Investigating Mechanisms of DNA Double Strand Break Joining of Switch Regions During IgH Class Switch Recombination

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citation</th>
<th>Panchakshari, Rohit. 2016. Investigating Mechanisms of DNA Double Strand Break Joining of Switch Regions During IgH Class Switch Recombination. Doctoral dissertation, Harvard University, Graduate School of Arts &amp; Sciences.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:33493369">http://nrs.harvard.edu/urn-3:HUL.InstRepos:33493369</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Investigating Mechanisms of DNA Double Strand Break joining of Switch regions during IgH Class Switch Recombination

A dissertation presented

by

Rohit Panchakshari

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Immunology

Harvard University
Cambridge, Massachusetts
April 2016
Investigating Mechanisms of DNA Double Strand Break joining of Switch regions during IgH Class Switch Recombination

Abstract

During B cell development, RAG endonuclease cleaves immunoglobulin heavy chain (IgH) V, D, and J gene segments and orchestrates their fusion as deletional events that assemble a V(D)J exon in the same transcriptional orientation as adjacent Cμ constant region exons. In mice, six additional sets of constant region exons (C_Hs) lie 100–200 kilobases downstream in the same transcriptional orientation as V(D)J and Cμ exons. Long repetitive switch (S) regions precede Cμ and downstream C_Hs. In mature B cells, class switch recombination (CSR) generates different antibody classes by replacing Cμ with a downstream C_H. Activation-induced cytidine deaminase (AID) initiates CSR by promoting deoxycytidine deamination lesions within S_μ and a downstream acceptor S-region; these lesions are converted into DNA double-strand breaks (DSBs) by general DNA repair factors which are then joined by end-joining pathways.

Productive CSR must occur in a deletional orientation by joining the upstream end of an S_μ DSB to the downstream end of an acceptor S-region DSB. However, the relative frequency of deletional to inversional CSR junctions has not been measured. Thus, whether orientation-specific joining is a programmed mechanistic feature of CSR as it is for V(D)J recombination and, if so, how this is achieved is unknown. To address this question, we adapt high-throughput genome-wide translocation sequencing (HTGTS) into a highly sensitive DSB end-joining assay and apply it to endogenous AID-initiated S-region in mouse B cells. We show that CSR is
programmed to occur in a productive deletional orientation and does so via an unprecedented mechanism that involves \textit{in cis} IgH organizational features in combination with frequent S-region DSBs initiated by AID. We further implicate ATM-kinase-dependent DSB-response (DSBR) factors including histone variant H2AX, 53BP1 and its associated effector protein Rif1 in enforcing this mechanism.

We go on to use HTGTS to study influence of different DSBR factor deficiencies on the structure of CSR junctions between AID-initiated DSBs in the 5' portion of the donor $S\mu$ region to those across the length of downstream acceptor S regions. Based on analyses of thousands of switch junctions, we find that absence of DSBR factors leads to varying increases in micro-homology (MH)-mediated junctions, with 53BP1-deficiency having the greatest increase. However, while translocation junctions between Cas-9/gRNA-induced DSB in \textit{c-myc} to AID-initiated S region DSBs in ATM- or 53BP1-deficient B cells show similar biases in MH-usage to those observed in the context CSR junctions, translocation junctions to other general DSBs genome-wide had no MH-usage increase in ATM-deficient cells and only a modest increase in 53BP1-deficient cells. We discuss these findings with respect to potential roles of AID-initiated DSBs in S regions to be especially prone to MH-usage potentially due to their increased resection along with their highly repetitive nature that provides abundant micro-homologous sequence.
TABLE OF CONTENTS

Abstract ........................................................................................................ iii

Table of Contents .......................................................................................... v

Attribution of Collaborator Contributions ................................................. vii

List of abbreviations ...................................................................................... viii

List of Figures ................................................................................................ x

Acknowledgements ......................................................................................... xi

INTRODUCTION ......................................................................................... 1 - 57

I.1 Thesis overview .......................................................................................... 1
I.2 B cell receptor / Antibody molecule .......................................................... 14
I.3 V(D)J Recombination in B cells ................................................................. 15
I.4 Secondary Immunoglobulin diversification: CSR ..................................... 18
I.5 Activation Induced Cytidine Deaminase (AID) ......................................... 25
I.6 Classical non-homologous end-joining (C-NHEJ) .................................... 27
I.7 C-NHEJ in V(D)J recombination ............................................................... 28
I.8 C-NHEJ in IgH CSR .................................................................................. 30
I.9 Alternative end-joining (A-EJ) in IgH CSR .............................................. 31
I.10 DNA Double Strand Break (DSB) response in CSR ................................. 33
I.11 LAM-HTGTS .......................................................................................... 40
I.12 References .............................................................................................. 43

CHAPTER 1: Orientation-specific joining of AID-initiated DNA breaks promotes
Antibody Class Switching ............................................................................. 58 - 92

1.1 Abstract ................................................................................................... 59
1.2 Introduction .............................................................................................. 59
1.3 Results ..................................................................................................... 61
1.4 Discussion .............................................................................................. 87
1.5 References .............................................................................................. 90

CHAPTER 2: Deficiencies for Dna Double-Strand Break response proteins promotes
microhomogy-biased IgH CSR junctions ...................................................... 93 - 124

2.1 Abstract ................................................................................................... 94
2.2 Introduction .............................................................................................. 94
2.3 Results ..................................................................................................... 98
TABLE OF CONTENTS (Continued)

2.4 Discussion .............................................................................................................. 117
2.5 References ............................................................................................................. 121

DISCUSSION ..................................................................................................................... 125 - 134

METHODS ......................................................................................................................... 135 – 155
M.1 Chapter 1 ................................................................................................................. 136 - 153
  M.1.1 Mice ...................................................................................................................... 136
  M.1.2 Plasmids and oligonucleotides ........................................................................... 136
  M.1.3 B cell culture, transduction and FACS analysis .................................................. 136
  M.1.4 Cell lines and nucleofection ............................................................................. 137
  M.1.5 DC–PCR ............................................................................................................... 138
  M.1.6 High-throughput genome-wide translocation sequencing (HTGTS) ................... 140
  M.1.7 Data analyses ...................................................................................................... 140
  M.1.8 S-region junction plotting .................................................................................. 142
  M.1.9 Calculation of joining orientation bias and acceptor S-region resection .......... 143
  M.1.10 Pipeline simulation for S region mappability .................................................... 144
  M.1.11 Normalization by “symmetrical joins” ............................................................... 146
  M.1.12 DNA-oligos ........................................................................................................ 148
  M.1.13 Coordinates used for plotting ......................................................................... 149
  M.1.14 Sequencing Data ............................................................................................... 150
  M.1.15 References ........................................................................................................ 151

M.2 Chapter 2 ..................................................................................................................... 153 - 156
  M.2.1 Material and methods ........................................................................................ 153
  M.2.2 Cell lines .............................................................................................................. 153
  M.2.3 Data analyses ...................................................................................................... 154
  M.2.4 Statistical analysis .............................................................................................. 154
  M.2.5 Sequencing Data ................................................................................................. 154
  M.2.6 References ........................................................................................................... 155
Attribution of collaborator contributions

Introduction: Rohit A. Panchakshari wrote this section along with conceptual guidance and discussion with Frederick W. Alt

Chapter 1: Junchao Dong, Rohit A. Panchakshari, Tingting Zhang, John P Manis and Frederick W. Alt designed the study; Junchao Dong, Rohit A. Panchakshari, Tingting Zhang, Jiazhi Hu and Sabrina Volpi performed experiments; Yui-ji Ho and Robin Meyers designed bioinformatics pipelines; Robin Meyers, Junchao Dong, Rohit A. Panchakshari and Zhou Du performed computational analyses of sequencing data; Junchao Dong, Rohit A. Panchakshari and Tingting Zhang contributed equally to this work.

Chapter 2: Junchao Dong, Rohit A. Panchakshari and Frederick W. Alt designed the study; Rohit A. Panchakshari and Junchao Dong, performed experiments; Yui-ji Ho and Robin Meyers designed bioinformatics pipelines; Robin Meyers, Junchao Dong, Rohit A. Panchakshari performed computational analyses of sequencing data; Rohit A. Panchakshari, Junchao Dong and Frederick W. Alt wrote the manuscript along with helpful discussions with Vipul Kumar (Alt lab), who also provided ATM−/−CH12F3 cell lines. Rohit A. Panchakshari and Junchao Dong contributed equally to this work.

Discussion: Rohit A. Panchakshari wrote this section along with helpful discussions with Junchao Dong and Frederick W. Alt
List of abbreviations

3C  chromosome configuration capture
3D  three-dimensional
53BP1 p53-binding protein 1
A-EJ alternative end-joining
AID activation induced cytidine deaminase
APE apurinic/apyrimidic endonuclease
ATM ataxia-telangiectasia mutated
BCR B cell receptor
BE break end
BER base excision repair
BRCA1 breast cancer gene 1
CRISPR clustered regularly interspaced short palindromic repeat
CSR class switch recombination
CtIP C-terminal binding protein (Ctbp)-interating protein
DC-PCR digestion circularization - polymerase chain reaction
DEL deletion
DNA-PKcs DNA-dependent protein kinase catalytic subunit
DSB double strand break
DSBR double strand break response
FISH fluorescence in situ hybridization
GC germinal center
HR homologous recombination
HTGTS high-throughput genome-wide translocations sequencing
Ig immunoglobulin
IgH immunoglobulin heavy
IgL immunoglobulin light
INV inversion
LAM-HTGTS linear amplification mediated - HTGTS
LIG4 DNA ligase 4
LPS lipopolysaccharide
MH microhomology
MMR mis-match repair
MRE11 meiotic recombination 11 homolog A
NBS1 Nijmegen breakage syndrome 1
NHEJ non-homologous end-joining
PARP1 poly-ADP-ribose polymerase 1
RAG recombinase activating gene
RPA replication protein A
RSS recombination signal sequence
SHM somatic hypermutaion
TAD topologically associated domain
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCR</td>
<td>T cell receptor</td>
</tr>
<tr>
<td>TDT</td>
<td>terminal deoxynucleotidyl transferase</td>
</tr>
<tr>
<td>UNG</td>
<td>uracil N glycosylase</td>
</tr>
<tr>
<td>V(D)J</td>
<td>variable-diversity-joining</td>
</tr>
<tr>
<td>XLF</td>
<td>XRCC4-like factor</td>
</tr>
<tr>
<td>XRCC4</td>
<td>X-ray repair cross complementing 4</td>
</tr>
</tbody>
</table>
List of Figures

Figure I.1: Structure of B cell receptor (BCR) / Antibody molecule. ................................................................. 14
Figure I.2: IgH constant region locus and secondary immune diversification reactions. ................................. 20
Figure I.3: Diagram illustrating two long-range joining events between donor and acceptor S region DSBs. ...... 23
Figure I.4: Repair pathways of AID mediated U:G lesion leading to different functional outcomes. .................. 25
Figure I.5: Sources of various DNA DSB damage and its repair mechanisms. ...................................................... 27
Figure I.6: Direct joining or microhomolgy (MH)-mediated joining during end-joining reactions. ..................... 28
Figure I.7: Diagram illustrating methodology of Linear amplification mediated – High Throughput Genome-wide Translocation Sequencing (LAM-HTGTS). ........................................................................................................ 42
Figure 1.1.1 S-region-dependent orientation-biased joining in CSR-stimulated B cells. ....................................... 62
Figure 1.1.1 Deletional CSR in in vitro activated B cells by DC-PCR; Sce1 DSBs within the Igh constant region locus in activated B cells join with orientation-independence. .................................................. 63
Figure 1.2.1 Genome-wide translocation junctions lack orientation bias; Statistical analyses for experimental replicates in Fig. 1.1 and 1.2; Orientation-biased joining between I-Sce1 break in place of S and AID-initiated S breaks in CH12F3 cells. ......................................................... 66
Figure 1.3.1 Joining between I-Sce1 break at S and AID-initiated S-region breaks in LPS-activated (lipopolysaccharide); ΔSµ2d/ΔSγ128d B cells and clustered I-Sce1 breaks in ΔSµ2/Iγ128d B cells in place of Sγ1; Inverted Sµ in CH12F3 cells support robust IgA CSR. .......................................................... 68
Figure 1.2 S-regions are not sufficient to promote orientation-biased CSR joining. ......................................... 71
Figure 1.3 Orientation-biased joining of AID-initiated endogenous S-region breaks. ........................................... 73
Figure 1.4.1 Orientation-biased joining between AID-initiated rSµ and downstream AID-initiated S-region breaks in anti-CD40/IL4-activated and LPS-activated Sµ-truncated B cells. ......................................................... 74
Figure 1.5.1 Orientation-biased joining between rSµ and AID-induced S DSBs in CSR-activated CH12F3 cells. 76
Figure 1.6.1 Level of junctions to downstream S-regions in wild-type and DSBR-deficient 5’Sµ HTGTS libraries correlate with CSR levels; 5’Sµ break site undergoes variable degrees of resection from stimulated DSBR-deficient B cells. ......................................................... 78
Figure 1.4 Mechanisitic roles of Igh organization and DSBR factors in deletional CSR. ................................. 80
Figure 1.7.1 Orientation-biased joining between rSµ and AID-induced S and Sc DSBs in wild-type, ATM-deficient, and 53BP1-deficient B cells. ......................................................................................... 82
Figure 1.8.1 Orientation-biased joining of I-Sce1 DSBs at Sγ1 to AID-induced S-region breaks in various DSBR-deficient backgrounds. ........................................................................................................ 83
Figure 1.9.1 Inhibition of resection in 53BP1-deficient B cells by ATMi (ATM inhibitor) does not rescue directional CSR joining to Sγ1. ......................................................... 85
Table 1.1 Statistical comparison for orientation bias and resection in Sγ1 and Sc junctions from wild-type and DSBR-deficient B cell libraries. ................................................................. 86
Figure 2.1 MH pattern of joining from rS μ to downstream S region DSBs in *in vitro* stimulated rS μ cells. …… 100
Figure 2.2 DSBR-deficient B cells show increased MH usage during S region DSB joining. ………………… 103
Table 2.1 Average length of MH ± s.e.m. from Sμ-Sγ1, Sμ-Sε and Sμ-Sα junctions. ……………………… 104
Figure 2.1.1 MH pattern in S-S junctions from primary SμαΔ B cells in WT, ATM or 53BP1-deficient background. ……………………………………………………………………………………… 105
Figure 2.2.2 Linear plot of S region DSB joining and MH pattern in activated WT, ATM- and 53BP1-deficient CH12F3 cells. …………………………………………………………………………………………. 106
Figure 2.3.3 MH pattern in joining ΔSγ12xIce to S region DSBs in WT, ATM or 53BP1-deficient B cells. …… 107
Figure 2.3 Translocation joining to S-region breaks in activated DSB response-deficient B cells are more prone to resection and MH-usage. …………………………………………………………………………………………………………… 109
Figure 2.4 MH-mediated repair does not affect deletion orientation-biased CSR. …………………………… 111
Figure 2.4.4 MH-usage is not affected regardless of S region sequence orientation. ………………………… 112
Table 2.2 Average length of MH ± s.e.m. …………………………………………………………………………………………… 113
Table 2.3 Information regarding replicates and number of junctions. ………………………………………………… 114
Table 2.4 Statistical comparisons of direct junctions between WT and various DSBR factor deficient background. ………………………………………………………………………………………………………. 116
Acknowledgements

Starting from an eleventh grade student aspiring about studying in United States, up till now when I am about to graduate from the Immunology PhD program at Harvard, this journey has been the most enriching and defining period of my life. I truly believe that I have been blessed to meet numerous people who have played pivotal roles in shaping me personally and professionally. Thus writing this acknowledgements section for my thesis is overwhelming but at the same time, a highly sought-after opportunity for expressing the utmost gratitude I feel towards my mentors, professors, family and friends.

I want to begin by thanking my mentor/guru Dr. Frederick W. Alt for giving me this tremendous opportunity to be in his lab for pursuing my thesis research. Fred leads by an example as an incredibly visionary scientist who has been a prolific contributor to the field of B cell biology. Throughout his career spanning multiple decades, he has always kept abreast with the technological innovations and has swiftly pioneered its use to identify and answer key questions in the field while maintaining his thorough scientific standards. His enthusiasm, passion and insightfulness towards science are infectious and inspirational. Fred has created a rigorous collaborative lab environment providing intellectual freedom to pursue science and nurture critical scientific thinking. He takes keen interest in scientific and professional development of all his trainees and provides timely advice in guiding the projects towards fruition. I have also learned a lot from his highly articulate and concise scientific writing style. I am deeply indebted towards him for training me under his tutelage and sincerely hope that I am able to integrate his teachings in my future scientific endeavors.

Apart from science I have had the good fortune of musically conversing with Fred, who is adept at playing banjo, guitar and also sings. Together with Fred and other colleagues in and outside the lab, I am part of the lab band called ‘Pogies’, where I play Tabla (Indian classical percussion instrument) with a few old Irish and Beatles songs. Such musical fusion experiments with Fred have also been very enjoyable.

During my rotation in Alt lab, I worked with a then postdoc Dr. Duane Wesemann, whom I would like to thank for an absolutely wonderful rotation experience. Duane since then has always been a great source for advice and encouragement. Richard Frock who is keen about technological advancements in conducting experiments and a wonderful friend, provided a lot of
advice. Similarly, Feilong Meng helped a lot with learning technical aspects such as performing genome-wide nuclear run-on sequencing (GRO-seq) assays. For two other projects, I also worked closely with Dr. Ming Tian, a senior scientist instrumental in intellectually guiding me in the lab. Ming’s lucid scientific thinking, precision in executing experiments, perceptive data interpretation and an unparalleled teaching ability have made a lasting impression on me and are qualities worth emulating. The project regarding directional joining of switch region DSBs during CSR, which formed the main crux of my thesis, has been a delightful collaboration with Junchao Dong. We worked closely on this project and I have enjoyed and learned a lot from collaborating with him, especially while having stimulating discussions about the results, proposing models and drafting the manuscripts.

