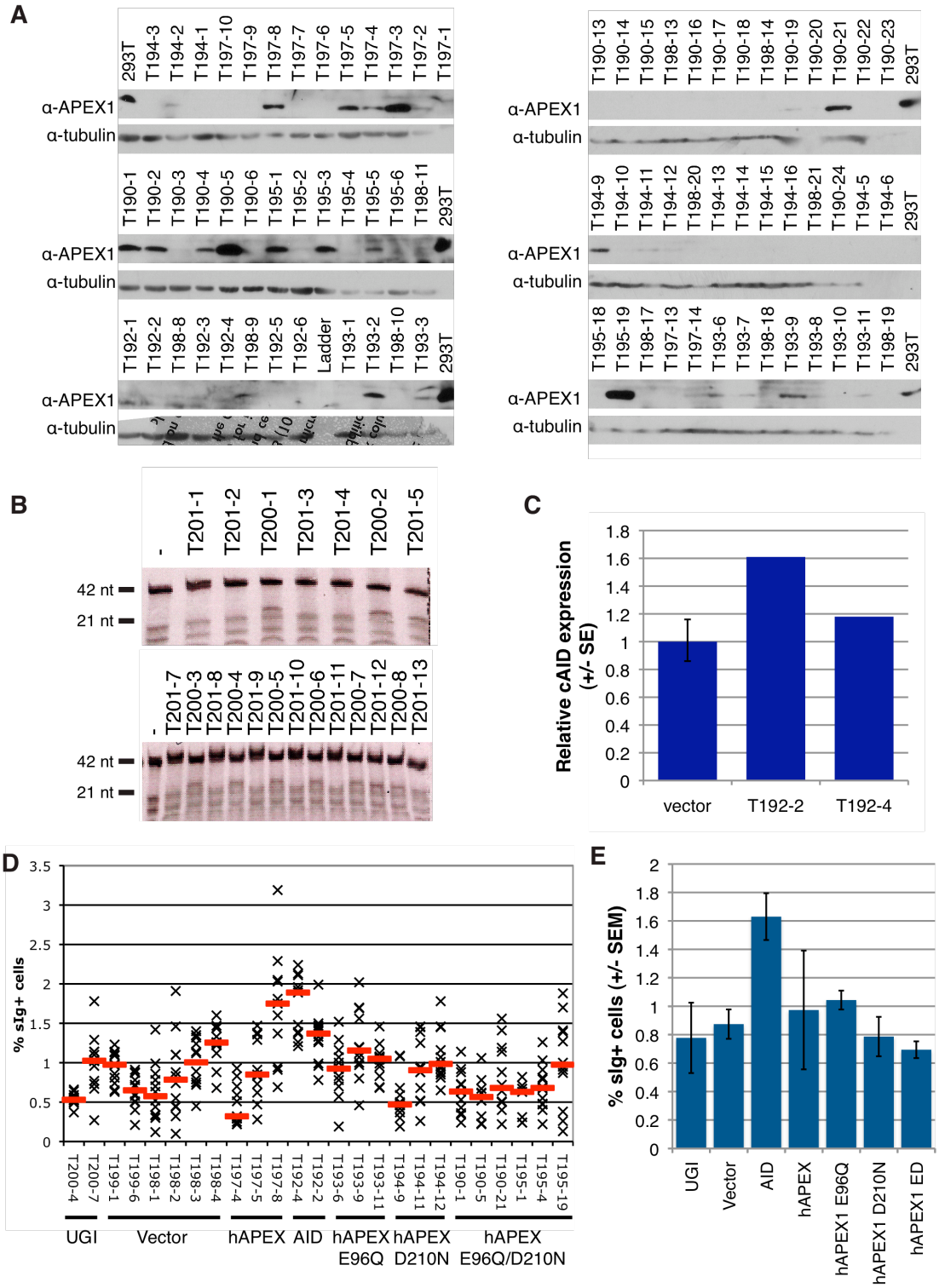

A
293T T194-3 T194-2 T194-1 T197-10 T197-9 T197-8 T197-7 T197-6 T197-5 T197-4 T197-3 T197-2 T197-1
α-APEX1
α-tubulin
T190-1 T190-2 T190-3 T190-4 T190-5 T190-6 T195-1 T195-2 T195-3 T195-4 T195-5 T195-6 T198-11 293T
T192-1 T192-2 T198-8 T192-3 T192-4 T198-9 T192-5 T192-6 Ladder T193-1 T193-2 T198-10 T193-3 293T
T190-13 T190-14 T190-15 T198-13 T190-16 T190-17 T190-18 T198-14 T190-19 T190-20 T190-21 T190-22 T190-23 293T
T194-9 T194-10 T194-11 T194-12 T198-20 T194-13 T194-14 T194-15 T194-16 T198-21 T190-24 T194-5 T194-6 293T
T195-18 T195-19 T198-17 T197-13 T197-14 T193-6 T193-7 T198-18 T193-9 T193-8 T193-10 T193-11 T198-19 293T
B
- T201-1 T201-2 T200-1 T201-3 T201-4 T200-2 T201-5
42 nt
21 nt
- T201-7 T200-3 T201-8 T200-4 T201-9 T200-5 T201-10 T200-6 T201-11 T200-7 T201-12 T200-8 T201-13
C
Relative cAID expression (+/- SE)
vector T192-2 T192-4
D
% sIg+ cells
UGI Vector hAPEX AID hAPEX E96Q hAPEX D210N hAPEX E96Q/D210N
E
% sIg+ cells (+/- SEM)
UGI Vector AID hAPEX hAPEX1 E96Q hAPEX1 D210N hAPEX1 ED

Figure AII-5. Dominant negative hAPEX1. New DT40 c18/c10 clones were generated to express UGI, human APEX1 and the human APEX1 catalytic site mutants E96Q, D210N, and E96Q/D210N. (**A**) Clones were screened by Western blot. (**B**) Uracil cleavage activity of UGI expressing clones. Uncleaved oligo runs at 42 nucleotides and cleaved product is 21 nt. Extra low-mass bands are because we did not order HPLC purified oligos for this assay. (**C**) Two representative AID expression clones selected for IGC analysis. (**D**) Subclones of selected transfectants were cultured for 28 days and analyzed for surface IgM expression by flow cytometry. The percentage of subclone cultures expressing sIgM are represented by an "x" and the median of subclones is presented as a red bar. (**E**) Mean of medians for each set of subclones +/- standard error.