Alt lab has also earned me great friendships. Vipul, Joyce, Sherry, Amelia and me were the graduate students in the lab and formed a nice team. This group of friends has been a great support when experiments were disheartening. Suvi Jain became a close friend right after joining the lab and has been very helpful and dependable in life inside as well as outside the lab.

Hwei-Ling Cheng, our lab administrator and a highly skilled bench scientist, taught me ES (embryonic stem) cell targeting and has been helpful in the lab. Pei-Yi Huang, our mice research technician, was very supportive with coordinating all the necessities while doing animal experiments. Judy Needham, Kristine King and Heather Hadeka’s help with all the official paperwork filing, organizing meetings, shipping etc. has been very valuable. I would also like to acknowledge all the advice, great humor and delicious food kindly provided by Connie Gee for all these years.

My undergraduate experience at Hampshire College has played a decisive role in my education; I extend my gratitude towards my professors Dr. Nancy Lowry, Dr. Merle Bruno and Dr. Christopher Jarvis. Chris was the chair of my advisory committee throughout my undergraduate years and was a terrific mentor who encouraged my penchant towards studying molecular biology and Immunology. Learning immunology in Fall 2006 from Dr. Richard Goldsby (Amherst College), was a life-changing experience that led me to choose this field for pursuing my graduate education. His advice and mentorship have been very critical in my development. I would like to acknowledge my computer science professor Dr. David Kulp (UMass, Amherst) for sparking my interest in Perl/Ruby coding and teaching me its vast applications in bioinformatics. My undergraduate thesis research was very exciting and
intellectually rewarding, and I sincerely thank Dr. Jesse Mager (UMass, Amherst) for having me in his lab. Jesse treated me like a graduate student, giving me freedom in terms of proposing hypotheses, designing and conducting my own experiments and problem solving. This helped me develop ability to troubleshoot and think critically about the science. He also taught me an essential aspect of research where success is determined by consistency in efforts, and perseverance is required to absorb sporadic setbacks in experiments and troubleshoot the protocols.

I did my undergrad summer research internships and also worked as a research technician after graduating in laboratory of Dr. Dipanjan Chowdhury at Dana Farber Cancer Institute. Being my first teacher for rigorous bench experiments, Dipanjan’s guidance has been crucial in beginning to develop my technical skill sets as well as a critical discerning mindset. I also would like to express my gratefulness towards my dissertation advisory committee members including Dr. Glenn Dranoff, Dr. Mike Carroll, Dr. Stephen Buratowski and Dr. John Manis for their valuable guidance and advice on my thesis research.

My family, mother (Shubhangi A. Panchakshari), father (Anil Y. Panchakshari), grandpa (Yashwant), grandma (Sindhu) and my brother (Aditya) have always been my greatest support and strength. My parents provided my brother and me the best of education and a great upbringing to make sure that we are not restrained in our aspirations. I have been very fortunate to get the love and affection of my grandpa and grandma who both have been very supportive and always proud of all my endeavors. This thesis would not have been possible without the support of family and no amount of written expression can do justice to my gratitude, love and respect towards them.

A person who is absolutely integral to my life is my loving wife Tejas Dani. She is my relentless source of encouragement and has stood by me through all the highs and lows. Tejas’s unconditional love and support has been crucial for my accomplishments. I am also blessed to have Deepa Dani and Girish Dani as my in-laws, who have always overwhelmed me with their affection and have been proud of all my achievements.

I am very fortunate to have a host family since I came to Hampshire College. Aaron Piziali, Kathryn Swaim and Thea-Francesca warmly welcomed me and have always showered upon me so much love and care that I found a home away from home. I am very thankful for their constant support throughout out my education.
Music (Tabla) has always provided me the inner strength and solace ever since I can remember. I am deeply grateful to all my gurus, Ustad Allarakha, Ustad Zakir Hussain, Pt. Surest Talwalkar, Pt. Kiran Deshpande, Pt. Bhanudas Pawar, Narendra Khot, Supreet Deshpande and Nitin Pawar for teaching me this art. I would like to acknowledge their indispensable role in my life. Supreet Deshpande and Narendra Khot have been a constant source of inspiration and guidance for learning how to pursue music ‘sadhana’ and science with equal rigor.
For my Aaji (Sindhu Yashwant Panchakshari), a loving grandmother!
INTRODUCTION
I.1 Thesis overview

The immunoglobulin (Ig) molecule is comprised of two identical pairs of immunoglobulin heavy chain (IgH) and light chain (either Igκ or Igλ, referred as IgL) [1]. The Ig molecule functions as B cell receptor (BCR) or is secreted as antibodies. The N-terminal portion of IgH and IgL chains have a variable amino acid sequence portion that contains the antigen-binding region of BCRs or antibodies. The IgH and IgL variable region is encoded in gene segments that are assembled de novo during B cell development by a programmed DNA rearrangement event termed V(D)J recombination [1]. The effector functions of an antibody molecule (e.g. complement activation, opsonization, antibody-dependent cell cytotoxicity (ADCC) etc.), are mediated by a set of different C-terminal "constant" region sequences of the IgH chain. In mouse, there are 8 different constant regions, each encoded by a different set of germline encoded constant region exons (C\text{H}x) that lie downstream of the V(D)J exon within the IgH locus [1].

B cell development creates a massive repertoire of BCRs providing ability to react against virtually any antigen; each individual B cell expresses a receptor with a unique specificity towards a foreign antigen [2, 3]. This is achieved in developing B cells initially in fetal liver and later in the bone marrow by a process of organized DNA recombination that assembles the V(D)J exon that encodes the Ig variable regions [3]. This process, termed V(D)J recombination, is initiated in progenitor B cells during which single germline encoded variable (V), diversity (D) and joining (J) coding segments are assembled together creating a combinatorially diverse repertoire of antigen binding variable region exons [3]. The recombination-activating genes 1 and 2 (RAG1 and RAG2) together form the lymphocyte-specific RAG endonuclease (RAG) which initiate the DNA cleavage during V(D)J
recombination [3]. RAG is a site-specific endonuclease which introduces two DNA double-strand breaks (DSBs) between coding segments and flanking V(D)J recombination signal sequences (RSSs), thus generating four broken ends containing a pair of coding ends and pair of signal ends (containing RSS) [3]. The RAG endonuclease complex owing it its synaptic function renders the joining phase of V(D)J recombination to be orientation-biased whereby, the two coding ends (containing V, D or J segments) are specifically joined with each other, and the signal ends (RSS containing break ends) are joined with each other, in both cases exclusively by the general cellular classical non-homologous end-joining (C-NHEJ) pathway of DSB repair [3, 4]. V(D)J recombination occurs exclusively in the G1 phase of the cell cycle, at least in part due to rapid degradation of RAG2 when cells exit G1 [4].

For initial expression of an IgH chain, transcription is initiated upstream of the assembled V(D)J exon and terminated downstream of the C\(\mu\) exons (first set of C\(\mu\) exons) resulting in a spliced transcript encoding the \(\mu\) IgH protein [5]. Similarly, for IgL chain expression, transcription of assembled VJ exon (IgL chains do not use D segments) through the downstream IgL constant region exon leads to expression of the IgL protein [5]. Pairing of the IgH and IgL chains leads to the formation of IgM BCR which is expressed on the surface of the resulting IgM\(^+\) B cell. Surface IgM\(^+\) B cells exit bone marrow to secondary lymphoid organs e.g. spleen, lymph nodes, where upon encountering cognate antigen and activation (by cytokines, T-cells etc.), undergo two additional Ig diversification processes, namely IgH class switch recombination (CSR) and IgH and IgL variable region somatic hypermutation (SHM) [6].

SHM occurs in specialized lymphoid structures called germinal centers (GCs), where IgH and IgL V(D)J exons accumulate point mutations as well as a low frequency small deletions and insertions [6]. In the GC reaction, B cells undergo rounds of SHM along with clonal expansion
and affinity-based selection which leads to positive selection of B cells that bear BCRs with increased antigen-binding affinity for their cognate antigen; in this manner SHM leads to affinity maturation of the antibody [6]. CSR changes antibody effector functions by replacing the Cμ exons with a different set of CΗ exons that lie 100-200kb downstream of Cμ in the IgΗ locus [1]. Both CSR and SHM are initiated by activation-induced cytidine deaminase (AID), a ssDNA deoxycytidine deaminating enzyme [1]. The resultant AID-initiated U:G lesion is processed by general DNA repair machinery of base excision repair (BER) and mismatch repair (MMR) pathways that, instead of normal repair, are tailored to introduce mutations or DSBs during SHM and CSR respectively [6].

The mouse IgΗ region contains of eight sets of constant CΗ genes: 5’--VDJ--Cμ--Cδ--Cγ3--Cγ1--Cγ2b--Cγ2a--Cε--Cα--3’, spanning ~200 kb at the telomeric end of mouse chr12. Each germline CΗ gene is organized into an individual transcription unit composed of a 5’ promoter, a noncoding intervening (I) exon, a switch (S) region, and a set of coding CΗ exons [1]. S regions are long repetitive DNA sequences that can vary from 4 kb to 12 kb in length and consist of tandem repetitive G-rich sequences (e.g. TGGGG, GGGGT, GGGCT, GAGCT) on the non-template (coding) strand [1]. Sμ, Sα, and Sε are comprised of pentameric repeats and are enriched in GAGCT motifs, with the AGCT palindrome representing a canonical RGYW (or WRC) AID hotspot motif. Sγ1, the largest S region (~12 kb) and similar to other Sγ sequences, contains multiple 49bp repeats that are rich in RGYW motifs (including AGCT) [1]. In vivo, CSR can be stimulated through both T cell dependent and T cell independent pathways, which can be mimicked in vitro by activating B cells with α-CD40 (mimicking T cell help given via CD40L) or bacterial lipopolysaccharide (LPS) combined with cytokines such as IL-4, TGF-β [1]. AID initiates double strand breaks (DSBs) in donor Sμ and downstream acceptor S region that
then undergo long-range deletional recombination repair resulting in CSR to produce different class of antibodies with different $C_H$ regions and thus different effector functions. Different combinations of activators and cytokines direct CSR to particular S regions by modulating activation of germline transcription which in still not fully characterized ways recruits AID [1].

S region DSBs generated in donor and acceptor S regions have various joining outcomes. They can be joined locally within the S region leading to intra-switch recombination (ISR) and internal deletions, or joining could occur in cis over relatively long chromosomal distances (100 to 200 kb) that leads to productive CSR by deletional recombination, where the upstream end of an $S_\mu$ DSB is joined to the downstream end of an acceptor S region DSB. However, long-range joining could also yield non-productive "inversional" CSR that inverts the intervening sequence between donor and acceptor DSB. Unlike RAG endonuclease in V(D)J recombination [3], AID during IgH CSR has not been shown to possess any synaptic activity to shepherd the donor S and acceptor S region DSB ends to join in a biased manner. Although CSR is commonly described as a deletional recombination, the relative frequency of deletional to inversional CSR junctions has not been measured. In this regard, inversional CSR joining has been found in transformed B cells [7-10], which also led to proposition that CSR unlike V(D)J might be an orientation-unbiased process [11]. Thus, whether deletion orientation-specific joining is a programmed mechanistic feature of CSR as it is for V(D)J recombination and, if so, how this is achieved is an unanswered important basic question.

AID-initiated donor and acceptor S region DSBs have to undergo long-range deletional end-joining to complete CSR. DSBs are repaired by two major pathways: homologous recombination (HR) which uses homologous sequence as template for error-free repair or by classical non-homologous end-joining (C-NHEJ) which repairs the break ends with no or
minimal processing. The C-NHEJ pathway operates throughout the cell cycle but predominates during G0 / G1 phase when intact homologous template (sister chromatid) is not available [12]. There are seven C-NHEJ pathway members including Ku70, Ku80 X-ray repair cross-complementing protein 4 (XRCC4), DNA ligase 4 (Lig4), DNA-dependent protein kinase, catalytic subunit (DNA-PKcs), Artemis, and XRCC4-like factor (XLF) [12]. Ku70, Ku80, XRCC4, and Lig4 are considered “core” C-NHEJ factors that are absolutely required for all C-NHEJ reactions including V(D)J recombination [12, 13]. C-NHEJ mediated DSB repair could result in generation of “direct” joins, which involves ligating blunt DNA ends (DSB ends either generated as blunt ends or blunted after end-processing) or, end-joining could proceed via pairing of short stretches (usually 1 bp to 4 bp) of complementary sequence (sticky ends, “micro-homology”, MH) present at the terminal of two breaks ends, or exposed in the single-strand overhangs up on undergoing end-processing [12]. Consistent with C-NHEJ mediated repair [12], donor and acceptor switch junctions during CSR displayed ~30-40% direct joins, while the remaining junctions utilized short micro-homology (MH) up to 1-4bp [14-17]. Correspondingly, C-NHEJ factors play important role in mediating joining of S regions during CSR as shown in B cells deficient any of the four core C-NHEJ factors, where CSR is substantially impaired with up to 50% reduction compared to WT levels [13]. In addition, core C-NHEJ-deficient activated B cells also displayed significant number of unrepaired IgH breaks along with IgH locus derived chromosomal breaks and translocations [16, 18]. However, presence of substantial CSR levels (up to 50% of WT levels) in various C-NHEJ factors deficient backgrounds indicated that alternative end-joining (A-EJ) pathways can also substantially contribute towards CSR joining, at least in the absence of C-NHEJ. Analyses of the S junctions cloned from B cells deficient in various core C-NHEJ factors revealed perturbed
junction profile, wherein, direct joins were almost non-existent and nearly all CSR junctions displayed MH-biased joining [15-17, 19]; a fraction of these MHs in CSR junctions were also longer than those typically associated with C-NHEJ [15]. Such long MH-mediated junctions are often considered a feature of A-EJ pathways.

Mammalian cells have evolved an elaborate response machinery to a DSB which broadly can be described as comprised of three stages: DSB sensing, signal transduction/amplification to recruit repair factors and effector proteins-mediated repair [20]. A heterotrimeric complex composed of the meiotic recombination 11 homolog A (Mre11)/Rad50/Nijmegen breakage syndrome 1 (Nbs1) up on detecting a DSB activates the ataxia telangiectasia-mutated (ATM) serine/threonine kinase [21]. Up on undergoing activation via DSBs, ATM initiates the Double Stand Break Response (DSBR) by phosphorylating a set of proteins that includes histone variant H2AX, mediator of DNA damage checkpoint 1 (MDC1) and the p53-binding protein 1 (53BP1), which form large foci extending up to megabases along chromatin flanking DSB [20, 22]. Apart from providing docking sites for repair protein complexes, formation of these ATM-dependent foci have also been proposed to tether DSB ends for re-joining via C-NHEJ [22, 23]. ATM also directly activates cell-cycle checkpoints in response to DSBs, including p53-dependent G1/S checkpoint that halts the cell cycle to allow DSB repair, thus preventing replication of damaged DNA [22, 23]. ATM-dependent DSBR factors have also been implicated in channeling repair of DSBs (including AID-initiated S region DSBs) into the C-NHEJ pathway versus HR or A-EJ by antagonizing 5’-3’ resection of DNA ends that results in generation of 3’ overhangs [13, 24-28]. In this regard, resection of S region DSBs due to the repetitive nature of S sequence could provide robust substrates for MH-mediated joining.
ATM-dependent DSBR is activated during IgH CSR in response to the S region DSBs mediated by AID; in this regard, Nbs1 and phosphorylated histone H2AX (γ-H2AX) have been shown to co-accumulate on the IgH locus mostly in G1 phase in response to AID-mediated breaks during class switching [29]. Correspondingly, deficiencies for DSBR factors including ATM, H2AX, or MDC1 lead to reduced CSR ranging from 30% to 50% of WT levels [30-33]. In addition, B cells deficient for these factors accumulate substantial levels of AID-dependent IgH locus breaks upon stimulation for CSR, indicating a role for DSBR factors in end joining during CSR [34]. However, of all the DSBR factors, 53BP1 deficiency most severely impairs CSR with residual switching observed at ~5-10% of WT CSR levels [35, 36]. Similar to ATM-, H2AX-deficiency and classical non-homologous end-joining (C-NHEJ)-deficiency, 53BP1-deficiency also leads to substantial unrepaired AID-dependent IgH breaks and translocations in activated B cells [16, 18, 34], but in contrast, leads to little instability elsewhere in the genome [34]. These findings led to the hypothesis of a specialized role for 53BP1 in AID-dependent S region DSB repair. In context of CSR, several other functions of 53BP1 have been proposed, including preventing resection of DSB ends [24, 27, 28, 37, 38] and influencing repair pathway choice between C-NHEJ and alternative end-joining (A-EJ) [24]. With respect to protecting break ends from resection, ATM-dependent phosphorylation of 53BP1 recruits Rif1, an effector protein that inhibits resection of DSB ends [39, 40]. Accordingly, Rif1-deficiency in B cells results in a substantial decrease in CSR, however it is not as severe as that associated with 53BP1-deficiency [41-43].

DSBs and their joining mechanisms are of considerable interest given the important role they play during adaptive immune system development and their potentially pathological outcomes in the case of illegitimate repair. In this regard, Alt lab first developed high-throughput
genome-wide translocation sequencing (HTGTS) to identify genome-wide DSBs based on the ability of genome-wide DSBs to translocate to a fixed "bait" DSB generated by the yeast I-SceI meganuclease after cleavage of its 18-bp recognition site inserted in the c-myc gene on chr15 [44]. Correspondingly, this method has evolved to become a landmark technique where both a variety of different types of ectopic bait DSBs (e.g. Cas9-gRNA DSBs) or endogenous DSBs (e.g. RAG or AID-initiated DSBs) have been used as bait to identify various recurrent DSBs or recurrent classes of DSBs in different mouse and human cell types [45-51]. We have adapted a more recently enhanced version, linear amplification mediated version of HTGTS (LAM-HTGTS) [48, 49, 52] into an assay that allows us to track joining of endogenous AID-initiated DSBs in 5’ portion of Sµ in activated B cells to DSBs that occur across the entire length of downstream S regions [48]. By applying this assay in B cells deficient for various ATM-dependent DSBR factors, we provide several insights into mechanism of CSR, including, nucleotide-resolution quantification of resection of endogenous S region DSBs, maintenance of deletion orientation-specific biased joining of donor and acceptor S regions. This assay also offers an unparalleled sensitivity and magnitude for analyzing thousands of unique S-S junctions, providing substantial advancement in understanding roles of ATM-dependent DSBR factors in end-joining during CSR.

The first chapter in the thesis presents our investigation of mechanism underlying the generation of a productive CSR event, which must occur in a deletional orientation by joining the upstream end of an Sµ DSB to the downstream end of an acceptor S-region DSB [3]. To calculate the relative frequency of deletional to inversional CSR junctions, we adapted HTGTS into a highly sensitive DSB end-joining assay and have applied it to endogenous AID-initiated S-region DSBs in activated mouse B cells. We show that CSR is programmed to occur in a
productive deletional orientation such that AID-initiated donor Sμ DSBs join acceptor S (including γ, ε and α) DSBs in preferentially (more than 90%) deletional orientation. This peculiar aspect is unique to joining of S region DSBs, as joining between two non-AID-initiated DSBs in the same locale (introduced by ISceI meganuclease) showed orientation-unbiased joining. We show that the IgH locus in cis configuration plays a role in mediating this orientation-biased CSR joining, as translocations of a bait DSB in c-myc locus on chr15 joined S region DSBs in chr12 IgH locus in an orientation-independent manner. We further found that S region transcription dependent chromatin structures like R-loops and/or the transcripts themselves do not play a role in maintaining CSR joining bias. However, given the requirement of S region DSBs to observe orientation-biased joining, the possibility of AID-associated origin of S region DSBs features including processing of AID-initiated U:G lesions, high-frequency DSBs and/or potential roles for AID/associated complexes remain to be tested.

We also applied our HTGTS technology of using donor 5’Sμ DSBs as bait in stimulated B cells deficient for various ATM-dependent DSB response factors (including ATM, H2AX, 53BP1 and Rif1). These studies revealed that deficiency for each of these different ATM-dependent-DSBR factors deficiency led to variable loss of orientation-biased joining in CSR, which roughly correlated with the amount of increased S region DSBs resection observed in that particular background. The residual CSR in 53BP1−/− B cells, showed almost orientation-independent joining. We have proposed a working model for eliciting deletional orientation-specific joining in CSR, in which, the organization of IgH locus within topologically associated domains ("TADs") fosters frequent S region synapsis [3,50,51] via Langevin motion [28,52]. Within IgH TAD, we postulate additional unknown IgH-specific organizational features, which could play a role in mediating S synapsis favorable for deletional joining. In addition, S
regions owing to their ability to undergo AID-initiated and/or multiple frequent DSBs, are also necessary for promoting deletional orientation-specific joining.

ATM-dependent DSBR functions potentially by tethering un-synapsed S region break ends, which promotes efficient rejoining and allows subsequent AID-initiated breakage to undergo joining to a synapsed S region. In absence of such tethering, S region break ends could potentially separate as chromosomal breaks that could frequently translocate with orientation-independence to other S region break ends within the TAD [3]. DSBR factors also prevent long end-resections that could cause S region break ends to be sequestered in resection complexes, preventing synapsis with other S region break ends and/or diminishing ability to be joined by C-NHEJ pathway. 53BP1-deficiency impairs both tethering for rejoining and activates resection of un-joined ends by failure to recruit downstream effector Rif1, leading to extreme resections and the most severe impairment of CSR and orientation-dependent joining; in addition, we do not rule out a putative specialized role for 53BP1 in stabilizing S regions synapsis. As common and unique outcomes of 53BP1-deficiency prominently affect both donor and acceptor S regions, they would be multiplicative and, thereby, explain the profound impact of 53BP1-deficiency on CSR.

In chapter 2, we describe use of HTGTS to study the influence of different DSBR factor deficiencies on the structure of CSR junctions between AID-initiated DSBs in the 5' portion of the donor Sμ region to those across the length of downstream acceptor S regions. The ATM-DSBR has been implicated in preserving integrity of DSB ends by antagonizing the 5’-3’ resection of DNA generates long 3’ overhangs, hence promoting repair via the C-NHEJ pathway that can only join DSBs with blunt ends or short overhangs (1-4 bp) [23-27]. In this regard resection of S region DSBs, due to the repetitive nature of S region sequences, also could provide
longer MHs, which are substrates for A-EJ. In this regard, CSR junctions from B cells-deficient for various core C-NHEJ factors consistently show mostly MH-mediated CSR junctions [15-17, 19, 53]. Given potential limitations of conventional CSR junction analysis methods for generating large amounts of data of recovering junctions from the repetitive S region core where most CSR occurs, we have used HTGTS to investigate in greater depth mechanisms by which S regions DSBs are end-joined during CSR.

With HTGTS, we have identified tens-of-thousands of unique S-S junctions, as opposed to the hundreds usually characterized in given experiments by conventional methods. In contrast to some prior conclusions, we found that absence of DSBR factors leads to varying increases in micro-homology (MH)-mediated switch junctions during CSR, with 53BP1-deficiency having the greatest increase. We also found a positive correlation between long resections of S region DSBs and MH-usage in CSR junctions, with residual CSR junctions in 53BP1 deficient B cells having the greatest resection, and displaying most severely reduced direct joining frequency and, correspondingly, the greatest frequency of MH-based junctions. Notably inhibiting long resections of S region DSBs in 53BP1<sup>−/−</sup> cells via treatment with an ATM-inhibitor substantially rescued direct CSR junctions with direct joins and caused a concomitant decrease in the frequency of MH-biased junctions to ATM-deficient levels. We note, however, that ATM deficiency alone leads to lower levels of increased S region DSBs ‘long’ resection, modestly decreased direct joining and associated increased MH-biased joining. We discuss these findings with respect to their implications for the impact of potential shorter resections within S regions on MH-mediated end-joining during CSR. Finally, by examining MH usage in deletional and inverted CSR joins from WT and 53BP1-deficient cells, we argue against a major role for MH-based joining mechanisms in contributing towards orientation-specific CSR joining.
While translocation junctions between Cas-9/gRNA-induced DSB in c-myc to AID-initiated S region DSBs in ATM- or 53BP1-deficient B cells show similar biases in MH-usage to those observed in the context of CSR junctions, translocation junctions to other general DSBs genome-wide had no substantial MH-usage increase in ATM-deficient cells and only a modest increase in 53BP1-deficient cells. We discuss these findings with respect to potential features of AID-initiated S regions DSB that may make them especially prone to MH-usage. Such feature may include an increased susceptibility to undergo resection in a DSBR deficient background along with the highly repetitive nature of S regions that provides abundant micro-homologies for CSR junctions.
I.2  B cell receptor / Antibody molecule

The antibody molecule is comprised of two immunoglobulin (Ig) heavy (IgH) chains (dark blue) and two light (IgL) chains (light blue). The rectangles represent Ig domains that constitute the structural units of the immunoglobulin heavy and light chains. The de novo assembled variable regions which bind the antigen consist of V\_H, D\_H and J\_H gene segments on IgH chain and V\_L and J\_L gene segments on the IgL. Complementarity-determining regions (CDRs) in variable region are indicated in dashed red boxes. The constant region of IgH chain determines the class and effector function of the antibody molecule.

The immunoglobulin (Ig) molecule is comprised of two identical pairs of immunoglobulin heavy chain (IgH) and light chain (either Igκ or Igλ, referred as IgL). The Ig molecule functions as B cell receptor (BCR) or is secreted as antibodies [1]. The N-terminal portion of IgH and IgL chains have a variable amino acid sequence portion that contains the antigen-binding region of BCRs or antibodies. The IgH and IgL variable region is encoded in gene segments that are assembled de novo during B cell development by a programmed DNA
rearrangement event termed V(D)J recombination [1]. The effector functions of the Ig are mediated by a set of different C-terminal "constant" region sequences of the IgH chain [1] (Fig I.1).

I.3 V(D)J Recombination in B cells

Lymphocytes including B and T cells can launch a specific immune response against virtually any antigen by generating a nearly infinite diversity of antigen receptors [2, 3], each individual B and T cell expresses a receptor with a unique specificity towards a foreign antigen. This is achieved in developing B cells initially in fetal liver and later in the bone marrow by a process of organized DNA recombination that assembles the V(D)J exon that encodes the Ig variable regions. This process, termed V(D)J recombination, is initiated in progenitor B cells during which single germline encoded variable (V), diversity (D) and joining (J) coding segments are assembled together creating a combinatorially diverse repertoire of antigen binding variable region exons [3].

The recombination-activating genes 1 and 2 (RAG1 and RAG2) together form the RAG endonuclease (RAG) which initiate the DNA cleavage during V(D)J recombination [3]. RAG is a site-specific endonuclease which introduces two DNA double-strand breaks (DSBs) in recombination signal sequences (RSS) lying adjacent to the two coding segments [3]. Canonical RSS consist of a highly conserved heptamer and nonamer sequence separated by a relatively non-conserved spacer of either 12 or 23 base pairs (bp) sequence (referred to as ‘‘12RSSs’’ and ‘‘23RSSs’’ respectively) [3]. RAG mediated joining of two coding segments follows the ‘‘12/23 rule’’; i.e. recombination between two coding segments occurs when they are flanked by ‘‘complementary’’ 12- and 23RSSs respectively [3]. This rule ensures the formation of a full V(D)J exon as in the IgH locus, V_{Hs} segments are flanked by 23RSSs, D_{Hs} on both sides by
12RSSs, and J_Hs by 23RSSs, preventing direct V_H-to-J_H joining [54]. After initiating DSBs at the border of two coding segments, the RAG proteins transiently retain the DNA ends in a post-cleavage synaptic complex [3] containing a pair of hairpin structure sealed coding ends (CEs, containing V, D or J segments) and a pair of blunt-ended signal ends (SEs, breaks ends containing RSSs). Within the post-cleavage complexes, there is an inherent directionality for joining, whereby, the coding end is specifically joined with the other coding end, and signal end joined with signal end [3, 54]. Hairpin sealed CEs are opened and processed before joining to yield coding joins, whereas the blunt SEs are joined directly to form direct signal joins [4]. The processing of CEs before joining expands the antigen receptor repertoire by diversification of V(D)J junctions [4, 55]. The end processing and joining phase of V(D)J recombination is exclusively mediated by ubiquitously expressed classical non-homologous end-joining (C-NHEJ) factors (described in later section) [4, 56, 57]. The combinatorial diversity arising from the assembly of numerous V, D, and J segments, and in addition the junctional diversity that arises from processing and joining of the coding segments, generates a massive repertoire of Ig variable region exons [2, 3].

IgH V(D)J recombination is “ordered” such that D_H-to-J_H joining occurs before V_H to DJ_H assembly. Developing B cells undergo D_H to J_H recombination on both IgH alleles first, and then join V_H segment to pre-rearranged DJ_H segments one at a time [58, 59]. If the first V_H to DJ_H rearrangement yields an in-frame V(D)J exons (productive), the resulting μ heavy chain protein generates a “feedback” signal to block V_H to DJ_H recombination on the other DJ_H rearranged allele. This feedback mechanism for ensuring mono-allelic productive V(D)J variable exon generation renders mono-specificity to B cells, this phenomenon is called “allelic exclusion” [60]. If the first V_H to DJ_H rearrangement produces an out-of-frame (non-productive)
variable exon, then the cell moves on to attempt productive recombination of V_H segment to the DJ_H on second allele. Expression of a productive μ heavy chain promotes progression of development to the pre-B stage where κ or λ light chain recombination occurs [54, 59]. V(D)J recombination is “stage specific” whereby IgH recombination yielding productive V(D)J variable exon (at pro-B cell stage) precedes IgL recombination that generates productive κ or λ light chain (pre-B cell stage) [9]. In addition, V(D)J recombination is also “lineage specific” [59] whereby, IgH locus undergoes productive recombination selectively in B cells and, similarly, T cell receptor (TCR) locus undergoes recombination specifically in T cells.

For initial expression of an IgH chain, transcription is initiated upstream of the assembled V(D)J exon and terminated downstream of the C_μ exons (first set of constant region (C_H) exons) resulting in a spliced transcript encoding the μ IgH protein [5]. Similarly, for IgL chain expression, transcription of assembled VJ exon (IgL chains do not use D segments) through the downstream IgL constant region exon leads to expression of the IgL protein [5]. Pairing of the IgH and IgL chains leads to the formation of IgM BCR which is expressed on the surface of the resulting mature naïve IgM+ B cells that now exit bone marrow to go to secondary lymphoid organs.
I.4 Secondary Immunoglobulin diversification: Class Switch Recombination (CSR)

Mature naïve B cells in the secondary lymphoid organs like spleen, lymph nodes, gut-associated lymphoid tissue (GALT) etc., up on encountering cognate antigen and activation (by cytokines, T-cells etc.), undergo two additional immunoglobulin (Ig) diversification processes, namely IgH class switch recombination (CSR) and IgH and IgL variable region Somatic Hypermutation (SHM). Both CSR and SHM are initiated by AID, a small 24-kDa protein discovered in a subtractive hybridization screen for genes that are upregulated in a cytokine stimulated B-cell line (CH12F3) undergoing CSR [61]. Soon it was shown that targeted deletion of AID in mice or mutations of the gene found in hyper-IgM patients led to a complete block in both CSR and SHM [62, 63].

SHM occurs in specialized lymphoid structures called germinal centers (GCs) [64], where AID deaminates deoxycytidine to convert in to uracil in IgH and IgL V(D)J exons, resultant U:G lesions in the DNA are processed by specific repair pathways (discussed below) leading to point mutations as well as a low frequency small deletions and insertions [6, 65, 66]. The mutation frequency over the V(D)J recombination assembled variable region exon on both IgH and IgL chains is approximately $10^{-3}$ mutations/bp per generation [64, 67]. In the GC reaction, B cells undergo rounds of SHM along with clonal expansion and affinity-based selection which leads to positive selection of B cells that bear BCRs with increased antigen-binding affinity for their cognate antigen; in this manner SHM leads to affinity maturation of the antibody [64] (Fig. I.2 A,B).

In order to elicit various effector functions (complement activation, neutralization, opsonization, antibody-dependent cell cytotoxicity (ADCC) etc.) of Ig molecule, IgH locus needs to undergo CSR, a long-range deletion recombination process which replaces Cµ constant exons
juxtaposed to VDJ exon with a set of downstream constant exons $C_H (\gamma, \varepsilon$ or $\alpha)$ [1]. This allows a B cell expressing IgM antibody to switch to production of either IgG, IgE or IgA antibody. The mouse IgH region consists of eight sets of constant $C_H$ genes: $5'-VDJ-C_\mu-C_\delta-C_\gamma3-C_\gamma1-C_\gamma2b-C_\gamma2a-C_\varepsilon-C_\alpha-3'$, spanning up to ~200 kb at the telomeric end of mouse chr12. There are B cell specific enhancers in IgH locus, namely, intronic $E_\mu$ enhancer ($iE_\mu$) which lies between the $J_H$ segments and the $C_\mu$ exons, and a 30 kb enhancer region termed as 3’ regulatory region (“IgH 3’RR”) lying downstream of the $C_\alpha$ exons [5, 68]. Each germline $C_H$ gene is organized into an individual transcription unit composed of a 5’ promoter, a noncoding intervening (I) exon, switch (S) intron region and coding $C_H$ exons [1]. When B cells are stimulated by a certain combination of cytokines and activators, germline non-coding transcription is initiated from a specific I promoter proceeding through the S region and terminated/polyadenylated downstream of the $C_H$ exons [1] (Fig. 1.2 A,C). In vivo, CSR can be stimulated through both T cell dependent and T cell independent pathways, which can be mimicked in vitro by activating B cells with $\alpha$-CD40 (mimicking T cell help given via CD40L) or bacterial lipopolysaccharide (LPS) combined with cytokines such as IL-4 and TGF-β [1]. Different combinations of activators and cytokines direct CSR to particular S regions by modulating activation of germline transcription of respective constant genes [1].

S regions are long repetitive DNA sequences that can vary from 4 kb to 12 kb in length and consist of tandem repetitive G-rich sequences (e.g. TGGGG, GGGGT, GGGCT, GAGCT) on the non-template (coding) strand. $S_\mu$, $S_\alpha$, and $S_\varepsilon$ are comprised of pentameric repeats and are enriched in GAGCT motifs, with the AGCT palindrome representing a canonical RGYW (or WRC) AID hotspot motif (discussed later). $S_\gamma1$, the largest S region (~12 kb) and similar to other $S_\gamma$ sequences, contains multiple 49bp repeats that are rich in RGYW motifs (including
AGCT) [1]. Although serving the same function, in some species the S region sequences can be quite divergent e.g. Xenopus S regions are AT-rich [12]. Most CSR junctions occur within and occasionally, just beyond the S region [14, 69]. Targeting studies have established the importance of S regions for efficient CSR, deletion of the core tandem Sµ repeats greatly reduced CSR [70, 71] and deletion of the 12kb Sγ1 abrogated IgG1 switching [72]. At least for CSR to IgG1, there is a direct correlation between length of the Sγ1 49-bp repeat sequence and its ability to promote CSR, as shown by replacement of endogenous Sγ1 with varying lengths of the core Sγ1 49-bp repeat units [73]. Thus S regions have evolved to provide a robust substrate for attracting AID, which initiates CSR by generation of double strand breaks (DSBs) in donor Sµ and downstream acceptor S region that consequently undergo long-range deletional recombination repair resulting in the production of different antibody class (Fig. I.2 A,C).