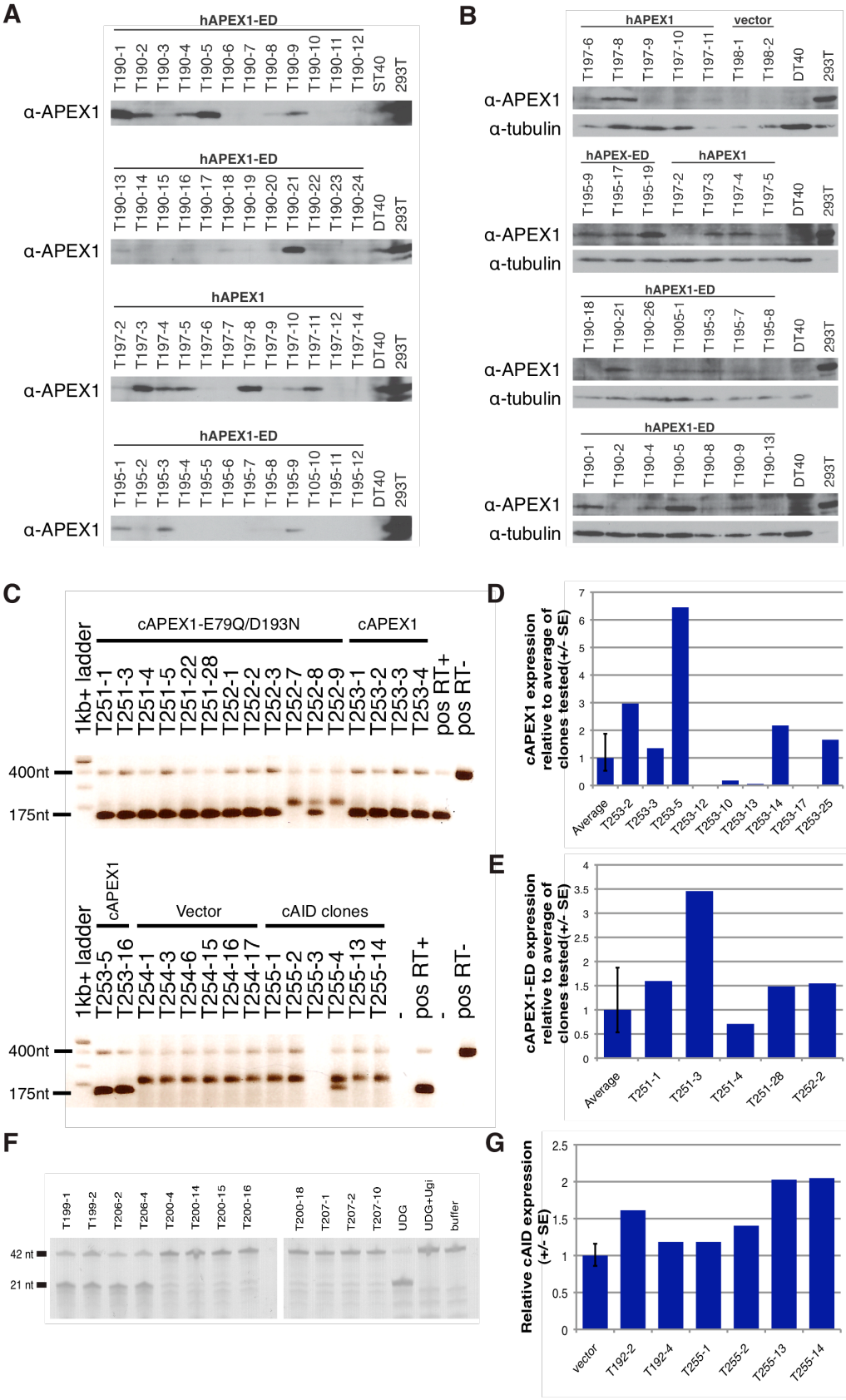

A
hAPEX1-ED
T190-1 T190-2 T190-3 T190-4 T190-5 T190-6 T190-7 T190-8 T190-9 T190-10 T190-11 T190-12 ST40 293T
α-APEX1
hAPEX1-ED
T190-13 T190-14 T190-15 T190-16 T190-17 T190-18 T190-19 T190-20 T190-21 T190-22 T190-23 T190-24 DT40 293T
α-APEX1
hAPEX1
T197-2 T197-3 T197-4 T197-5 T197-6 T197-7 T197-8 T197-9 T197-10 T197-11 T197-12 T197-14 DT40 293T
α-APEX1
hAPEX1-ED
T195-1 T195-2 T195-3 T195-4 T195-5 T195-6 T195-7 T195-8 T195-9 T105-10 T195-11 T195-12 DT40 293T
α-APEX1
B
hAPEX1
vector
T197-6 T197-8 T197-9 T197-10 T197-11 T198-1 T198-2 DT40 293T
α-APEX1
α-tubulin
hAPEX-ED
hAPEX1
T195-9 T195-17 T195-19 T197-2 T197-3 T197-4 T197-5 DT40 293T
α-APEX1
α-tubulin
hAPEX1-ED
T190-18 T190-21 T190-26 T1905-1 T195-3 T195-7 T195-8 DT40 293T
α-APEX1
α-tubulin
hAPEX1-ED
T190-1 T190-2 T190-4 T190-5 T190-8 T190-9 T190-13 DT40 293T
α-APEX1
α-tubulin
C
cAPEX1-E79Q/D193N
cAPEX1
1kb+ ladder T251-1 T251-3 T251-4 T251-5 T251-22 T251-28 T252-1 T252-2 T252-3 T252-7 T252-8 T252-9 T253-1 T253-2 T253-3 T253-4 pos RT+ pos RT-
400nt
175nt
cAPEX1
Vector
cAID clones
1kb+ ladder T253-5 T253-16 T254-1 T254-3 T254-6 T254-15 T254-16 T254-17 T255-1 T255-2 T255-3 T255-4 T255-13 T255-14 - pos RT+ - pos RT-
400nt
175nt
D
cAPEX1 expression relative to average of clones tested(+/- SE)
0 1 2 3 4 5 6 7
Average T253-2 T253-3 T253-5 T253-12 T253-10 T253-13 T253-14 T253-17 T253-25
E
cAPEX1-ED expression relative to average of clones tested(+/- SE)
0 0.5 1 1.5 2 2.5 3 3.5 4
Average T251-1 T251-3 T251-4 T251-28 T252-2
F
T199-1 T199-2 T206-2 T206-4 T200-4 T200-14 T200-15 T200-16 T200-18 T207-1 T207-2 T207-10 UDG UDG+Ugi buffer
42 nt
21 nt
G
Relative cAID expression (+/- SE)
0 0.5 1 1.5 2 2.5
vector T192-2 T192-4 T255-1 T255-2 T255-13 T255-14