**Figure I.2: IgH constant region locus and secondary immune diversification reactions.**

a. Organization of the IgH constant region; each constant (C) region is preceded by a switch (S) region and a noncoding “I” exon. Blue oval between V(D)J exon and Iµ denotes IgH intronic enhancer (iEµ). Blue oval downstream of Cα represents IgH 3’ regulatory region (IgH 3’RR). Corresponding arrows depict transcription along VDJ exons and Cµ and Cy1 exons. b. AID generates point mutations and/or small deletions/insertions at the V(D)J exon during somatic hypermutation (SHM). c. AID-initiated DSBs in Sµ and Sγ1 could lead to intra-switch recombination (left) or could result in CSR by long-range deletional recombination generating IgG1 and excision circle byproduct (right). d. Working model for mechanism of AID recruitment. Figure: Alt FW et al., Cell 2013.
Direct evidence for the mechanistic requirement for germline transcription in CSR came from gene targeting studies where deletion of corresponding I exon promoters abrogated CSR to IgG1 [74] and IgG2b [75]. It was also shown that integrity of 3′IgH regulatory region (3′RR) containing a super-enhancer is necessary for CSR by regulating induction of germline transcription (GLT) of the upstream C_H genes [76-78]. Hence germline transcription through C_H genes has been linked to determining accessibility of S regions to undergo CSR. AID is a ssDNA specific deoxycytidine deaminase that lacks activity on dsDNA [6]. Owing to the specialized strand-biased sequence composition of G-rich sequences on the non-template strand of S regions, transcription through S regions creates R-loops whereby, template strand binds the nascent RNA transcript while looping out the non-template strand as ssDNA [72, 79-81]. However, inversion of endogenous S region experiments show that R-loop formation is not absolutely required for S-region function in CSR [72]. In this regard, Xenopus S_M regions, which are A/T-rich and do not form R loops up on transcription still support substantial CSR when replaced for mouse S regions in activated B cells [82].

In \textit{in vitro} T7 polymerase driven dsDNA transcription assays, ssDNA-binding protein, RPA, was shown as a cofactor that allowed AID to efficiently deaminate transcribed DNA (non-template strand), including the RGYW synthetic sequences and several tested variable region exons [83]. Addition of RNA exosome complex to the above \textit{in vitro} assay allowed AID access to the deamination of template strand [84] presumably due to destabilizing/degradation of transcription dependent transient RNA-DNA hybrids or R-loops (mammalian S region). High density RNA polymerase II (RNAP II) was detected in S regions suggesting the possibility of transcription machinery pausing in S regions to mediate AID-recruitment [85, 86]. In support of
this model, AID was also shown to associate with Spt5, a component of stalled RNA pol II complex [87]. Thus transcription of S regions is crucial in recruiting AID through various mechanisms (Fig. I.2 D).

Individual I exons are replete with multiple stop codons in all three reading frames and therefore germline transcripts of constant genes are unable to code for any peptide of significant length [1, 88]. Nevertheless, primary germline transcripts undergo processing, including splicing of intronic sequences including S regions, which fuses I exon to the C_H exons. Given the conserved structural organization and splicing patterns of all germline transcripts in C_H genes, the hypothesis of transcripts themselves or splicing per se having some role in the CSR process, potentially via mediating AID recruitment, has been proposed since long time. A number of studies [89-95] have tried to tackle this question but given various caveats, have not been able to convincingly show role of splicing and/or germline transcript in mediating CSR, potentially via regulating AID recruitment. However, recently a provocative finding showed spliced intronic S region RNA forms G-quadruplex structure which binds AID and recruits it to S region in IgH locus [96].

Altogether, transcription through S regions and its resultant associated features attracts and recruits AID via RNA Pol II pausing mechanisms, generating transcription dependent higher order chromatin structures such as R-loops, and potentially in trans by generating G-quadruplex structures of S region RNA, resulting in S-region DSBs leading to CSR.
Figure I.3: Diagram illustrating two long-range joining events between donor and acceptor S region DSBs.

Top, Diagram showing organization of IgH locus, rectangles denote V(D)J exons and constant exons. AID mediated DSBs are denotes in S regions (ovals) preceding C\(\mu\) and C\(\gamma_1\) exons. Arrow denote initiation of transcription. Bottom, two outcomes of long-range joining between S\(\mu\) and S\(\gamma_1\) DSBs leading to deletional joining (productive CSR, left) or inversional joining (non-productive CSR, right). Figure: Dong J, Panchakshari RP, Zhang T et al., Nature 2015.

S region DSBs generated in donor and acceptor S regions have various joining outcomes. They can be joined locally within the S region leading to intra-switch recombination (ISR) and internal deletions. ISR can result either from joining of two separate DSBs within the same S region or by re-joining of a single DSB post-resection. Other outcome includes joining occurring in cis over relatively long chromosomal distances (100 to 200 kb) that leads to productive CSR through deletional recombination by joining the upstream end of an S\(\mu\) DSB to the downstream end of an acceptor S region DSB. The remaining joining outcome is non-productive "inversional" CSR that inverts the intervening sequence between donor and acceptor DSB which has been found in transformed B cells [7, 9, 10, 97] (Fig. I.3).

Productive CSR requires distant double strand break ends in donor and acceptor S regions to be physically in orientation-specific proximity ("synapsed") for joining. Unlike RAG holding the DSBs in a post-cleavage complex, AID has no known activity of chaperoning two
different S-region DSBs to join in productive CSR manner. A key insight in to understanding long-range S region DSBs synapsis was provided by the experiments where endogenous $S\mu$ and $S\gamma 1$ regions replaced with I-SceI targets drove substantial recombinational class-switching upon expression of I-SceI meganuclease in activated B cells [98]. High-throughput analyses of joining frequencies between these widely separated IgH locus ISceI-generated DSBs indicated high level joining, similar results of robust long-range joining between these IgH locus ISceI-generated DSBs were seen in other cell types (e.g. T cells) [46]. Several recent studies have indicated that chromatin is organized into megabase or sub-megabase topological domains that contribute to increased frequency of interactions between sequences within them [99-102]; accordingly, DSBs in such domains were found to undergo high frequency joining within them [45, 59]. Recent high-resolution in situ Hi-C has revealed that mammalian genomes are divided into contact domains at an average scale of 185 kb [103]. The three-dimensional organization of the IgH locus as studied by 3C (chromosome configuration capture) assays have demonstrated that in IgH locus, $E\mu$ enhancer directly interacts with the downstream 3′RR and this long-range interaction could potentially facilitate donor and acceptor S region DSBs synapsis [104, 105]. Thus CSR has evolved to exploit general cellular mechanisms such as the general DSB response factors (DSBR, described below) and features of three-dimensional spatial genome organization to promote synapsis and frequent joining of long-range S region DSBs.
I.5 Activation Induced Cytidine Deaminase (AID)

Figure I.4: Repair pathways of AID mediated U:G lesion leading to different functional outcomes.

AID deaminates cytidine (C) to uridines (U). The U:G lesion is repaired by base excision repair (BER) or mismatch repair (MMR) pathway which results in SHM and CSR as follows: Replication over the U/G lesion produces transition mutations at C:G base pairs. Uracil-DNA-Glycosylase (UNG) of the BER pathway excises the U creating an abasic site. Replication over the abasic site generates transition and transversion mutations at C/G base pairs. AP endonuclease 1 (APE1) may create a nick at the abasic site. Nicks on both DNA strands may lead to DSBs in S regions which are processed by end-joining repair pathway that leads to CSR. MSH2-MSH6 (mismatch repair pathway) –mediated repair of U:G mismatch leads to spreading of mutations to A/T base pairs. Figure: Di Noia J and Neuberger MR, Ann. Rev. Immuno. 2007.

As mentioned above, AID was discovered to be almost exclusively upregulated in stimulated B cells [61]. Deletion of AID in mice and its mutations discovered in human patients confirmed AID’s requisite role in initiating SHM and CSR [62, 63]. AID in spite of sharing sequence similarity to APOBEC family deaminases (specifically APOBEC1 which deaminates C_{6666} to U in the mRNA encoding apolipoprotein B polypeptide) does not act on RNA [106,
*In vitro* studies revealed that purified AID acts specifically on single-stranded DNA substrates specifically targeting cytidine in RGYW hotspot motif. [6]. Evidence for DNA deamination by AID was largely pioneered by Dr. Michael Neuberger’s group who proposed AID-mediated deoxyctydine deamination induced U:G mismatch in the DNA would be a key intermediate and its differential processing by repair pathways would lead to secondary antibody diversification reactions like SHM and CSR.

Uracil in DNA is most commonly removed by the Base Excision Repair (BER) pathway [108]. Neuberger et al. proposed that AID-initiated generation of uracil up on undergoing removal by uracil-DNA glycosylase (UNG) creates abasic site which followed by cleavage by an apurinic/apyridimic AP endonuclease would lead to creation of a nick; AID activity on RGYW motifs (e.g. palindromic AGCT) on both DNA strands would create staggered nicks on opposite strands potentially leading to DSBs [109]. The evidence for role of UNG in this model came from studies in UNG−/− mice and human patients with UNG mutations showing substantially decreased CSR [110, 111]. Proposed role of AP-endonuclease (APE1) was also verified in CH12F3 B cell lymphoma line, where deletion of APE1 caused significant reduction in IgA CSR [112]. AID-mediated deamination of deoxyctydine also creates a U:G mismatch which can be recognized by the MSH2/MSH6 mismatch recognition heterodimer [113]. Accordingly, CSR is almost entirely ablated in mice double deficient for UNG and MSH2 [114]. Thus during CSR, AID-initiated U:G mismatch up on processing by general DNA repair pathways of base excision repair (BER) and mismatch repair (MMR) get converted into DSBs (Fig. I.4).
I.6 Classical non-homologous end-joining (C-NHEJ)

DNA double strand break (DSB) is repaired by two major pathways: Homologous recombination (HR) which uses homologous sequence as template for error-free repair or by classical non-homologous end-joining (C-NHEJ) which repairs the break ends with no or minimal processing. C-NHEJ pathway operates throughout the cell cycle but predominates during G0 / G1 phase when intact homologous template (sister chromatid) is not available [13] (Fig. I.5). Repair mediated by C-NHEJ can happen via: a) direct joining of blunt DNA break ends or ends with overhangs blunted after processing (either by removal of overhands or fill-in...
by DNA polymerases). b) Micro-homology (MH)-mediated repair where sticky ends (short complementary / micro-homologous sequence) embedded within break ends are exposed via processing and annealed at micro-homologous sequence. c) Insertions: DNA polymerases mediated de novo synthesis that adds several nucleotides at break ends before ligation thus leading to insertions at the repair junction. C-NHEJ joins DSBs most frequently (~40%) by direct joining and can also use micro-homology (1-4bp) for joining [12] (Fig. I.6).

**Figure I.6: Direct joining or microhomolgy (MH)-mediated joining during end-joining reactions.**
Variety of DSB ends could be processed and repaired by end-joining pathways. MH-mediated joining is facilitated by complementary base-pairing interactions between microhomologous sequence present in single-strand overhangs. C-NHEJ joins both MH-bearing and blunt (direct) ends, whereas A-EJ shows a strong preference for MH-mediated joining. Figure: Boboila et al, J. Ex. Med., 2010.

I.7  **C-NHEJ in V(D)J recombination**

Seven C-NHEJ pathway members have been identified namely: Ku70, Ku80 X-ray repair cross-complementing protein 4 (XRCC4), DNA ligase 4 (Lig4), DNA-dependent protein kinase, catalytic subunit (DNA-PKcs), Artemis, and XRCC4-like factor (XLF) [12]. Recently another protein, paralog of XRCC4 and XLF (PAXX), has been discovered and proposed to be an additional member of C-NHEJ pathway [115, 116]. Ku70, Ku80, XRCC4, and Lig4 are
evolutionarily conserved as they have homologs in yeast; these four factors are considered “core” C-NHEJ factors that are absolutely required for all C-NHEJ reactions including the joining of both hair pinned sealed CEs and blunt RSS signal ends during V(D)J recombination. C-NHEJ mediated repair entails: DSB recognition / protection by Ku70/Ku80 heterodimer, processing of DNA ends if necessary, and end ligation by the XRCC4/Lig4 complex [12].

V(D)J recombination is almost entirely blocked in mice deficient for core C-NHEJ factors Ku70, Ku80, XRCC4, or Lig4 leading to a block in development of B and T cells and causing severe combined immunodeficiency (SCID) phenotype [13]. This exclusive C-NHEJ dependent repair during V(D)J recombination is mediated by the RAG2 component of the RAG endonuclease which excludes both the HR and the alternative end-joining (A-EJ, discussed below) pathways from V(D)J repair in WT cells [56, 57]. In V(D)J recombination, hairpin-sealed CEs need to undergo opening and further processing before they can be joined, in this regard, DNA-dependent protein kinase catalytic subunit (DNA-PKcs) and Artemis, an endonuclease activated by DNA-PKcs, serve these specialized functions in C-NHEJ that mediate processing of hairpin-sealed CEs. Accordingly, complete deficiency for DNA-PKcs or Artemis in mice also results in a block in B and T cell development and a SCID phenotype [13]. Processing of DSBs via C-NHEJ also may involve de novo insertions by various DNA polymerases, such as DNA polymerase μ and DNA polymerase λ [12]. Terminal deoxynucleotidyl transferase enzyme (TdT), a lymphocyte specific polymerase [55], adds non-templated nucleotides to coding ends prior to their ligation, thereby greatly increasing junctional diversity leading to generation of a massive repertoire of antibody and TCR [2, 55].
I.8 C-NHEJ in IgH CSR

As mentioned above, RAG endonuclease complex shepherds repair of V(D)J recombination specifically to C-NHEJ, however, no such chaperoning activity of AID has been described during IgH CSR. In addition, unlike V(D)J recombination, AID-initiated S region DSBs are not precisely site-specific, but can occur at a variety of RGYW hotspot motifs along the highly repetitive long switch regions [1]. Although S regions are repetitive, the extent of homology between different S regions is not sufficient to provide a robust template for HR pathway [1]. Consistent with C-NHEJ mediated repair [12], donor and acceptor switch junctions during CSR displayed ~30-40% direct joins, while the remaining junctions utilized short micro-homology (MH) up to 1-4bp [14-17].

Correspondingly, C-NHEJ factors play an important role in mediating joining of S regions during CSR; Ku70- or Ku80-deficient B cells while blocked for V(D)J recombination, upon rescue of B cell development with pre-rearranged IgH and IgL are substantially impaired for CSR with up to 50% reduction observed compared to WT levels, these activated B cells also display significant unrepaired IgH breaks along with IgH locus derived chromosomal translocations [17]. Similarly deficiency for either XRCC4 or Lig4 in activated B cells results in a major reduction in CSR (~30-50% levels of WT) along with the accumulation of substantial levels of AID-dependent IgH locus chromosomal breaks [16]. Similar findings of significantly reduced levels of CSR in a Lig4-deficient condition were reported in CH12F3 B cell lymphoma line [19]. Ig heavy and light chain knock-in DNA-PKcs-null mice B cells displayed modest reduction in CSR to all isotypes except IgG1 [117], other models with DNA-PKcs mutants showed variable decreases in levels of CSR [118-120]. Fluorescence in situ hybridization (FISH) assays for IgH locus chromosomal breaks in DNA-PKcs-deficient, and to lesser extent in
Artemis-deficient stimulated B cells, revealed CSR derived S region break burden along with IgH locus translocations [121]. Hence DNA-PKcs and Artemis are also required for joining a subset of S region DSBs during CSR which potentially require further processing.

I.9 Alternative end-joining (A-EJ) in IgH CSR

Early studies with Ku70 deficiency in yeast cells [122] and Ku80- and XRCC4 deficiency in mammalian cell lines [123] showed efficient repair of linearized plasmid substrate, albeit in a MH-biased joining manner. Studies with repair of non-chromatinized DNA substrates (transiently expressed linear plasmids) and other biochemical assays established end-joining activity in cells lacking C-NHEJ ligation complex XRCC4/Lig4 to depend on the X-ray repair cross-complementing protein 1 (XRCC1)/DNA ligase 3 (Lig3) complex [124]; whereas in absence of C-NHEJ DSB recognition factor Ku70, cells utilized poly-ADP-ribose polymerase 1 (Parp1), a factor with known roles in single-strand break DNA repair to play a role in end-joining [125]. Definitive evidence for role of A-EJ pathway in repairing endogenous DSBs in the context of chromosomal substrates was provided by pro-B cell lymphomas derived from mice deficient for the C-NHEJ ligation factors XRCC4 or Lig4 and lacking the tumor suppressor p53 [126]. In spite of lack of XRCC4 or Lig4, B cells in these mice contained translocations that joined RAG-initiated IgH DSBs to DSBs around the c-myc gene which led to generation of oncogenic IgH:c-myc translocations associated with pro-B cell lymphomagenesis. More than a dozen such A-EJ mediated oncogenic IgH:c-myc translocation junctions were isolated and were found to have occurred by MH-mediated end joining [126]. Additional evidence came from studies showing repair of exogenous DSBs generated by the yeast meganuclease I-SceI in
chromosomally integrated reporter constructs in Ku80-deficient or XRCC4-deficient cells to be mediated by A-EJ [127, 128].