FIGURE AII-6. Express dominant negative hAPEX1 and cAPEX1, part 1 – screen clones. (**A** and **B**) DT40 clones expressing human APEX1 and catalytic mutants E96N/D210N were screened by western blotting. (**C**) Transgenic chicken APEX1 expressing clones were screened for mRNA expression RT-PCR. Primers were designed to generate a shorter, 175 bp spliced product (400 bp unspliced) that that in FigureAI-4. (**D**) Relative levels of transgenic chicken APEX1 expression as determined by quantitative RT-PCR. Levels are normalized to GAPDH and are reported relative to the calculated mean of all expressing clones (+/- standard error). Clone T253-12 did not express detectable transgenic cAPEX1. (**E**) Transgenic cAPEX-E96Q/D210N levels were normalized to GAPDH and are also presented relative to the average of expressing clones. (**F**) UNG activity was assayed in UGI-expressing DT40 clones. Assay was performed as in Figure AII-5, except that five times as much lysate was used per reaction. Bill Brown performed the UNG activity assay pictured using lysates that were prepared, quantified, and normalized by me. (**G**) AID expression (relative to GAPDH) was determined by quantitative RT-PCR and is presented normalized to the mean (+/- standard error) of vector-transfected DT40 clones.

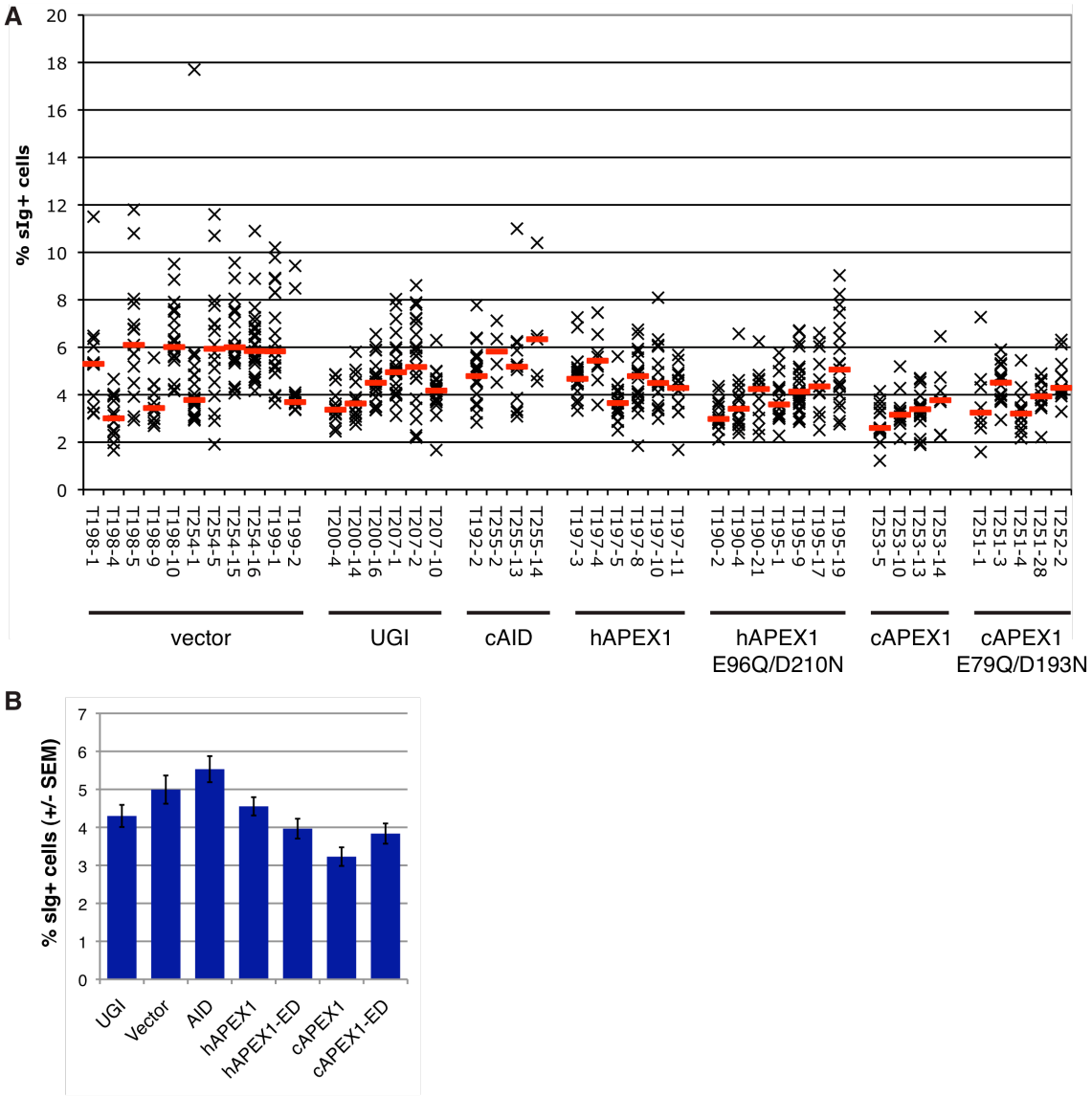


FIGURE AII-7. Express dominant negative hAPEX1 and cAPEX1, part 2 – assay IGC. (**A**) IGC was assayed in clones screened in **Figure AI-6**. The % of sIgM + cells for each subclone is represented as an "x." The median value for all subclones is depicted as a red bar. (**B**) Mean percentage +/- standard error of data presented in A is reported.

Table AII-1. Key plasmids generated/used in APEX1 studies

pRH	AKA	Plasmid name	Reference
99	pSO57	pIRES-HYG	Clontech
502	pSO56	pHAPEX1-pIRES-HYG	Appendix II
2962	pSO111	pHAPEX1-E96Q-pIRES-HYG	Appendix II
2963	psO195	pHAPEX1-D210N-pIRES-HYG	Appendix II
2964	pSO203	pHAPEX1-E96Q/D210N-pIRES-HYG	Appendix II
2968	pSO205	pHAPEX1-E96Q/D210N-pIRES-HYG	Appendix II
501	pSO55	pCAPEX1-pIRES-HYG	Appendix II
2965	pSO119	pCAPEX1-E79Q-pIRES-HYG	Appendix II
2966	pSO120	pCAPEX1-D193N-pIRES-HYG	Appendix II
2967	pSO121	pCAPEX1-E79Q/D193N in pIRES-HYG	Appendix II

Table AII-2. Key cell lines generated for APEX1 studies

Name	Clone	Cell line	Plasmid	Expressing
T198	1	DT40 c18/c10	pRH_99	vector-IRES-HYG
T198	2	DT40 c18/c10	pRH_99	vector-IRES-HYG
T198	3	DT40 c18/c10	pRH_99	vector-IRES-HYG
T198	4	DT40 c18/c10	pRH_99	vector-IRES-HYG
T198	5	DT40 c18/c10	pRH_99	vector-IRES-HYG
T198	9	DT40 c18/c10	pRH_99	vector-IRES-HYG
T198	10	DT40 c18/c10	pRH_99	vector-IRES-HYG
T198	12	DT40 c18/c10	pRH_99	vector-IRES-HYG
T254	1	DT40 c18/c10	pRH_99	vector-IRES-HYG
T254	5	DT40 c18/c10	pRH_99	vector-IRES-HYG
T254	15	DT40 c18/c10	pRH_99	vector-IRES-HYG
T254	16	DT40 c18/c10	pRH_99	vector-IRES-HYG
T199	1	DT40 c18/c10	pRH_300	vector (HYG)
T199	2	DT40 c18/c10	pRH_300	vector (HYG)
T199	6	DT40 c18/c10	pRH_300	vector (HYG)
T206	2	DT40 c18/c10	pRH_300	vector (HYG)
T206	4	DT40 c18/c10	pRH_300	vector (HYG)
T206	5	DT40 c18/c10	pRH_300	vector (HYG)
T206	6	DT40 c18/c10	pRH_300	vector (HYG)
T200	4	DT40 c18/c10	pRH_301	ugi (HYG)
T200	7	DT40 c18/c10	pRH_301	ugi (HYG)
T200	9	DT40 c18/c10	pRH_301	ugi (HYG)
T200	10	DT40 c18/c10	pRH_301	ugi (HYG)
T200	14	DT40 c18/c10	pRH_301	ugi (HYG)
T200	15	DT40 c18/c10	pRH_301	ugi (HYG)
T200	16	DT40 c18/c10	pRH_301	ugi (HYG)
T200	18	DT40 c18/c10	pRH_301	ugi (HYG)
T207	1	DT40 c18/c10	pRH_301	ugi (HYG)
T207	2	DT40 c18/c10	pRH_301	ugi (HYG)
T207	3	DT40 c18/c10	pRH_301	ugi (HYG)
T207	5	DT40 c18/c10	pRH_301	ugi (HYG)
T207	10	DT40 c18/c10	pRH_301	ugi (HYG)
T192	2	DT40 c18/c10	pRH_1212	cAID (HYG)
T192	4	DT40 c18/c10	pRH_1212	cAID (HYG)
T192	11	DT40 c18/c10	pRH_1212	cAID (HYG)
T255	1	DT40 c18/c10	pRH_1212	cAID (HYG)
T255	2	DT40 c18/c10	pRH_1212	cAID (HYG)
T255	4	DT40 c18/c10	pRH_1212	cAID (HYG)
T255	13	DT40 c18/c10	pRH_1212	cAID (HYG)
T255	14	DT40 c18/c10	pRH_1212	cAID (HYG)

Table AII-2, continued

Name	Clone	Cell line	Plasmid	Expressing
T197	3	DT40 c18/c10	pRH_502	hAPE (HYG)
T197	4	DT40 c18/c10	pRH_502	hAPE (HYG)
T197	5	DT40 c18/c10	pRH_502	hAPE (HYG)
T197	8	DT40 c18/c10	pRH_502	hAPE (HYG)
T197	10	DT40 c18/c10	pRH_502	hAPE (HYG)
T197	11	DT40 c18/c10	pRH_502	hAPE (HYG)
T197	14	DT40 c18/c10	pRH_502	hAPE (HYG)
T193	6	DT40 c18/c10	pRH2962	hAPE-E96Q (HYG)
T193	9	DT40 c18/c10	pRH2962	hAPE-E96Q (HYG)
T193	11	DT40 c18/c10	pRH2962	hAPE-E96Q (HYG)
T194	9	DT40 c18/c10	pRH2963	hAPE-D210N (HYG)
T194	11	DT40 c18/c10	pRH2963	hAPE-D210N (HYG)
T194	12	DT40 c18/c10	pRH2963	hAPE-D210N (HYG)
T190	1	DT40 c18/c10	pRH_2968	hAPE-E96Q/D210N (HYG)
T190	2	DT40 c18/c10	pRH_2968	hAPE-E96Q/D210N (HYG)
T190	4	DT40 c18/c10	pRH_2968	hAPE-E96Q/D210N (HYG)
T190	5	DT40 c18/c10	pRH_2968	hAPE-E96Q/D210N (HYG)
T190	21	DT40 c18/c10	pRH_2968	hAPE-E96Q/D210N (HYG)
T195	1	DT40 c18/c10	pRH_2964	hAPE-E96Q/D210N (HYG)
T195	4	DT40 c18/c10	pRH_2964	hAPE-E96Q/D210N (HYG)
T195	9	DT40 c18/c10	pRH_2964	hAPE-E96Q/D210N (HYG)
T195	17	DT40 c18/c10	pRH_2964	hAPE-E96Q/D210N (HYG)
T195	19	DT40 c18/c10	pRH_2964	hAPE-E96Q/D210N (HYG)
T253	2	DT40 c18/c10	pRH_501	cAPE (HYG)
T253	3	DT40 c18/c10	pRH_501	cAPE (HYG)
T253	5	DT40 c18/c10	pRH_501	cAPE (HYG)
T253	10	DT40 c18/c10	pRH_501	cAPE (HYG)
T253	13	DT40 c18/c10	pRH_501	cAPE (HYG)
T253	14	DT40 c18/c10	pRH_501	cAPE (HYG)
T253	25	DT40 c18/c10	pRH_501	cAPE (HYG)
T251	1	DT40 c18/c10	pRH2967	cAPE-E79Q/D193N (HYG)
T251	3	DT40 c18/c10	pRH2967	cAPE-E79Q/D193N (HYG)
T251	4	DT40 c18/c10	pRH2967	cAPE-E79Q/D193N (HYG)
T251	5	DT40 c18/c10	pRH2967	cAPE-E79Q/D193N (HYG)
T251	28	DT40 c18/c10	pRH2967	cAPE-E79Q/D193N (HYG)
T252	2	DT40 c18/c10	pRH2967	cAPE-E79Q/D193N (HYG)

REFERENCES

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

Di Noia J. M. and M. S. Neuberger (2004). Immunoglobulin gene conversion in chicken DT40 cells largely proceeds through an abasic site intermediate generated by excision of the uracil produced by AID-mediated deoxycytidine deamination. Eur J Immunol **34**(2): 504-508.