Role of A-EJ in mediating repair of physiologically generated DSBs came from analyzing cytokine stimulated XRCC4- or Lig4-deficient B cells which underwent IgH CSR at a robust level of up to 50% of CSR seen in WT B cells [16]. Similar findings of reduced yet substantial levels of CSR in a Lig4-deficient condition were reported in CH12F3 B cell lymphoma line [19]. In addition, 50% reduced CSR was also detected in B cells deficient for Ku or both Ku and Lig4 factors [17]. Presence of substantial CSR levels (up to 50% of WT levels) in various C-NHEJ factors deficient backgrounds indicated that alternative end-joining (A-EJ) pathways can also substantially contribute towards CSR joining, at least in the absence of C-NHEJ.

Analyses of the S junctions cloned from B cells deficient in various core C-NHEJ factors revealed perturbed junction profile, wherein, direct joins were almost non-existent and nearly all CSR junctions displayed MH-biased joining [15-17, 19, 129]; a fraction of these MHs in CSR junctions were also longer than those typically associated with C-NHEJ [15]. CSR switch junctions analysis in patients with defective DNA ligase IV activity also displayed a marked shift to the usage of microhomology at the Sμ–Sα junctions [130]. In accordance with earlier studies mentioned above, such MH-mediated junctions are often considered a feature of A-EJ pathways. In this regard, joining of S region DSBs due to the repetitive nature of S sequence could provide robust substrates for MH-mediated joining.

However, end joining mediated by A-EJ pathway in context of CSR is not necessarily always MH-mediated, as S-region junctions in Ku70-deficient or Ku70/Lig4 double-deficient mature B cells [17] all show presence of relatively significant proportion of direct joins.
compared to XRCC4- or Lig4- single deficient cells, indicating that MH is not a strict requirement of A-EJ. This has led to the proposal that at least two distinct sub-pathways of A-EJ that can mediate CSR [17]. One sub-pathway mediates repair in the absence of C-NHEJ ligation complex Lig4 or Xrcc4, where it relies on Ku for DSB recognition and either Lig1 or Lig3 for ligation and is strongly biased toward MH-usage [16, 17]. The other form of A-EJ occurs in the absence of C-NHEJ DSB recognition factor Ku, or the combined absence of Ku and Lig4, which preferentially repairs DSBs in MH-biased manner too, but still also frequently uses relatively higher direct joining [17].

I.10 DNA Double Strand Break response (DSBR) in CSR

DNA consistently incurs damage via various endogenous (e.g. RAG, AID, reactive oxygen species etc.) as well as exogenous sources (e.g. natural irradiation from environment) [12]. Of the many DNA insults, DNA double-strand break (DSB) could potentially be the most hazardous one because if unrepaired, it can result in cell death, and if illegitimately repaired can cause loss of genetic information or chromosomal translocations, one of the causative events in carcinogenesis. V(D)J recombination and IgH CSR, both are cell intrinsically programmed endogenous mutagenic events. Mammalian cells have evolved an elaborate response machinery to a DSB which broadly can be described as comprised of three stages: DSB sensing, signal transduction/amplification to recruit repair factors and effector proteins-mediated repair [20].

Initial DSB detection is mediated by a heterotrimeric MRN complex composed of the meiotic recombination 11 homolog A (Mre11)/Rad50/Nijmegen breakage syndrome 1 (Nbs1) [21]. This complex is required for activation of several phosphatidylinositol 3-kinase-like serine/threonine kinases (PIKKs) including ATM and DNA-PKcs [131]. Up on undergoing
activation via DSBs, ATM initiates the Double Stand Break Response (DSBR) by phosphorylating a set of proteins that includes histone variant H2AX, mediator of DNA damage checkpoint 1 (MDC1) and the p53-binding protein 1 (53BP1), which form large foci extending up to megabases along chromatin flanking DSB [20, 22].

One of the upstream events following ATM activation is the phosphorylation of the histone variant H2AX at Ser139 [132]. This DSB-induced phosphorylation of H2AX occurs within minutes after DNA damage and it rapidly spreads over large chromatin domains (>500 kb) flanking the DNA breaks [20]. MDC1 directly binds phosphorylated H2AX (known as γH2AX) and is required for amplification and further propagation of γH2AX foci surrounding DSB [133-135]. Along with formation of γH2AX in the chromatin surrounding the break, DSBR also entails mediating histone ubiquitylation in DSB vicinity by MDC1-dependent recruitment of RNF8/RNF168 factors, primary outcome of which is recruitment and/or retention of DSB repair/signaling factors including 53BP1, RIF1, PTIP, RAP80, BRCA1 etc. [136]. Thus, ATM-dependent DSBR leads to the assembly of macromolecular repair factors foci over large chromatin regions flanking the DSB [20, 22]. ATM also directly activates cell-cycle checkpoints in response to DSBs, including p53-dependent G1/S checkpoint that halts the cell cycle to allow DSB repair thus preventing replication of damaged DNA [22, 23].

ATM-dependent DSBR is activated in response to AID-initiated S region DSBs during CSR, as shown by Nbs1 and phosphorylated histone H2AX (γ-H2AX) co-accumulation on the IgH locus mostly in G1 phase in activated class switching B cells. [29]. Deficiencies for DSBR factors including ATM, H2AX, or MDC1 lead to reduced CSR ranging from 30% to 50% of WT levels (discussed below). ATM-dependent DSBR factors have also been implicated in channeling repair of DSBs (including AID-initiated S region DSBs) into the C-NHEJ pathway.
versus HR or A-EJ by antagonizing 5’-3’ resection of DNA ends that results in generation of 3’ overhangs [24-26, 28, 38, 137]. In this regard, resection of S region DSBs due to the repetitive nature of S sequence could provide longer MHs which are substrates for A-EJ.

ATM

ATM plays an important role during class-switching as ATM deficiency is shown to reduce CSR levels to roughly 50% of WT levels [31, 32]. ATM deficiency in stimulated B cells undergoing CSR displays substantial number of unreppaired AID-dependent IgH breaks, along with increased general chromosomal instability and significant IgH locus translocations [34]; thus suggesting a role of ATM-dependent DSBR response in end joining of CSR breaks. Correspondingly, in vitro activated ATM−/− B cells undergoing class switching exhibit substantially higher Sμ:c-myc translocations compared to activated WT B cells [138]. Along with the downstream DSBR response, ATM is proposed to be involved in potential tethering of DSBs to potentiate synapsis and prevent translocations [23]. ATM via DSBR-led recruitment of γH2AX and 53BP1 limits exonuclease mediated DNA end resection of DSBs [24-27], which suppresses DNA repair by both HR and A-EJ pathways, and promotes C-NHEJ repair. However, ATM also plays an antagonistic pro-resection role, via activation of CtBP-interacting protein (CtIP), where G1-arrested developing lymphocytes from Artemis−/−H2AX−/− background were shown to display increased exonucleolytic processing of V(D)J-associated DNA breaks [26]. Correspondingly, inhibition of ATM or its substrate CtIP, partially inhibited resection and partly rescued class-switching defects in H2AX- or 53BP1-deficient cells, an activity that appears to be potentially G1 cell cycle phase specific [24, 26, 27, 137].
Analysis of Sμ–Sα CSR junctions from A-T (ataxia telangiectasia) patients’ B cells showed strong bias towards MH-usage, where mean junctional homology was 7.2 bp compared to 1.8bp in WT human B cells [139]. Although relatively lesser compared to Sμ–Sα junctions, Sμ–Sγ junctions from A-T patients’ B cells also displayed significant MH-biased joining, mean junctional homology of 2.5 bp vs. 1.2 bp in WT human B cells. [139, 140]. The dissimilarity of magnitude of microhomology usage in Sμ–Sα and Sμ–Sγ junctions in A-T patients’ B cells is attributed to the higher degree of sequence homology between Sμ and Sα when compared with between Sμ and Sγ [141]. In mice, however, only one of the two ATM−/− mouse studies [31] found a small but significant increase in microhomology length at Sμ–Sγ1 junctions (2.6 bp vs. 1.2bp in WT), similar to the Sμ–Sγ junctions from A-T patients [140].

H2AX

Robust γH2AX accumulation is observed at the IgH locus in wild type G1 phase B cells undergoing CSR [29, 137]. Correspondingly, H2AX deficiency in mature activated B cells led to reduced CSR (~30% compared to wild type levels), thus indicating a role of H2AX in CSR. Similar to ATM-deficiency [30], H2AX absence also led to substantial unrepaired AID-dependent IgH breaks and showed IgH locus translocations along with increased general chromosomal instability [34, 138]. The IgH breaks in CSR-activated H2AX−/− B cells were observed on both chromatids, indicating that they were caused prior to the S phase, consistent with the evidence that S region DSBs are generated in the G1 phase [29, 34, 142].

It has been shown that γH2AX protects V(D)J recombination-generated DNA break ends from resection by CTBP-interacting protein (CtIP) in G1-arrested B cells, thereby, obviating
MH-mediated repair and promoting C-NHEJ mediated repair, and preventing aberrant repair of these V(D)J DSBs that could generate translocations [26]. Similarly protective function of γH2AX against resection of ISceI generated breaks in IgH locus was also observed in AID$^-$H2AX$^-$ B cells expressing ISceI meganuclease [38]. Resection as measured by ChIP-seq of replication protein A (RPA), a ssDNA-binding protein, showed dramatic increase as significant RPA ChIP-seq signal was detected at IgH switch regions undergoing CSR in the absence of H2AX [137]. However, in spite of resection inhibition function of γH2AX, unlike ATM, the amount of microhomology usage at Sμ–Sγ1 junctions was shown to be similar between H2AX$^-$ and wild type activated mouse B cells [30].

53BP1

In response to a DSB, 53BP1 rapidly accumulates on the chromatin surrounding the break site [143-145]. This is driven by a signaling cascade initiated by the ATM-mediated phosphorylation of the histone variant H2AX (γH2AX), followed by the recruitment of MDC1 and activation of RNF8–RNF168-dependent chromatin ubiquitylation cascade [146]. A key substrate of RNF168 is histone H2A, which is ubiquitylated on Lys13 and/or Lys15 (to form H2AK13ub or H2A15ub, respectively) in response to DNA damage [146]. Ubiquitination-dependent recruitment (UDR) motif present on carboxy-terminal of 53BP1 interacts with H2AK15ub [147, 148]. In addition, 53BP1 is also recruited to DSB sites by recognizing monomethylated and especially dimethylated H4K20 via its Tudor domain. The stable accumulation of 53BP1 at sites of DNA damage requires binding to these two separate histone modifications [39, 40].
One of the main 53BP1 functions during DSBR is blocking resection of breaks ends and promoting C-NHEJ, specifically highlighted in repair settings including class switch recombination (CSR), V(D)J recombination, telomere dysfunction and BRCA1-deficient cells [40]. 53BP1-deficient mice have been shown to display increased exonucleolytic processing of V(D)J-associated DNA breaks [26, 149-151]. In the context of telomere dysfunction, deletion of the shelterin protein, telomeric repeat binding factor 2 (TRF2) from mouse cells results in deprotected telomeres, which activates the ATM kinase signaling cascade and are processed by C-NHEJ pathway to form telomere fusions primarily in G1 phase [152]. Absence of 53BP1 leads to substantial resection of deprotected telomeres, resulting in significant decrease (~100-fold) in the rate of these telomere fusions [153]. In the absence of BRCA1, 53BP1 inhibits HR mediated DSB repair and inappropriately channels DSBs toward NHEJ repair, resulting in formation of radial chromosomes and other toxic chromosomal aberrations [25, 154, 155]. In this regard, absence of 53BP1 rescues embryonic lethality, DNA damage sensitivity and PARPi (PARP1 inhibitor) sensitivity (i.e. formation of PARPi-induced lethal chromosomal aberrations) associated with BRCA1 deficiency [25, 155].

Of all the DSBR factors, 53BP1 deficiency most severely impairs CSR with residual switching observed at ~5-10% of WT CSR levels [35, 36]. Similar to ATM-, H2AX-deficiency and classical non-homologous end-joining (C-NHEJ)-deficiency, 53BP1-deficiency also leads to substantial unrepaired AID-dependent IgH breaks and translocations in activated B cells [16, 18, 34], but in contrast, leads to little instability elsewhere in the genome [34]. These findings led to the hypothesis of a specialized role for 53BP1 in AID-dependent S region DSB repair. Additionally, only a subtle 2-fold decrease in CSR efficiency was observed in 53BP1−/−AID−/− B cells compared to AID−/− B cells, when DNA breaks were generated via I-SceI meganuclease
cutting at single I-SceI restriction sites knocked-in adjacent to Sμ and Sγ1 switch regions [24], similar findings by Zhang T., Alt Lab, unpublished data.

In context of CSR, several other functions of 53BP1 have been proposed, including preventing resection of S region DSB ends [24, 27, 38, 137] and influencing repair pathway choice between C-NHEJ and alternative end-joining (A-EJ) [24]. Experimental system using mouse B cells with I-SceI meganuclease cut site inserted at both Sμ and Sγ1, depicted greater resection of sequences involving I-SceI to I-SceI joining in AID−/−53BP1−/− background and a concomitant decrease in precise I-SceI to I-SceI direct joining [24, 38]. Resection of S region DSBs was dramatically increased in absence of 53BP1, as observed by RPA ChIP-seq signal detected at IgH switch regions undergoing CSR [137]. Similar to H2AX-deficiency, inhibition of ATM or its substrate CtIP also partially inhibited resection and relatively rescued class-switching defects in 53BP1-deficient cells [27, 38, 137].

ATM-dependent phosphorylation of 53BP1 recruits Rif1, an effector protein that inhibits resection of DSB ends [41-43, 156]. Accordingly, Rif1-deficiency in B cells results in a substantial decrease in CSR, however it is not as severe as that associated with 53BP1-deficiency [42, 43]. Thus the ability of 53BP1 to promote C-NHEJ in context of CSR is in part explained by its ability to block the 5′-3′ end resection at S region DSBs. Due to the repetitive nature of S regions, resection of S sequence in 53BP1-deficient background might lead to more chances of exposing microhomology in break ends for joining, a feature commonly associated with A-EJ pathway mediated joining.

Studies analyzing very limited number of Sμ–Sγ1 switch junctions from 53BP1−/− activated mouse B cells reported indistinguishable joining profile in terms of microhomology
usage compared to wild type B cells [36, 37]. Surprisingly, Sμ–Sγ1 junctions in RNF8-deficient B cells (RNF8/RNF168 mediate histone ubiquitylation that facilitates 53BP1 recruitment at DSB site) showed increased MH-usage [157]. In addition, a patient bearing a mutation in ubiquitin ligase RNF168, suffering from the RIDDLE syndrome (radiosensitivity, immunodeficiency, dysmorphic features, and learning difficulties), also displayed Sμ–Sα junctions with significant microhomology-biased joining pattern [158]. Thus a new comprehensive switch junction analysis is warranted in activated 53BP1−/− B cells.

I.11 Linear amplification mediated - High-Throughput Genome-wide Translocations Sequencing (LAM-HTGTS)

DSBs and their joining mechanisms are of considerable interest given the important role they play during adaptive immune system development and their potentially pathological outcomes in the case of illegitimate repair. In this regard, Alt lab first developed high-throughput genome-wide translocation sequencing (HTGTS) to identify genome-wide DSBs based on the ability of genome-wide DSBs to translocate to a fixed "bait" DSB generated by the yeast I-SceI meganuclease after cleavage of its 18-bp recognition site inserted in the c-myc gene on chr15 [44]. Correspondingly, this method has evolved to become a landmark technique where both a variety of different types of ectopic bait DSBs (e.g. Cas9-gRNA DSBs) or endogenous DSBs (e.g. RAG or AID-initiated DSBs) have been used as bait to identify various recurrent DSBs or recurrent classes of DSBs in different mouse and human cell types [44-51].

The ability of recurrent clusters of DSBs across the genome to be revealed by HTGTS results from cellular heterogeneity in 3D genome organization [45, 49, 51, 59]. The incidence of
translocations between two sites in the genome is a function of the frequency at which DSB ends at the two sites are available to be translocated and the frequency of their physically proximity [59]. Owing to the hierarchy in genome organization [101-103], within a *cis* chromosome, translocation frequency is further enhanced between DSBs that lie within topological domains/contact loops due to increased interaction or other processes [45, 48, 50, 59, 98]. Altogether, HTGTS serves as a remarkably sensitive DSB detection technology.