McNeill D. R. and D. M. Wilson, 3rd (2007). A dominant-negative form of the major human abasic endonuclease enhances cellular sensitivity to laboratory and clinical DNA-damaging agents. Mol Cancer Res **5**(1): 61-70.

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

Rada C., J. M. Di Noia and M. S. Neuberger (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.

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

Saribasak H., N. N. Saribasak, F. M. Ipek, J. W. Ellwart, H. Arakawa and J. M. Buerstedde (2006). Uracil DNA glycosylase disruption blocks Ig gene conversion and induces transition mutations. J Immunol **176**(1): 365-371.

APPENDIX III: THE HETEROZYGOUS RAD50-Q372X MUTATION CONFERS SENSITIVITY TO DOUBLE-STRAND DNA DAMAGE

In this appendix I present experiments that test the impact of the RAD50-Q372X mutation on DNA damage repair. RAD50-Q372X was identified during experiments presented as part of Chapter III. I performed all experiments and data collection, unless noted otherwise in figure legends. I would like to acknowledge Brian Ruis for technical assistance with generating the targeting vectors used in these experiments, and Yongbao Wang for assistance with gamma irradiation experiments.

A *RAD50* truncation mutation Q372X in a patient with common variable immunodeficiency

Steven M Offer[1], Qiang Pan-Hammarström[2], Lennart Hammarström[2] and Reuben S Harris[*,1]

[1]*Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, Minneapolis, Minnesota 55455, USA;*

[2]*Division of Clinical Immunology, Department of Laboratory Medicine, Karolinska Institutet at Karolinska University Hospital Huddinge, SE-141 86 Stockholm, Sweden*

[*]Correspondence: Reuben S. Harris, Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, Minneapolis, Minnesota 55455, USA.

Tel: +1 612 624 0457, Fax: +1 612 625 2163.

Email: rsh@umn.edu

Common variable immunodeficiency (CVID) is characterized by severe deficiency of IgA, IgG, and often IgM antibodies, resulting in recurrent bacterial infections. Although mutations in *TACI*, *BAFF-R*, *ICOS*, and *CD19* underlie a subset of CVID cases, the majority of the molecular defects remain undefined. We report the detection of a heterozygous *RAD50* nonsense mutation Q372X in a Swedish CVID patient. A patient-derived B lymphocyte line and a purposefully engineered HCT116 RAD50-Q372X heterozygous cell line were sensitive to ionizing radiation-induced DNA double-strand breaks. Together with a 50% reduction in RAD50 and MRE11 protein levels, we conclude that this patient is haploinsufficient for MRE11/RAD50/NBS1-mediated recombination repair and that mutations in this pathway may contribute to the class switch recombination defects observed in CVID and, potentially also, the related syndrome immunoglobulin A deficiency (IgAD).

Keywords: class switch recombination; common variable immunodeficiency syndrome; CVID; RAD50; recombination repair.

Introduction

Common variable immunodeficiency (CVID) is the most common primary immunodeficiency syndrome.(Cunningham-Rundles and Bodian 1999) It is characterized by low levels of IgA, IgG, and IgM. Proportionately diminished antibody responses often result in recurrent bacterial infections of the respiratory and gastrointestinal tracts.(Cunningham-Rundles and Bodian 1999) CVID has been documented in the same pedigrees as IgA deficiency (IgAD) strongly suggesting common genetic components.(Vorechovsky et al. 1995) This relationship is further supported by observations showing that IgAD patients sometimes progress to CVID.(Carvalho Neves Forte et al. 2000; Espanol et al. 1996; Ishizaka et al. 1989; Johnson et al. 1997) Despite this shared component, only 10-20% of CVID/IgAD cases have shown familial inheritance and only a minor fraction of causal defects have been discovered.(Castigli et al. 2005; Cunningham-Rundles and Bodian 1999; Grimbacher et al. 2003; Pan-Hammarström et al. 2007; Salzer et al. 2005; van Zelm et al. 2006; Vorechovsky et al. 1995; Warnatz et al. 2009) During screens for genetic variations that associate with CVID and/or IgAD, we discovered a CVID patient with a heterozygous premature stop codon in exon 8 of the *RAD50* gene (Q372X).(Offer et al.) Here we show that cells carrying this mutation have reduced levels of RAD50 and MRE11 and increased sensitivity to ionizing radiation. These data suggest that the recombination repair complex – RAD50/MRE11/NBS1 – can influence class switch recombination and that defects in this pathway may contribute to CVID or IgAD.

Materials and methods

CVID patient

The male patient, currently 30 years of age, was diagnosed with CVID at the age of 10, having had symptoms including recurrent *H. influenza* infections since the age of 8. At diagnosis he exhibited reduced levels of IgG (2 g/L; normal 7-15 g/L), IgA (0.2 g/L; normal 0.5-3.1 g/L),

and IgM (0.2 g/L; normal 0.3-2.0 g/L). Mutations in known CVID genes *TACI*,(Castigli et al. 2005; Pan-Hammarström et al. 2007; Salzer et al. 2005) *BAFF-R*,(Warnatz et al. 2009) *CD19*,(van Zelm et al. 2006) and *ICOS*(Grimbacher et al. 2003) were not detected by DNA sequencing (data not shown).

***RAD50* sequencing**

DNA was isolated from whole blood by the phenol:chloroform method at the Karolinska Institutet (Stockholm, Sweden). Primers were designed to amplify the 25 mRNA coding exons of the *RAD50* gene and sequencing was done by the Broad Institute Center for Genotyping and Analysis. Genetic variants were identified using the PolyPhred and PolyDhan programs. Primers were synthesized by IDT (Coralville, IA, USA) and sequences are available upon request.

RAD50-Q372X heterozygous cell lines

EBV-immortalized lymphoblast lines were generated at the Karolinska Institutet (Stockholm, Sweden). The HCT116-*RAD50*-c.1114C>T (Q372X) knock-in cell line and a wildtype knockin control line were generated using recombinant adeno-associated virus (rAAV)-assisted gene targeting, as previously described.(Ruis et al. 2008) Briefly, targeting arms were amplified from genomic DNA by PCR and cloned into pJET1.2 (CloneJet PCR kit, Fermentas). Site directed mutagenesis was used to introduce the *RAD50* c.1114C>T stop mutation and targeting arms were assembled into pNeDaKO.(Kohli et al. 2004) rAAV was produced in HEK-293T cells and used to transduce HCT116 cells. The resulting G418-resistant clones were screened for targeted insertion by PCR and confirmed by Southern blotting. The introduction of the mutation was confirmed by sequencing. Plasmids and cell lines used in these studies can be found in **Tables AIII-1 and AIII-2**.