We have adapted a more recently enhanced version, linear amplification mediated version of HTGTS (LAM-HTGTS) (Fig. I.7) [49, 52] into an assay that allows us to track joining of endogenous AID-initiated DSBs in 5’ portion of S\(\mu\) in activated B cells to DSBs that occur across the entire length of downstream S regions [48]. By applying this assay in B cells deficient for various ATM-dependent DSBR factors, we provide several insights into mechanism of CSR, including, nucleotide-resolution quantification of resection of endogenous S region DSBs, maintenance of deletion orientation-specific biased joining of donor and acceptor S regions. We additionally highlight various salient features of this assay as a highly sensitive method for analyzing thousands of unique switch junctions, providing substantial advancement in understanding roles of ATM-dependent DSBR factors in end-joining during CSR.
Brief methodology of LAM-HTGTS: Junctions from bait DSB are directly amplified from sheared genomic DNA using single biotinylated primer by LAM-PCR, followed by streptavidin selection based enrichment and bridge adapter ligation to allow for exponential amplification and Miseq labeling of enriched products. Inset (bottom right) shows final annotated sequenced product. Figure: Hu et al., Nature Protocols, 2016.
I.12 References


CHAPTER 1

ORIENTATION-SPECIFIC JOINING OF AID-INITIATED DNA BREAKS PROMOTES ANTIBODY CLASS SWITCHING

Modified from original publication:

1.1 Abstract

During B cell development, RAG endonuclease cleaves immunoglobulin heavy chain (IgH) V, D, and J gene segments and orchestrates their fusion as deletional events that assemble a V(D)J exon in the same transcriptional orientation as adjacent Cμ constant region exons\(^1,2\). In mice, six additional sets of constant region exons (C\(_{HS}\)) lie 100–200 kilobases downstream in the same transcriptional orientation as V(D)J and Cμ exons\(^2\). Long repetitive switch (S) regions precede Cμ and downstream C\(_{HS}\). In mature B cells, class switch recombination (CSR) generates different antibody classes by replacing Cμ with a downstream C\(_{H}\)\(^2\). Activation-induced cytidine deaminase (AID) initiates CSR by promoting deamination lesions within S\(\mu\) and a downstream acceptor S-region\(^2,3\); these lesions are converted into DNA double-strand breaks (DSBs) by general DNA repair factors\(^3\). Productive CSR must occur in a deletional orientation by joining the upstream end of an S\(\mu\) DSB to the downstream end of an acceptor S-region DSB. However, the relative frequency of deletional to inversional CSR junctions has not been measured. Thus, whether orientation-specific joining is a programmed mechanistic feature of CSR as it is for V(D)J recombination and, if so, how this is achieved is unknown. To address this question, we adapt high-throughput genome-wide translocation sequencing (HTGTS)\(^4\) into a highly sensitive DSB end-joining assay and apply it to endogenous AID-initiated S-region in mouse B cells. AID. We further implicate ATM-dependent DSB-response factors in enforcing this mechanism and provide an explanation of why CSR is so reliant on the 53BP1 DSB-response factor.

1.2 Introduction

B cell development creates a massive repertoire of B cell receptor (BCR) providing ability to react against virtually any antigen; each individual B cell expresses a receptor with a
unique specificity towards a foreign antigen. The enormous repertoire of BCRs is created during B development in the bone marrow by a de novo assembly of the variable region portion of the immunoglobulin (Ig) molecule\(^1\). This process termed as V(D)J recombination is initiated in progenitor B cells during which single germline encoded variable (V), diversity (D) and joining (J) coding segments are assembled together creating a massive combinatorially diverse repertoire of antigen binding (Fab) region\(^1\)\(^2\). Successful assembly of variable regions on both immunoglobulin heavy chain (IgH) and light chain (IgL) locus via V(D)J recombination results in expression of BCR on the surface of a mature B cells. Pairing of the IgH and IgL chains leads to the formation of IgM BCR which is expressed on the surface of the resulting IgM\(^+\) B cell. Surface IgM\(^+\) B cells after exiting bone marrow to secondary lymphoid organs e.g. spleen or lymph nodes, upon encountering cognate antigen and activation (by cytokines, T-cells etc.), undergo two additional Ig diversification processes, namely IgH class switch recombination (CSR) and IgH and IgL variable region somatic hypermutation (SHM)\(^3\).

CSR changes antibody effector functions by replacing the C\(\mu\) exons with a different set of C\(\text{H}\) exons that lie 100-200kb downstream of C\(\mu\) in the IgH locus\(^2\). The mouse IgH region consists of eight sets of constant C\(\text{H}\) genes: 5′--VDJ--C\(\mu\)--C\(\delta\)--C\(\gamma\)\(3\)--C\(\gamma\)\(1\)--C\(\gamma\)\(2\)\(b\)--C\(\gamma\)\(2\)\(a\)--C\(\varepsilon\)--C\(\alpha\)--3′, spanning ~200 kb at the telomeric end of mouse chr12. Each germline C\(\text{H}\) gene is organized into an individual transcription unit composed of a 5′ promoter, a noncoding intervening (I) exon, switch (S) intron region and coding C\(\text{H}\) exons\(^2\). Activation induced cytidine deaminase (AID) directs CSR to particular isotype by generating DSBs in the donor S\(\mu\) and acceptor S regions which undergo long-range deletional recombination repair by non-homologous end-joining pathway leading to class switch recombination\(^2\). Joining between donor and acceptor S region DSBs could have two outcomes: a) in cis joining over relatively long chromosomal
distances (100 to 200 kb) that leads to productive CSR by deletional recombination by joining the upstream end of an Sµ DSB to the downstream end of an acceptor S region DSB, b) long-range *in cis* joining could also lead to non-productive "inversional" CSR which inverts the intervening sequence between donor and acceptor DSB. Unlike RAG endonuclease in V(D)J recombination, AID during IgH CSR has not been shown to possess any synaptic activity to shepherd the donor S and acceptor S region DSB ends to join in a biased manner. Although CSR is commonly described as a deletional recombination process, the relative frequency of deletional to inversional CSR junctions has not been measured\(^2\). In this study we address this long-standing question in the field, whether deletion orientation-specific joining is a programmed mechanistic feature of CSR and, if so, how this is achieved.

### 1.3 Results

Most chromosomal DSB ends join to ends of separate DSBs genome-wide without orientation (end) specificity\(^4,5\). Similarly, non-productive ‘inversional’ CSR joins have been found in transformed B cells\(^6-9\), suggesting CSR may also not be orientation-specific\(^10\) (Fig. 1.1a). To address this possibility, we employed digestion–circularization PCR (DC–PCR, Fig. 1.1.1a) to identify the orientation of CSR joins between Sµ and Sγ1 in purified mouse B cells stimulated with anti-CD40 plus IL4 to activate AID-targeting to Sγ1 and Scε, and class-switching to IgG1 (and IgE). Most Sµ to Sγ1 junctions identified by this semi-quantitative approach were deletional (Fig. 1.1.1b).
Figure 1.1 S-region-dependent orientation-biased joining in CSR-stimulated B cells.

a, Top, Igh C_{H} locus with AID targeting (lightning bolts) in S_{\mu} and S_{\gamma1}. Bottom, productive (left) CSR via deletional S_{\mu} to S_{\gamma1} joins plus excision circles; and inversionsal (right) CSR. 3’RR: 3’ regulatory region b, c, Joining outcomes from 3’ (red arrow) or 5’ (blue arrow) bait broken ends (BEs; bolts/gaps) to prey broken ends\(^5\). CEN: centromere, TEL: telomere d, Top, joining between I-SceI DSBs at S_{\gamma1} and S_{\mu} in \( \Delta S_{\mu}^{2,d}\Delta S_{\gamma1}^{2,d} \) B cells\(^12\). Middle, location of bait BE junctions to I-SceI-generated \( \Delta S_{\mu}^{2,d} \) or AID-initiated Sc prey broken ends in either + (red) or – (blue) orientation (five experiments) as a percentage of total Igh junctions (2 bins). Bottom, junction distribution to the 10-kb region encompassing I-SceI prey DSBs at S_{\mu} and AID-initiated prey DSBs at Sc. e, Results from \( S_{\gamma1}^{2,d/+} \) B cells\(^12\) plotted as in d. Grey box indicates prey junctions not assignable to single core-S_{\mu} sequence. Numbers in parenthesis denote total unique junctions. Statistical analyses in Fig. 1.2.1h, j.
Deletional CSR in \textit{in vitro} activated B cells by DC-PCR; \textit{Sce}1 DSBs within the \textit{Igh} constant region locus in activated B cells join with orientation-independence.

\textbf{a}, Schematic representation of DC–PCR assay. \textbf{b}, DC–PCR results from anti-CD40/IL4-activated wild-type and 53BP1\textsuperscript{-/-} B cells. \textbf{c}, Schematic representation of the HTGTS method. \textbf{d}, \textbf{e}, HTGTS libraries analyses of anti-CD40/IL4-stimulated \textit{Igh}\textsuperscript{36k} B cells with 3\textsuperscript{-}broken end (\textbf{d}, red arrow, \(n = 3\)) or 5\textsuperscript{-}broken end (\textbf{e}, blue arrow, \(n = 3\)) primers. BE, broken end. \textbf{f}, HTGTS libraries with 5\textsuperscript{-}broken end primer (blue arrow, \(n = 3\)) from \(\Delta S\mu 2.4/\Delta S\gamma 2.4\) B cells stimulated with anti-CD40/IL4. \textbf{g}, Bar graph depicting deletion/inversion and excision circle/inversion ratio between two I-SceI sites and between I-SceI and S-region in wild-type versus 53BP1\textsuperscript{-/-} backgrounds.
To confirm DC–PCR findings and analyse potential mechanisms, we used HTGTS, an unbiased genome-wide approach that identifies ‘prey’ DSB junctions to a fixed ‘bait’ DSB with nucleotide resolution. We refer to broken ends of bait *Igh* DSBs as 5′- and 3′-broken ends; specific primers allow use of each as bait (Fig. 1.1b, c). Prey junctions are denoted + if prey is read from the junction in a centromere-to-telomere direction and – if in the opposite direction (Fig. 1.1b, c). The + and – outcomes for intra-chromosomal joining of broken ends of different DSBs on the same chromosome include rejoining of a DSB subsequent to resection, or joining the broken ends of two separate DSBs to form intra-chromosomal inversions, deletions, or excision circles (Fig. 1.1b, c). To assess the relative frequency at which non-AID-initiated *Igh* DSBs join in deletional versus inversional orientation, we expressed I-SceI endonuclease in anti-CD40/IL4-activated, AID-deficient B cells in which I-SceI targets were inserted upstream of *Sμ* and downstream of *Sγ1* (*Igh^{1.96k}* allele; Fig. 1.1.1d, e), or in AID-sufficient B cells in which *Sγ1* and *Sμ* were replaced with I-SceI targets (Δ*Sμ^{2.6k}/ΔSγ1^{2.6k}* allele; Fig. 1.1d and Fig. 1.1.1f). HTGTS with primers respectively, captured junctions involving 3′- or 5′-broken ends I-SceI bait DSBs in the *Sγ1* locale revealed that a major class of recovered junctions were re-joins of bait DSBs following resection (Fig. 1.1d and Fig. 1.1.1d–f). A second major class of bait junctions in the *Sγ1* locale involved intact or resected 3′- or 5′-broken ends of I-SceI-generated DSB in the *Sμ* locale, which comprised relatively similar numbers of deletional (+) and inversional (−) junctions for bait 3′-broken ends (Fig. 1.1d and Fig. 1.1.1d) and similar numbers of excision circle (−) versus inversional (+) junctions for bait 5′-broken ends (Fig. 1.1.1e, f). As expected, bait 3′- and 5′-broken ends from the *Sγ1* locale recovered similar levels of + and – junctions genome-wide (Fig. 1.2.1a–d). We conclude that joining between two I-SceI DSBs in different
Igh S-region locations in CSR-activated B cells lacks any notable preference for or against inversional versus deletional joins.

In AID-deficient Igh \( ^{I-96k} \) B cells, I-SceI 5′- and 3′-broken end baits downstream of Sy1 did not capture Igh DSB hotspots beyond I-SceI-generated broken ends upstream of S\( \mu \) (Fig. 1.1.1d, e). In contrast, I-SceI 5′- and 3′-broken ends from the \( \Delta S\mu^{2.d}/\Delta Syl^{2.d} \) allele in AID-sufficient B cells joined frequently to AID-initiated Sc Sc junctions (Fig. 1.1.1d and Fig. 1.1.1f), with the majority (\ (~80%) of 3′ and 5′ \( \Delta S\mu^{2.d}/\Delta Syl^{2.d} \) broken ends joins distributed across the 4-kilobase (kb) Sc in orientations that generate, respectively, excision circles (Fig. 1.1d) or deletions (Fig. 1.1.1f). We also performed HTGTS on activated, I-SceI-expressing B cells in which only Sy1 was replaced by an I-SceI cassette (\( \Delta Syl^{2.d} \) allele; Fig. 1.1e). Beyond break-site junctions, major Igh hotspot regions of 3′ \( \Delta Syl^{2.d} \) broken ends were \( S\mu \) and Sc (Fig. 1.1e and Fig. 1.2.1j). Junctions occurred broadly across \( S\mu \), with 80% in a deletional orientation; while 90% of Sc junctions were in the reciprocal excision circle orientation (Fig. 1.1e; Fig. 1.2.1j). CH12F3 B lymphoma cells in which S\( \alpha \) was replaced with an I-SceI site had a similar orientation bias of S\( \alpha \) I-SceI 3′-broken end joining to S\( \mu \) DSBs (Fig. 1.2.1n–q). Joining of the 5′-broken ends of \( \Delta S\mu^{2.d} \) (on the \( \Delta S\mu^{2.d}/\Delta Syl^{2.d} \) allele) to AID-initiated DSBs in Sy3, Sy2b and Sy2a in LPS (lipopolysaccharide) plus anti-IgD-dextran-activated B cells were similarly orientation-biased (Fig. 1.3.1a–c). However, joining of the 5′-broken ends of \( \Delta S\mu^{2.d} \) across an array of 28 × I-SceI sites replacing Sy1 \( ^{13} \) was not orientation-biased (Fig. 1.3.1d, e). Together, these findings suggest that orientation-specific CSR joining requires an S-region sequence and/or unique aspects of S-region DSBs.
Figure 1.2.1 Genome-wide translocation junctions lack orientation bias; Statistical analyses for experimental replicates in Fig. 1.1 and 1.2; Orientation-biased joining between I-SceI break in place of Sγ and AID-initiated Sμ breaks in CH12F3 cells.

a, b, Circos plots for translocation junctions across the whole genome from 3′-broken end (a, n = 4) or 5′-broken end (b, n = 3) HTGTS with anti-CD40/IL4 stimulated ΔSγ1 ΔSγ2/I B cells. c, d, Bar graphs depicting genome-wide percentage of junctions from pooled 3′- and 5′-broken end libraries plotted separately in − or + orientations. Error bars are s.d. e, Joining from ΔSγ1 ΔSγ2 3′-broken end to AID off-target DSBs in Il4ra gene on chromosome 7. f, Bar graph showing the number of junctions (average ± s.d.) recovered from Igh1.96k AID−/− 3′-broken end HTGTS libraries (n = 3) at the break site and the upstream Sμ carrier break as a percentage of
Figure 1.2.1 (Continued)

the total number of junctions mapped to the 200 kb Igh constant region. Right panel shows the percentage of junctions mapping at SμI^d (average ± s.d.) over the total Igh junctions that are mapped in the deletion (Del) or inversionsal (Inv) orientation. The numbers above the bar graph (average ± s.d.) denote the ratio of deletional to inversionsal junctions. g, Bar graph showing the percentage of junctions (average ± s.d.) recovered from the Igh^106k.AID^+– 5′-broken end HTGTS libraries (n = 3). h, Bar graph showing the percentage of junctions (average ± s.d.) recovered from the ΔSμj2d /ΔSγl2d 3′-broken end libraries (n = 4). i, Bar graph showing the percentage of junctions (average ± s.d.) recovered from the ΔSμj2d /ΔSγl2d 5′-broken end libraries (n = 3). j, Bar graph showing the percentage of junctions (average ± s.d.) recovered from the wild-type ΔSγl2d 3′-broken end libraries (n = 3). k, Bar graph showing the percentage of junctions (average ± s.d.) recovered from the ΔSα1d CH12F3 3′-broken end libraries (n = 3) and ΔSα1d Sμ(INV) CH12F3 cells 3′-broken end libraries (n = 3). l, Bar graphs depicting percentage of trans junctions mapping to Sμ in – and + orientations from libraries of ΔSμj2d /ΔSγl2d B cells (n = 3) cloning from ΔSγl2d 3′-broken ends. m, Bar graphs depicting percentage of trans junctions mapping to Sμ in – and + orientations and to Se in – and + orientations from libraries of e-myc25.d 5′-broken ends (n = 3). n, o, HTGTS library analyses of ΔSα1d CH12F3 cells stimulated with anti-CD40, IL4 and TGFβ and nucleofected with I-Sce1 expression plasmid. Cells were harvested on day 3 post-stimulation for 3′-broken ends (n, n = 6) and 5′-broken ends (o, n = 6) libraries. p, 3′- and 5′-broken end libraries are normalized with ‘symmetric junctions’ (see Methods). q, Bar graph showing percentage of junctions from ΔSα1d CH12F3 cells (n = 6) from 3′- and 5′-broken end primers.
Figure 1.3.1 Joining between I-SceI break at S\(\mu\) and AID-initiated S-region breaks in LPS-activated (lipopolysaccaride); \(\Delta S\mu^{2*}/\Delta S\gamma^{2*}\) B cells and clustered I-SceI breaks in \(\Delta S\mu^{2*}/\Delta S\gamma^{2*}\) B cells in place of S\(\gamma\); Inverted S\(\mu\) in CH12F3 cells support robust IgA CSR.

a, Diagram of \(Igh\) locus organization in \(\Delta S\mu^{2*}/\Delta S\gamma^{2*}\) B cells highlighting AID-initiated breaks in S\(\gamma3\), S\(\gamma2b\) and S\(\gamma2a\) regions upon LPS stimulation and potential outcomes of CSR in the form of deletion (red, –) and...
Figure 1.3.1 (Continued)

inversional joining (blue, +). b, Plots showing enlarged distribution of pooled prey junctions in a 20-kb region flanking Sγ3 and Sγ2b and Sγ2a from HTGTS libraries of ΔSμγ2d/ΔSγγ2d B cells (n = 3) stimulated with LPS
and anti-IgD-dextran and infected with I-SceI-expressing retrovirus. c, Bar graph from three independent
ΔSμγ2d/ΔSγγ2d 5'-broken end libraries showing the percentage of junctions mapped at different S-regions. d,
Diagram of IgH locus organization in ΔSμγ2d ΔSγγ28d B cells highlighting joining outcomes of I-SceI-mediated
bait DSBs at ΔSμγ2d to clustered I-SceI DSBs at ΔSγγ28d in the form of deletion (red, −) and inversional joining
(blue, +). e, Pooled prey junctions from independent ΔSμγ2d ΔSγγ28d B cell libraries (n = 2, left panel,
emulsion-mediated PCR; n = 2, right panel, linear-amplification-mediated HTGTS). f, Southern blot for Sμ
inversion on the productive allele of CH12F3 cells with non-productive allele deleted. g, IgA CSR on day 3 for
CH12F3 (non-productive ΔSμ-Sα) cells stimulated with anti-CD40, IL4 and TGFβ. h, IgA CSR on CH12F3
(productive allele Sμ(INV), non-productive allele ΔSμ-Sα) cells stimulated with anti-CD40, IL4 and TGFβ.
Two independent clones of CH12F3 (Sμ(INV), non-productive ΔSμ-Sα).
Mammalian S-regions are G-rich on the non-template strand, giving AID-initiated 5’ and 3’ S-region broken ends a potential end-sequence bias. Also, when transcribed in the sense direction, S-regions generate stable R-loops\textsuperscript{14,15}, which could differentially affect 5’ and 3’ S-region broken end structure. To test the potential roles of S-regions in orientation-specific CSR, we used a Cas9/gRNA approach to invert S\textsubscript{\(\mu\)} on the productive allele of CH12F3 B cells, which modestly reduced CSR (Fig. 1.3.1f–h). We then assayed CH12F3 cells in which S\textsubscript{\(\alpha\)} was replaced with an I-SceI site and S\textsubscript{\(\mu\)} was in a normal or inverted orientation. These assays revealed that joins of I-SceI-generated 3’-broken ends at the S\textsubscript{\(\alpha\)} locale to S\textsubscript{\(\mu\)} DSBs were similarly biased for deletional junctions independent of S\textsubscript{\(\mu\)} orientation (Fig. 1.2a–c). Consistent with low-level \textit{trans} CSR\textsuperscript{16}, HTGTS libraries from activated AS\textsubscript{\(\mu\)}\textsuperscript{2\textsubscript{d}}/AS\textsubscript{\(\gamma\)}\textsubscript{I}\textsuperscript{2\textsubscript{d}} B cells contained numerous junctions from AS\textsubscript{\(\gamma\)}\textsubscript{I}\textsuperscript{2\textsubscript{d}} 3’-broken ends across the \textit{trans} S\textsubscript{\(\mu\)}; which, in contrast to \textit{cis} AS\textsubscript{\(\gamma\)}\textsubscript{I}\textsuperscript{2\textsubscript{d}} 3’-broken end S\textsubscript{\(\mu\)} junctions, occurred in + and – orientations at a similar frequency (Fig. 1.2d). Likewise, bait 3’-broken ends from the AS\textsubscript{\(\gamma\)}\textsubscript{I}\textsuperscript{2\textsubscript{d}} I\textsubscript{gh} allele identified approximately equal numbers of (+) versus (−) junctions to AID off-target DSBs in \textit{I\textsubscript{l4ra}} on chromosome 7 (Fig. 1.2.1e). Finally, translocations between bait 5’ I-SceI DSB broken ends in \textit{c-\textit{myc}}\textsuperscript{4} and prey AID-initiated S\textsubscript{\(\mu\)} and S\textsubscript{\(\varepsilon\)} broken ends in CSR-activated B cells lacked orientation bias (Fig. 1.2e). We conclude that orientation-dependent CSR joining does not require orientation-associated features of S\textsubscript{\(\mu\)} sequence, transcription, or transcripts. Moreover, AID-initiated DSBs \textit{per se} are not sufficient to promote orientation specificity, as demonstrated by orientation-independence of DSB joining to them \textit{in trans}. Thus, beyond S-region sequences and/or high frequency AID-initiated DSBs within them, aspects of I\textsubscript{gh} locus organization \textit{in cis} must play a critical role in promoting orientation-dependent CSR joining.
Figure 1.2 S-regions are not sufficient to promote orientation-biased CSR joining.

a, b, Top, HTGTS in CSR-activated CH12F3 cells for joining between 3′-broken ends of I-SceI DSB in the cassette replacing Sα and AID-initiated Sμ DSBs in which Sμ is in the normal (a; WT) or inverted (b; Inv) orientation. Bottom, junctions from I-SceI DSB 3′-broken ends to normal (left) or inverted (right) Sμ plotted as in Fig. 1.1d. e, Plot showing ratio (average ± s.d.) of + (deletional) to – (inversional) joins (n = 3), with significance calculated by unpaired two-tailed t-test (P = 0.307). NS, not significant. d, Top, HTGTS junctions from ΔSγI2∆ 3′-broken ends to AID-induced Sμ broken ends on trans chromosome in + (red) and – orientations in activated ΔSγI2∆ΔSμ2/d cells. Bottom, distribution of junctions in 10-kb trans Sμ (n = 5). e, Joining of bait I-SceI 5′-broken ends of c-myc23/d cassette4 on chromosome 15 and AID-induced Sμ and Sc Igh breaks in + or – orientations as in Fig. 1.1d. Top, linear distribution of junctions in Sμ (left) or Sc (right) (n = 4). Statistical analysis in Fig. 1.2.1k–m. (number in brackets in d and e are biological replicates)
We tested whether joining between two sets of endogenous AID-initiated S-region DSBs is orientation-dependent. Use of core S-regions DSBs as HTGTS bait is confounded by their highly repetitive nature. Therefore, we used as bait a 150-base-pair (bp) sequence at the 5' end of Sμ (5'Sμ), which retains 14 of approximately 500 Sμ AID-target motifs (Fig. 1.3a, left panel). HTGTS of anti-CD40/IL4-stimulated B cells with the 5'Sμ broken end primer revealed break-site junctions, as well as Sγ1 and Sε junctions (Fig. 1.3b, c). Consistent with AID-initiation, bait junctions were enriched at AID-targets within the 5'Sμ bait (Fig. 1.3a, right panel). 5'Sμ broken end junctions spread broadly over prey S-regions, with up to 95% in a deletional orientation (Fig. 1.3c). For comparison, we tested a 150-bp 5' remnant of Sμ (rSμ; Fig. 1.4.1a, left panel), retained when the rest of Sμ was deleted17. B cells homozygous for rSμ have reduced IgG1 CSR but nearly normal IgE CSR18. HTGTS with either 5’ rSμ or 3’ rSμ broken end primers of anti-CD40/IL4- and LPS/anti-IgD-dextran-stimulated B cells, respectively, revealed junctions to Sγ1 and Sε and to Sγ3, Sγ2b, and Sγ2a (Fig. 1.4.1). 5’ rSμ broken end junctions spread over target S-regions, with >90% in a deletional orientation (Fig. 1.4.1b, f); while >90% of 3’ rSμ broken end junctions were in the complementary excision circle orientation (Fig. 1.4.1c, g). Within the bait rSμ, junctions again were enriched at AID targets (Fig. 1.4.1a). Consistent with IgH class-switching patterns, rSμ HTGTS junctions occurred more frequently to Sε than those from the 5'Sμ bait in the context of full-length Sμ (Fig. 1.4.1b). Analyses of rSμ-mutant CH12F3 cells gave similar results (Fig. 1.5.1a–c). Thus, AID-initiated Sμ DSB joining to all downstream acceptor S-regions is strongly biased towards the deletional orientation.
Figure 1.3 Orientation-biased joining of AID-initiated endogenous S-region breaks.

a, Left, 150-bp 5’Sµ sequence used as HTGTS bait. Red arrow denotes 5’Sµ primer. Red and blue vertical lines indicate AGCT or other AID-targeting motifs, respectively. Right, distribution and frequency of 5’Sµ break points in junctions to downstream S-regions recovered from anti-CD40/IL4-stimulated wild-type B cells. Asterisks indicate positions of AGCT or other RGYW motifs.

b, Junctional outcomes from 5’Sµ AID-initiated broken end junctions to AID-initiated DSBs in Sγ1 and Sc including deletions (−) or inversions (+); long resections indicated by grey arrows. Break site 5’Sµ resections also are depicted.

c–e, Linear distribution of pooled junctions along 200-kb Cµ locus (left) or at Sγ1 and Sc (middle and right) recovered from anti-CD40/IL4-stimulated wild-type (n=5), ATM−/−(n=3) (d) or 53BP1−/−(n=8) (e) B cells. Grey boxes indicate repetitive sequences with junctions mapping to multiple locations in Sγ1; asterisks indicate G-rich Sγ1 regions devoid of AID motifs and junctions.
Figure 1.4.1 Orientation-biased joining between AID-initiated rSμ and downstream AID-initiated S-region breaks in anti-CD40/IL4-activated and LPS-activated Sμ-truncated B cells.

a, 150-bp rSμ sequence used as HTGTS bait with red arrow indicating rSμ 5′-broken end HTGTS primer; red and blue vertical lines indicate canonical AGCT or other RGYW AID-targeting motifs, respectively. Distribution of rSμ break points in junctions to downstream S-regions recovered from anti-CD40/IL4-stimulated rSμ B cells. b, HTGTS analyses of anti-CD40/IL4-activated rSμ B cells, 5′ rSμ AID-initiated broken end (red primer, n = 3) junctions to AID-initiated DSBs in Sγ1 and Sε which includes deletion (− orientation, red) or inversions (+ orientation, blue). c, HTGTS analyses of anti-CD40/IL4-activated rSμ B
Figure 1.4.1 (Continued)

cells, 3′ rSμ AID-initiated broken end (blue primer, \( n = 3 \)) junctions to AID-initiated DSBs in Sγ1 and Sε which includes excision circle (± orientation, blue) or inversions (− orientation, red). d, Bar graph showing the percentage of junctions (average ± s.d.) from anti-CD40/IL4-activated rSμ 5′-broken end libraries mapped to Sγ1 and Sε. e, Bar graph showing the percentage of junctions (average ± s.d.) from anti-CD40/IL4-activated rSμ 3′-broken end libraries mapped to Sγ1 and Sε. f, HTGTS analyses of LPS-activated rSμ B cells, 5′ rSμ (red primer, \( n = 3 \)) AID-initiated broken end junctions to AID-initiated DSBs in Sγ3 and Sγ2b and Sγ2a which includes deletional joining (− orientation, red) or inversions (± orientation, blue). g, HTGTS analyses of LPS-activated rSμ B cells, 3′ rSμ (blue primer, \( n = 3 \)) AID-initiated broken end junctions to AID-initiated DSBs of above LPS stimulated cells in Sγ3 and Sγ2b and Sγ2a which includes excision circle (± orientation, blue) or inversions (− orientation, red). h, i, Percentage of junction distribution at Sγ3, Sγ2b and Sγ2a in both orientations from both 5′-broken end libraries (h) and 3′-broken end libraries (i) are shown as average ± s.d. from three independent experiments.
Figure 1.5.1 Orientation-biased joining between rSµ and AID-induced Sα DSBs in CSR-activated CH12F3 cells.

a, Diagram outlining potential junction outcomes from 5’ rSµ (red primer) or 3’ rSµ (blue primer) AID-initiated broken end junctions to AID-initiated DSBs in Sα upon anti-CD40, IL4 and TGFβ stimulation of ΔSµ CH12F3 cells. b, c, Top panel shows HTGTS libraries analyses of day 2 (b) and day 3 (c) stimulated CH12F3 (non-productive allele ΔSµ-Sα, productive allele ΔSµ) cells cloning from 5′-broken end rSµ (red primer, n = 3), whereas lower panel shows HTGTS libraries cloning from 3′-broken end rSµ (blue primer, n = 3). d, Bar graph shows percentage of junctions (average ± s.d.) for 5′-broken end and 3′-broken end libraries indicated in b and c. e, Bar graph with percentage of junctions (average ± s.d.) from rSµ libraries mapped to prey Sα in the deletion (DEL) or inversion (INV) for 5′-broken end libraries and in excision circle (EC) or inversion orientation for 3′-broken end libraries.
CSR DSBs generate a DSB response (DSBR) in which ATM activates histone H2AX and 53BP1 in chromatin flanking DSBs, thereby contributing to end-joining\textsuperscript{19-21}. ATM or H2AX deficiency moderately reduces CSR (Fig. 1.6.1a)\textsuperscript{2,19}. However, 53BP1 deficiency causes a more drastic reduction (Fig. 1.6.1a), suggesting specialized CSR roles\textsuperscript{2,19,22}, such as promoting S-region synapsis or protecting S-region DSBs from resection\textsuperscript{11,23-25}. To elucidate influences on orientation-specific CSR, we employed HTGTS to assay joining of AID-initiated 5′Sμ broken ends to AID-initiated Sγ1 and Sε DSBs in anti-CD40/IL4-activated ATM-, H2AX-, and 53BP1-deficient B cells, as well as in B cells deficient for Rif-1, a 53BP1-associated factor that mediates resection blocking\textsuperscript{26,27}. ATM-, H2AX-, and Rif1-deficient B cells had reduced Sγ1 and Sε junctions compared to wild type; 53BP1-deficient B cells had a greater reduction, with most localizing to the break-site region (Fig. 1.3d, e and Fig. 1.6.1b–d). Most break-site junctions were resections, which were longest (up to about 6 kb) for 53BP1 deficiency (Fig. 1.6.1e, f). Compared to wild type, bait 5′Sμ junctions to Sγ1 and Sε DSBs in different DSBR-deficient backgrounds had varying decreases in orientation specificity, with H2AX deficiency having the smallest and 53BP1 deficiency the largest (Figs 1.3d, e and 1.4a; Fig. 1.6.1c, d and Table 1.1a, b). Indeed, residual junctions of 5′Sμ to Sγ1 and Sε locales in 53BP1-deficient B cells showed relatively normalized inversion:deletion ratios (Fig. 1.4a), a finding confirmed by DC–PCR (Fig. 1.1.1b). Finally, 53BP1-deficiency did not impact joining orientation of 5′Sμ and 3′Sγ1 I-SceI-generated broken ends in AID-deficient Igh\textsuperscript{K-96k} B cells (Fig. 1.1.1g).
Figure 1.6.1 Level of junctions to downstream S-regions in wild-type and DSBR-deficient 5’Sμ HTGTS libraries correlate with CSR levels; 5’Sμ break site undergoes variable degrees of resection from stimulated DSBR-deficient B cells.

a, Table showing IgG1 and IgE CSR levels of splenic B cells from various genotypes (with number of replicates indicated) activated in vitro with anti-CD40 and IL4. FACS was performed on day 4 and values
indicate average ± s.d. WT, wild type. b, Left panel shows bar graph for percentage of junctions (average ± s.d.) recovered from wild-type 5′Sμ 5′-broken end libraries mapped to the Sμ, Sγ1 and Sc over the total number of junctions identified from the 200-kb Igh constant region. Remaining panels shows the similar results from different DSBR-deficient backgrounds using the same 5′-broken end primer. c, d, 5′Sμ 5′-broken end HTGTS libraries analyses from H2AX−/− and RIF1fl/flCD19cre B cells are shown respectively. e, Diagram of potential junction outcomes from 5′Sμ AID-initiated 5′-broken end junctions to AID-initiated DSBs in Sγ1 and Sc. f, Data from HTGTS libraries mapped to the 20-kb region flanking 5′Sμ break site from B cells stimulated with anti-CD40/IL4 in wild-type and DSBR-deficient backgrounds.
Figure 1.4 Mechanistic roles of Igh organization and DSBR factors in deletional CSR.

a, Ratios of inversional to deletional 5'Sµ joins to Sy1 (left) or Sc (right) in wild-type and mutant cells. Average ± s.d. calculated from at least three separate experiments. b, Sy1 and Sc resection junctions mapping to deletional (−) resection regions where the incidence of wild-type junctions decreases to background plotted as the percentage of total junctions in the deletional orientation from cells in a. For panels a and b, more than 1,000 unique junctions (up to tens of thousands in some cases) were analysed. Statistical significance or insignificance of key comparative results calculated by unpaired two-tailed t-test from at least three biological repeats (Table 1.1). Ai: ATM inhibitor c, d, Working model for orientation-biased joining (c) and functions of DSBR proteins (d) in maintaining directional CSR.
Owing to the potential difficulty in measuring relative resection of recurrent re-joins at or near the break-site (discussion section), we focused on prey S-region broken end resections. Because S-regions are long and AID-initiated DSB locations within them are diverse, we estimated relative resection by quantifying bait broken end to prey broken end junctions downstream of S-region positions where the incidence of wild-type junctions decreases to background (Fig. 1.3b–e). Based on this ‘long’ S-region resection assay, ATM- and H2AX-deficient cells had modest resection increases, Rif1-deficient cells slightly greater increases, and 53BP1-deficient B cells far greater increases that were also apparent as a ‘flattening’ of Sγ1 and Sε junction profiles relative to other backgrounds (Figs 1.3c–e and 1.4b; Fig. 1.6.1c, d and Table 1.1c,d). HTGTS assays of rSμ bait broken end junctions to Sγ1 and Sε (Fig. 1.7.1) and I-SceI-generated 3′ ΔSγ12ΔI broken end bait junctions to Sμ and Sε (Fig. 1.8.1) gave similar results. In H2AX- or Rif1-deficient B cells, a large fraction of 5′Sμ junctions were within S-regions, with the main difference from wild type being a subset of junctions extending beyond S-regions, probably reflecting extensive resection of broken ends not rapidly fused (Fig. 1.4b and Fig. 1.6.1c, d). Treatment of 53BP1-deficient activated B cells with ATM kinase inhibitor substantially diminished very long S-region resections, but did not alter orientation-dependent joining (Fig. 1.4a, b; Table 1.1 and Fig. 1.9.1a–f). This finding may reflect shorter resections in inhibitor-treated 53BP1-deficient versus ATM-deficient B cells that are not revealed by our long resection assay. Another possibility would involve a putative specialized role for 53BP1 in stabilizing synapsed S-regions23.
Figure 1.7.1 Orientation-biased joining between rSµ and AID-induced Sγ and Sε DSBs in wild-type, ATM-deficient, and 53BP1-deficient B cells.