Immunoblots and ionizing radiation sensitivity experiments

Cell extracts were prepared using RIPA buffer, fractionated by SDS PAGE, and transferred to nylon membrane for immunoblotting. The membranes were probed with antibodies to RAD50 (Novus), MRE11 (Genetex), or tubulin (Bethyl) followed by appropriate HRP-conjugated secondary antibodies (BioRad). Radiation sensitivity assays were conducted using a cesium irradiator for the indicated doses and assaying live cells by 7-AAD staining and flow cytometry (lymphoblasts) or crystal violet colony staining of plated dilutions (HCT116 cells).(Ruis et al. 2008)

Results

During screens for genetic variations that associate with CVID and/or IgAD, we discovered a CVID patient with a heterozygous premature stop codon in exon 8 of the *RAD50* gene (c.1114C>T; NM_005732, numbered from the first nucleotide of the start codon).(Offer et al.) This mutation converts glutamine 372 to a stop codon (Q372X). DNA sequences from available family members showed that allele was inherited from the father. The father had normal serum levels of all antibody isotypes at the time the patient was diagnosed (Materials and methods).

Prior studies have shown that all three components of the MRE11/RAD50/NBS1 (MRN) recombinase complex are subject to coordinated post-translation regulation.(Stewart et al. 1999) We therefore hypothesized that the Q372X mutation would result in 50% lower MRN levels and a diminished capacity to repair DNA double-strand breaks. To test this hypothesis, we used EBV to immortalize B cells from the patient and examined the resulting line by immunoblotting and ionizing radiation sensitivity experiments.

Immortalized lymphoblasts derived from this CVID patient exhibited an approximate 2-fold reduction in RAD50 protein levels in comparison to three similarly derived cell lines from

individuals with wild-type *RAD50* (**Figure AIII-1A**). MRE11 levels were also reduced, consistent with prior observations indicating that the molar ratios of each MRN component are regulated coordinately.(Stewart et al. 1999) Compared to a control EBV immortalized cell line, the Q372X cells were modestly, but significantly, sensitive to higher doses of ionizing radiation (**Figure AIII-1B**). This impairment was not as profound as that observed for cells derived from an ataxia telangiectasia patient with *ATM* mutations.

Since patient-derived cell lines are by definition non-isogenic and potentially variable, we asked if we could recapitulate the sensitivity to ionizing radiation by introducing the *RAD50* c.1114C>T mutation into the endogenous locus of an immortalized human cell line HCT116 (**Figure AIII-2A**). HCT116 is a near-diploid line that has been used extensively for gene targeting experiments, and particularly for studies of DNA metabolism proteins.(Ruis et al. 2008) Targeting events were identified by insertion-specific PCR products, and they were confirmed by Southern blotting with a flanking probe (Materials and methods). The specific incorporation of the c.1114C>T mutation was confirmed by sequencing insertion-specific PCR products. Relative to the control parental line, non-targeted sister clones, or targeted sister clones with wildtype *RAD50*, cells containing the heterozygous c.1114C>T mutation displayed an increased sensitivity to ionizing radiation (**Figure AIII-2B** and data not shown). The sensitivity was significant but not as pronounced as that for a DNA-PKcs-deficient cell line, which is defective for recombination by the nonhomologous end-joining pathway.(Ruis et al. 2008)

Discussion

In the context of a large genetic study, we recently identified a single Swedish CVID patient with RAD50-c.1114C>T transition mutation, predicted to encode a premature Q372X translational stop codon.(Offer et al.) This mutation was not found in 882 chromosomes from

other patients or 1900 control chromosomes, but it is noteworthy that other potentially detrimental variations were observed in *RAD50*, *MRE11*, and *NBS1*. These studies suggested that defects in the MRN recombination complex might have a functional consequence and potentially underlie a subset of CVID and/or IgAD cases.

Here, we report the first functional data from cell-based experiments. An EBV-immortalized RAD50-Q372X patient cell line showed a 2-fold reduction in RAD50 and MRE11 protein levels and a modest, but significant, sensitivity to ionizing radiation. To confirm the functional significance of RAD50-Q372X, we used rAAV-mediated gene targeting to render the human HCT116 cell line heterozygous for this mutation and, again, ionizing radiation sensitivity was observed. We conclude that cells carrying the RAD50-Q372X allele are haploinsufficient, incapable of wildtype levels of recombination repair.

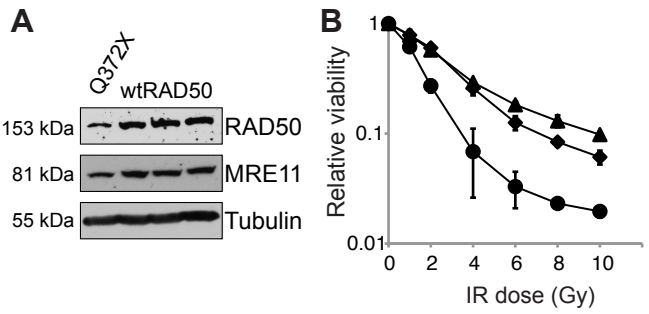
Recently, a RAD50-deficient patient was reported with an NBS-like phenotype.(Waltes et al. 2009) Interesting comparisons can be made between the heterozygous RAD50-Q372X allele characterized here and this other patient who carried biallelic *RAD50* mutations. Notably, our RAD50-Q372X patient exhibits markedly reduced serum IgG, IgA, and IgM antibody titers, while the biallelic RAD50 patient appeared immune-competent. Furthermore, our RAD50-Q372X patient does not exhibit any developmental abnormalities that would be characteristic of Nijmegen breakage syndrome. A likely explanation is that RAD50-Q372X causes a simple haploinsufficiency, a relatively mild defect in comparison to amino acid substitution mutations that may have a dominant-negative poisoning effect on the entire MRN recombinational repair complex.

A combination of detailed genetic and biochemical studies will be needed to further investigate the contribution of MRE11/RAD50/NBS1 pathway defects to immunodeficiency syndromes such as CVID and IgAD. The obvious necessity for these proteins in class switch recombination is compatible with a causal role. However, the fact that the father of the CVID

patient was asymptomatic but also heterozygous for the RAD50-Q372X allele suggests that many defects in this recombination protein complex will be asymptomatic (*i.e.*, incompletely penetrant). These observations are fully compatible with the complex etiology of CVID and IgAD, in which there are likely both common and specific causal genetic defects as well as clear modifying polymorphisms and potentially even environmental factors.(Cunningham-Rundles and Bodian 1999; Hammarström et al. 2000)

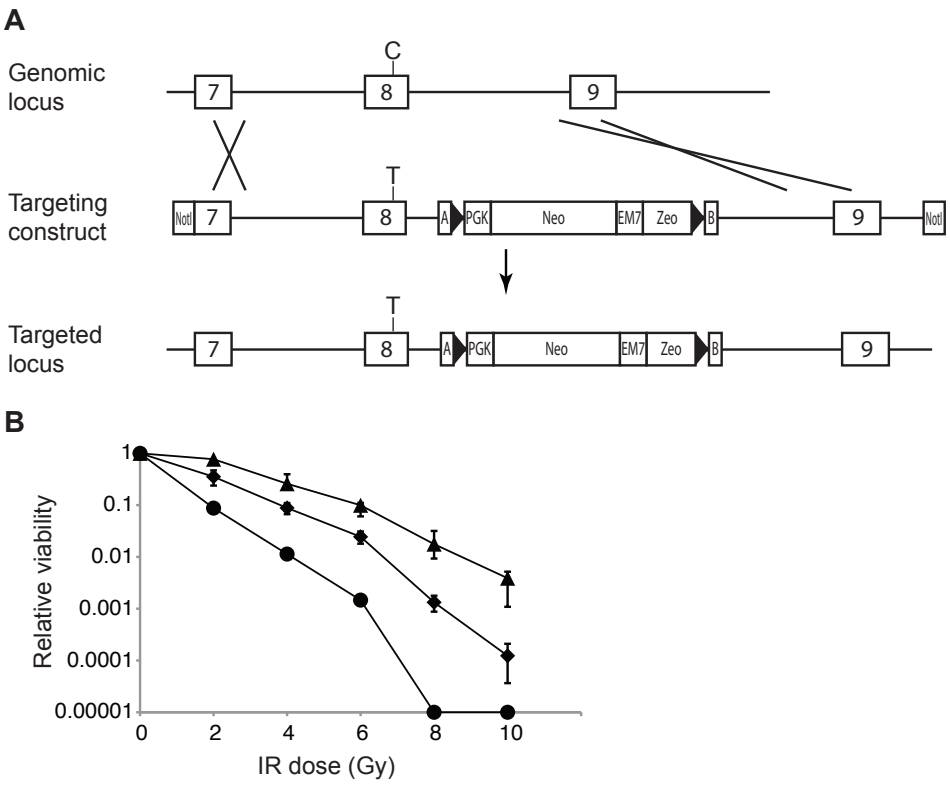
Acknowledgements

We thank E. Hendrickson for rAAV targeting vectors and control cell lines, Y. Wang for assistance with IR experiments, J. Petrini for DNA constructs, and the Broad Institute for DNA sequencing. This work was supported by grants from the National Institutes of Health (R21 AI079743 and R01 GM80437 to R.H. and U19 AI067152 to L.H.) and the Swedish Cancerfound (to Q.P.). S.M.O. was supported in part by a University of Minnesota Graduate School Doctoral Dissertation Fellowship and by the Curtis L. Carlson University of Minnesota/Karolinska Institutet Medical Research and Education Program Award.



Offer et al., Figure 1.

Figure AIII-1. Patient-derived RAD50-Q372X cells show reduced protein expression and increased sensitivity to ionizing radiation. (**A**) Immunoblot of lysates from lymphoblast cell lines derived from the CVID patient carrying RAD50-Q372X and three CVID patients with no mutations in RAD50, MRE11, or NBS1 (wildtype, wtRAD50). (**B**) Viability of patient derived cell lines after exposure to ionizing radiation. Lymphoblast cell lines from the RAD50-Q372X patient (diamonds) and a CVID patient with no mutations in RAD50, MRE11, or NBS1 (triangles) were exposed to the indicated doses of ionizing radiation and cell viability was assessed by 7-AAD staining and flow cytometry 10 days post-exposure. The Coriell cell line GM08436A (filled circles), containing heterozygous truncating mutations in ATM, was included as a positive control (Coriell Cell Repositories). Error bars denote the SEM of triplicate treatments.



Offer et al., Figure 2.

Figure AIII-2. Ionizing radiation sensitivity of HCT116 cells engineered to be heterozygous for the RAD50-Q372X allele. (**A**) The targeting strategy used to introduce the c.1114C>T mutation into the endogenous *RAD50* locus. Exons 7, 8, and 9 are depicted by boxes, and the position of c.1114C>T in exon 8 is marked. (**B**) Ionizing radiation sensitivity profiles G418 resistant targeted clones with c.1114C>T (diamonds) or a wildtype *RAD50* exon 8 (triangles). It is notable that both of these clones have properly integrated drug resistance cassette, and they only differ by whether or not the flanking C-to-T mutation was co-recombined. The parental HCT116 cell line shows an ionizing sensitivity profile superimposable with that of the wildtype exon 8 clone (data not shown). DNA-PKcs-deficient cells(Ruis et al. 2008) (circles) are extremely IR sensitive. Error bars represent the standard error of the mean sensitivity of four independent subclones (each assayed in duplicate).

Table AIII-1. Key plasmids generated/used in RAD50 studies

pRH	AKA	Plasmid name	Reference
99	maxi-pSO_57	pIRES-HYG	Clontech
2979	pSO699	pHRAD50-IRES-HYG	Appendix III
2980	pSO719	pHRAD50-IRES-HYG	Appendix III
2981	pSO700	pHRAD50-Q372X-IRES-HYG	Appendix III
2982	pSO720	pHRAD50-Q372X-IRES-HYG	Appendix III
2977	pSO_493	pBlueKS+::hRAD50	Appendix III
2978	pSO_496a	pBlueKS+::hRAD50-Q372X	Appendix III
2969	PSO_651	rAAV:: RAD50 targeting	Appendix III
2970	PSO_472	pCSII::MCS	Appendix III
2971	PSO_488(3)a	pCSII::hRAD50	Appendix III
2972	PSO_571	pCSII::hRAD50-Q372X	Appendix III
2983		pBabeHAER-hRAD50	J. Petrini

Table AIII-3. Key cell lines generated for RAD50 studies

Name	Clone	Cell line	Notes
CVID-8		EBV-immortalized lymphoblast	Int. radiosensitive from Hammarström Lab
CVID-51		EBV-immortalized lymphoblast	Int. radiosensitive from Hammarström Lab
CVID-69		EBV-immortalized lymphoblast	Non-radiosensitive from Hammarström Lab
CVID-72		EBV-immortalized lymphoblast	RAD50-Q372X patient from Hammarström Lab
GM08436A		EBV-immortalized lymphoblast	$ATM^{-/-}$ Corriel line from Hendrickson Lab
GM00130C		EBV-immortalized lymphoblast	$ATM^{+/+}$ Corriel line from Hendrickson Lab
cSO_131	parental	HCT116	Targeted, but NOT carrying Q372X mutation
cSO_131	8	HCT116	Targeted, but NOT carrying Q372X mutation
cSO_131	9	HCT116	Targeted, but NOT carrying Q372X mutation
cSO_131	10	HCT116	Targeted, but NOT carrying Q372X mutation
cSO_131	11	HCT116	Targeted, but NOT carrying Q372X mutation
cSO_9F	parental	HCT116	Q372X targeted
cSO_9F	7	HCT116	Q372X targeted
cSO_9F	8	HCT116	Q372X targeted
cSO_9F	11	HCT116	Q372X targeted
cSO_9F	12	HCT116	Q372X targeted
DNA-PKcs-	-	HCT116	DNA-PKcs null line from Hendrickson Lab
cSO_2f12	-	HCT116	Q372X mistarget
cSO_2e12	-	HCT116	Q372X mistarget
cSO_4D	-	HCT116	Q372X mistarget
cSO_2f1	-	HCT116	Q372X mistarget
cSO_4G	-	HCT116	Q372X mistarget
cSO_2a2	-	HCT116	Q372X mistarget
cSO_4G(2)	-	HCT116	Q372X mistarget

REFERENCES

Carvalho Neves Forte W., F. Ferreira De Carvalho Junior, N. Damaceno, F. Vidal Perez, C. Gonzales Lopes and R. A. Mastroti (2000). Evolution of IgA deficiency to IgG subclass deficiency and common variable immunodeficiency. Allergol Immunopathol (Madr) **28**(1): 18-20.

Castigli E., S. A. Wilson, L. Garibyan, R. Rachid, F. Bonilla, L. Schneider and R. S. Geha (2005). TACI is mutant in common variable immunodeficiency and IgA deficiency. Nat Genet **37**(8): 829-834.

Cunningham-Rundles C. and C. Bodian (1999). Common variable immunodeficiency: clinical and immunological features of 248 patients. Clin Immunol **92**(1): 34-48.

Espanol T., M. Catala, M. Hernandez, I. Caragol and J. M. Bertran (1996). Development of a common variable immunodeficiency in IgA-deficient patients. Clin Immunol Immunopathol **80**(3 Pt 1): 333-335.

Grimbacher B., A. Hutloff, M. Schlesier, E. Glocker, K. Warnatz, R. Drager, H. Eibel, B. Fischer, A. A. Schaffer, H. W. Mages, R. A. Kroczek and H. H. Peter (2003). Homozygous loss of ICOS is associated with adult-onset common variable immunodeficiency. Nat Immunol **4**(3): 261-268.

Hammarström L., I. Vorechovsky and D. Webster (2000). Selective IgA deficiency (SIgAD) and common variable immunodeficiency (CVID). Clin Exp Immunol **120**(2): 225-231.

Ishizaka A., M. Nakanishi, S. Yamada, Y. Sakiyama and S. Matsumoto (1989). Development of hypogammaglobulinaemia in a patient with common variable immunodeficiency. Eur J Pediatr **149**(3): 175-176.

Johnson M. L., L. G. Keeton, Z. B. Zhu, J. E. Volanakis, M. D. Cooper and H. W. Schroeder, Jr. (1997). Age-related changes in serum immunoglobulins in patients with familial IgA deficiency and common variable immunodeficiency (CVID). Clin Exp Immunol **108**(3): 477-483.

Kohli M., C. Rago, C. Lengauer, K. W. Kinzler and B. Vogelstein (2004). Facile methods for generating human somatic cell gene knockouts using recombinant adeno-associated viruses. Nucleic Acids Res **32**(1): e3.

Offer S. M., Q. Pan-Hammarström, L. Hammarström and R. S. Harris Unique DNA Repair Gene Variations Associate with the Primary Antibody Deficiency Syndromes IgAD and CVID. Hum Mutat **in review**.

Pan-Hammarström Q., U. Salzer, L. Du, J. Bjorkander, C. Cunningham-Rundles, D. L. Nelson, C. Bacchelli, H. B. Gaspar, S. Offer, T. W. Behrens, B. Grimbacher and L. Hammarström (2007). Reexamining the role of TACI coding variants in common variable immunodeficiency and selective IgA deficiency. Nat Genet **39**(4): 429-430.

Ruis B. L., K. R. Fattah and E. A. Hendrickson (2008). The catalytic subunit of DNA-dependent protein kinase regulates proliferation, telomere length, and genomic stability in human somatic cells. Mol Cell Biol **28**(20): 6182-6195.

Salzer U., H. M. Chapel, A. D. Webster, Q. Pan-Hammarström, A. Schmitt-Graeff, M. Schlesier, H. H. Peter, J. K. Rockstroh, P. Schneider, A. A. Schaffer, L. Hammarström and B. Grimbacher (2005). Mutations in TNFRSF13B encoding TACI are associated with common variable immunodeficiency in humans. Nat Genet **37**(8): 820-828.

Stewart G. S., R. S. Maser, T. Stankovic, D. A. Bressan, M. I. Kaplan, N. G. Jaspers, A. Raams, P. J. Byrd, J. H. Petrini and A. M. Taylor (1999). The DNA double-strand break repair gene hMRE11 is mutated in individuals with an ataxia-telangiectasia-like disorder. Cell **99**(6): 577-587.

van Zelm M. C., I. Reisli, M. van der Burg, D. Castano, C. J. van Noesel, M. J. van Tol, C. Woellner, B. Grimbacher, P. J. Patino, J. J. van Dongen and J. L. Franco (2006). An

antibody-deficiency syndrome due to mutations in the CD19 gene. N Engl J Med **354**(18): 1901-1912.

Vorechovsky I., H. Zetterquist, R. Paganelli, S. Koskinen, A. D. Webster, J. Bjorkander, C. I. Smith and L. Hammarström (1995). Family and linkage study of selective IgA deficiency and common variable immunodeficiency. Clin Immunol Immunopathol **77**(2): 185-192.

Waltes R., R. Kalb, M. Gatei, A. W. Kijas, M. Stumm, A. Sobeck, B. Wieland, R. Varon, Y. Lerenthal, M. F. Lavin, D. Schindler and T. Dork (2009). Human RAD50 deficiency in a Nijmegen breakage syndrome-like disorder. Am J Hum Genet **84**(5): 605-616.

Warnatz K., U. Salzer, M. Rizzi, B. Fischer, S. Gutenberger, J. Bohm, A. K. Kienzler, Q. Pan-Hammarström, L. Hammarström, M. Rakhmanov, M. Schlesier, B. Grimbacher, H. H. Peter and H. Eibel (2009). B-cell activating factor receptor deficiency is associated with an adult-onset antibody deficiency syndrome in humans. Proc Natl Acad Sci U S A.