a, Diagram of potential junction outcomes from 5′ rSµ AID-initiated broken end junctions to AID-initiated DSBs in Sγ1 and Sε as described earlier. b–d, Linear plots of pooled junctions across the 200-kb Igh constant region (first panel), the 20-kb region flanking rSµ break site (second panel), the 20-kb region flanking downstream Sγ1 (third panel) and Sε (last panel) from three independent wild-type (b), ATM- (c), or 53BP1- (d) deficient 5′-broken end rSµ libraries. e, Bar graphs showing Inversion/Deletion (INV/DEL) Bias ratios of HTGTS Sγ1 (left panel) and Sε (right panel) junctions in different genotypes showing average ± s.d. from three independent libraries for each genotype. p-values calculated by unpaired two-tailed t-test f, Bar graphs showing percentage of long resection of Sγ1 (left) and Sε (right) junctions in different genotypes as average ± s.d. n.s., not significant. p-values calculated by unpaired two-tailed t-test g, Bar graphs (average ± s.d.) showing statistical analyses of reproducibility for experiments mentioned above.
Figure 1.8.1 Orientation-biased joining of I-SceI DSBs at Sγ1 to AID-induced S-region breaks in various DSBR-deficient backgrounds.

a. Schematic illustration of the ΔSγ1Δ2 allele and joining outcomes from 3′-broken end (red arrow) to AID-initiated upstream Sμ and downstream Sε DSBs. b. Linear distribution of junctions between ΔSγ1Δ2 3′-broken end to AID-induced Sμ/Sε region DSBs in anti-CD40/IL4-stimulated wild-type (b, n = 4), ATM−/− (c, n = 3), H2AX−/− (d, n = 3), and 53BP1−/− (e, n = 4) cells across the 200-kb Igh region (first panels), 10-kb Sμ (second panels) and Sε (last panels). f. Bar graphs (average ± s.d.) showing the percentage of junctions mapped at
Figure 1.8.1 (Continued)

$\Delta SyI^{2'd}$ (break site), S$_{\mu}$ and S$_{\epsilon}$ over the total number of junctions in the 200-kb $\text{Igh}$ constant region in different genotypes. g, Bar graphs (average ± s.d.) showing the percentage of junctions from above various genotypes of $\Delta SyI^{2'd}$ 3'-broken end libraries mapped to S$_{\mu}$ and S$_{\epsilon}$ as average ± s.d. h, Bar graph (average ± s.d.) showing comparison of percentages of junctions cloned using $\Delta SyI^{2'd}$ 3'-broken end involving resection of S$_{\mu}$ (top panel) and S$_{\epsilon}$ (bottom panel) breaks in indicated backgrounds.
Figure 1.9.1 Inhibition of resection in 53BP1-deficient B cells by ATMi (ATM inhibitor) does not rescue directional CSR joining to Sy1.

a–d, Linear plots of pooled junctions across the 200-kb Igh constant region (left panels), the 20-kb region flanking downstream Sy1 (middle panels) and Se (right panels) from wild type plus DMSO control (a, n = 2), wild type plus ATMi (b, n = 3), 53BP1−/− plus DMSO (c, n = 3) and 53BP1−/− plus ATMi (d, n = 3) libraries are shown as above. e, Bar graph showing the percentage of Sμ, Sy1 and Se junctions (average ± s.d.) from wild type plus DMSO 5′Sμ libraries (left) and wild type plus ATMi 5′-broken end libraries (right, n = 3). f, Bar graph showing the percentage of Sμ, Sy1 and Se junctions (average ± s.d.) from 53BP1−/− plus DMSO (left) and 53BP1−/− plus ATMi (right) 5′Sμ libraries.
Table 1.1 Statistical comparison for orientation bias and resection in Sγ1 and Sε junctions from wild-type and DSBR-deficient B cell libraries

a, b, Unpaired two-tailed t-test for the degree of bias in the Sγ1 (a) and Sε (b) junctions between WT and DSBR-deficient B cells with full-length Sm was performed for experiments described in Figs 3 and 4. Orientation bias was calculated as described in the Methods for individual libraries. Numbers in parenthesis indicate times of independent experiments performed for each genotype. N/A, not available. c, d, Unpaired two-tailed t-test for the level of resections in the Sγ1 (c) and Sε (d) junctions was performed for experiments described above. Percentage of resection was calculated as described in the Methods for individual libraries. Numbers in parenthesis indicate number of times of independent experiments performed for each genotype.

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>ATM&lt;sup&gt;+&lt;/sup&gt;</th>
<th>H2AX&lt;sup&gt;+&lt;/sup&gt;</th>
<th>53BP1&lt;sup&gt;+&lt;/sup&gt;</th>
<th>53BP1&lt;sup&gt;+&lt;/sup&gt; + Ai</th>
<th>RIf1&lt;sup&gt;+&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>a WT</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>ATM&lt;sup&gt;+&lt;/sup&gt; (n=3)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>H2AX&lt;sup&gt;+&lt;/sup&gt; (n=3)</td>
<td>0.0057</td>
<td>0.0031</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>53BP1&lt;sup&gt;+&lt;/sup&gt; (n=3)</td>
<td>0.001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>53BP1&lt;sup&gt;+&lt;/sup&gt; + Ai (n=3)</td>
<td>&lt;0.0001</td>
<td>0.0039</td>
<td>0.001</td>
<td>0.1481</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>RIf1&lt;sup&gt;+&lt;/sup&gt; (n=3)</td>
<td>&lt;0.0001</td>
<td>0.181</td>
<td>0.0094</td>
<td>&lt;0.0001</td>
<td>0.0029</td>
</tr>
<tr>
<td>b WT</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>ATM&lt;sup&gt;+&lt;/sup&gt; (n=3)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>H2AX&lt;sup&gt;+&lt;/sup&gt; (n=3)</td>
<td>0.0034</td>
<td>0.0291</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>53BP1&lt;sup&gt;+&lt;/sup&gt; (n=3)</td>
<td>&lt;0.0001</td>
<td>0.0003</td>
<td>0.0002</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>53BP1&lt;sup&gt;+&lt;/sup&gt; + Ai (n=3)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>RIf1&lt;sup&gt;+&lt;/sup&gt; (n=3)</td>
<td>0.0012</td>
<td>0.9274</td>
<td>0.1154</td>
<td>0.0003</td>
<td>N/A</td>
</tr>
<tr>
<td>c WT</td>
<td>&lt;0.0002</td>
<td>0.0037</td>
<td>0.3012</td>
<td>&lt;0.0001</td>
<td>0.0002</td>
<td>0.0008</td>
</tr>
<tr>
<td></td>
<td>ATM&lt;sup&gt;+&lt;/sup&gt; (n=3)</td>
<td>&lt;0.0001</td>
<td>0.0002</td>
<td>0.0002</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>H2AX&lt;sup&gt;+&lt;/sup&gt; (n=3)</td>
<td>0.0037</td>
<td>0.3012</td>
<td>0.7974</td>
<td>0.0005</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>53BP1&lt;sup&gt;+&lt;/sup&gt; (n=3)</td>
<td>&lt;0.0001</td>
<td>0.158</td>
<td>0.7974</td>
<td>0.0005</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>53BP1&lt;sup&gt;+&lt;/sup&gt; + Ai (n=3)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>RIf1&lt;sup&gt;+&lt;/sup&gt; (n=3)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>d WT</td>
<td>&lt;0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>ATM&lt;sup&gt;+&lt;/sup&gt; (n=3)</td>
<td>&lt;0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>H2AX&lt;sup&gt;+&lt;/sup&gt; (n=3)</td>
<td>&lt;0.0001</td>
<td>0.0279</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>53BP1&lt;sup&gt;+&lt;/sup&gt; (n=3)</td>
<td>&lt;0.0001</td>
<td>0.0001</td>
<td>0.0003</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>53BP1&lt;sup&gt;+&lt;/sup&gt; + Ai (n=3)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>RIf1&lt;sup&gt;+&lt;/sup&gt; (n=3)</td>
<td>0.0001</td>
<td>0.0199</td>
<td>0.058</td>
<td>0.0062</td>
<td>N/A</td>
</tr>
</tbody>
</table>
1.4 Discussion

We demonstrate that CSR is mechanistically programmed to occur in a productive, deletional orientation. Based on our findings, we propose a working model for orientation-specific CSR, in which a key component is the organization of S-regions within topologically-associated domains (TADs) that promote their frequent S-region synapsis\textsuperscript{2,12,13} via Langevin motion\textsuperscript{2,13,28} (Fig. 1.4c). Within such TADs, we implicate additional Ig\textit{h}-specific organizational features, not yet fully elucidated, in playing a fundamental role in mediating synapsis in an orientation that promotes deletional joining (Fig. 1.4c). We find that functions of such organizational features are complemented by S-regions, potentially associated with their ability to promote AID-initiated DSBs, multiple frequent DSBs, or both. Our studies also implicate DSBR factors in enforcing this mechanism (Fig. 1.4d). The broader DSBR probably contributes by tethering un-synapsed S-region DSBs for efficient re-joining, keeping them from separating into chromosomal breaks that could frequently translocate with orientation independence to S-region broken ends within the TAD\textsuperscript{2,20}; this function would also allow subsequent AID-initiated breakage and joining to a synapsed S-region (Fig. 1.4c).

DSBR factors also prevent long end-resections that could cause S-region broken ends to linger in resection complexes, preventing synapsis with other S-region broken ends and/or diminishing ability to be joined by classical NHEJ (Fig. 1.4d). Different DSBR factors have differential impact in tethering versus resection inhibition and, thus, may impact orientation dependence via different routes. For example, ATM deficiency inhibits resection by impairing CtIP activation\textsuperscript{29}, but promotes resection via other nucleases by impairing inhibitory activities of H2AX, 53BP1 and, indirectly, Rif1\textsuperscript{26,27} (Fig. 1.4d). 53BP1-deficiency is unique in that it both impairs tethering for rejoining and activates resection of un-joined ends by failure to activate
Rif1, leading to extreme resections and the greatest impairment of CSR and orientation-dependent joining (Fig. 1.4d). As common and unique impacts of 53BP1 deficiency markedly affect both donor and acceptor S-regions, they would be multiplicative and, thereby, explain the profound impact of 53BP1-deficiency on CSR.

Discussion for breaksite analysis

Comparison of I-Sce1 3′-BE (1.1.1d) or 5′-BE (1.1.1e) derived junctions at IgH1-96k Sγ1x1 breaksite (lower right panels in d and e) versus junctions to upstream Sμ1x1 break (lower left panels in d and e) reveals a striking difference; namely, it seems that a large fraction of 5′ or 3′BEs from Sγ1x1 breaksite that are re-joined undergo resections extending to several kb. However, such a high degree of apparent resection is not observed in junction profiles involving Sμ1x1 5′ or 3′BE joining to prey Sμ1x1 BEs (1d,e, lower left panels) which present much tighter, sharper peaks. We have seen this broader break-site junction pattern with other sets of nuclease bait and prey junctions. Our interpretation of this finding is that the HTGTS approach, like all prior PCR based (and other) approaches cannot readily measure total break-site junctions because a large number represent perfect re-joins (which could also be cut again). Thus, the actual level of break-site junctions likely is much higher than indicated by the break-site rejoining peak. In this regard, the resection tail from break-site rejoins is real; but it is the tail of a much larger peak of rejoined DSBs at or near the break site that cannot be readily quantified (mostly perfectly rejoined DSBs but others joined in a few limited patterns close to the break site that also could yield many copies of a given junction that cannot be distinguished from a PCR repeat). Therefore, level of break-site resections can be artificially inflated as a proportion of re-joined junctions recovered at the break-site. In this regard, due to bait to prey junctional
diversity, the junction distribution at the prey $S\mu^{1xl}$ DSBs presents a much more accurate representation of resection frequency and length. Prior PCR-based methods for analyzing resection of with respect to rejoining single nuclease DSBs have similar limitations in accurately estimating bait cutting and an additional limitation because they cannot capture long-range resections as they are restricted in size by the need for two sets of PCR primers within several hundred bp.

The general issues discussed above regarding the impact on break-site rejoining analyses with respect to inability to determine perfectly rejoined break-site DSBs also could apply to 5’Sμ or rSμ break-site, albeit to a lesser extent considering that there are multiple different AID-initiated DSBs in rSμ and these would more rarely be expected to be perfectly rejoined due to anticipated end structure. However, there could also be a large number joins immediately adjacent to the break-site that would be identical and their frequency could vary in different backgrounds (e.g. WT versus 53BP1-deficient) due to different degrees of resection in a given background (which would alter numbers of such repetitive joins). While long resections of break-site rejoins of AID-initiated DSBs in 5’Sμ and rSμ do appear to correlate in general with the long resection levels determined by joining to downstream bait S region DSBs, we chose to use the latter as resection measure as described in the text to avoid any potential ambiguities associated with breaksite rejoin measurements in different WT and mutant backgrounds.
1.5 References


CHAPTER 2

DEFICIENCIES FOR DNA DOUBLE-STRAND BREAK RESPONSE PROTEINS PROMOTES MICROHOMOMOGY-BIASED IGH CSR JUNCTIONS

Rohit A. Panchakshari*, Junchao Dong*, Vipul Kumar and Frederick W. Alt

* Equal contribution
2.1 Abstract

During B cell class switch recombination (CSR), activation-induced cytidine deaminase (AID) initiates DNA double strand breaks (DSBs) in switch (S) regions that are joined by end-joining. Deficiency for the ATM kinase and its various DNA DSB response (DSBR) substrates, including histone H2AX and 53BP1, impairs CSR to variable extents in association of varying degrees of increased end-resection. We report use of a high throughput genome-wide DSB joining assay to study influence of different DSBR factor deficiencies on the structure of CSR junctions between AID-initiated DSBs in the 5' portion of the donor Sμ region to those across the length of downstream acceptor S regions. Based on analyses of thousands of junctions, we find that absence of DSBR factors leads to varying increases in micro-homology (MH)-mediated junctions, with 53BP1-deficiency having the greatest increase. However, while translocation junctions between cas-9/gRNA-induced DSBs in c-myc to AID-initiated S region DSBs in ATM- or 53BP1-deficient B cells show similar biases in MH-usage to those observed in the context CSR junctions, translocation junctions to other DSBs genome-wide had no MH-usage increase in ATM-deficient cells and only a modest increase in 53BP1-deficient cells. We discuss these findings with respect to potential roles of AID-initiated DSBs in S regions to be prone to MH-usage potentially due to their increased resection along with their highly repetitive nature that provides abundant micro-homologous sequence.

2.2 Introduction

Mature B cells undergo immunoglobulin heavy chain (IgH) class switch recombination (CSR) to mediate different antibody effector functions. CSR occurs by replacing the initially expressed μ constant gene (Cμ) that is juxtaposed to the rearranged V(D)J segments with a
downstream constant gene\(^1\). In mouse IgH locus, six C\(_{\text{H}}\) genes, C\(_{\gamma}3\), C\(_{\gamma}1\), C\(_{\gamma}2\)b, C\(_{\gamma}2\)a, C\(_{\varepsilon}\) and C\(_{\alpha}\), lie up to 200kb downstream of the C\(_{\mu}\) gene\(^1\). Each C\(_{\text{H}}\) gene is an independently transcribed unit containing a non-coding ‘intervening’ (I) exon, an intronic switch sequence (S), followed by coding constant gene exons. S region sequences are long (4-12 kb) repetitive DNA sequences consisting of tandem G-rich repeats on the non-template strand\(^2\). Stimulating B cells with different combinations of activators and cytokines direct CSR to particular C\(_{\text{H}}\) gene by modulating its germline transcription to recruit activation-induced cytidine deaminase (AID)\(^1,2\). AID initiates CSR by introducing multiple DNA lesions into donor S\(_{\mu}\) and acceptor S regions\(^1,2\), which are subsequently converted to staggered double-strand breaks (DSBs) by the general base excision and/or mismatch repair mechanisms\(^3\). Productive CSR is completed by joining the upstream end of donor S\(_{\mu}\) DSB to the downstream end of an acceptor S-region DSB via nonhomologous end-joining (NHEJ) pathway. We have previously shown CSR to be a deletional recombination process in which productive deletion-joining occurs at a frequency at least 10-times higher than the non-productive events resulting from inversion of intervening sequence between S\(_{\mu}\) and acceptor S-region\(^4\).

DSBs generated during CSR triggers ataxia telangiectasia-mutated (ATM)-dependent DSB response (DSBR). Upon activation, ATM kinase phosphorylates a series of downstream substrates including histone variant H2AX, MDC1, 53BP1 etc. that assemble into macromolecular foci over large chromatin region surrounding DSBs to amplify damage signaling and tether DSB ends for efficient repair\(^5,6\). Deficiency of ATM or H2AX leads to impaired CSR at 30%-50% of corresponding wild type cells and accumulation of substantial AID-dependent Ig\(H\) breaks, indicating a role of ATM/H2AX in the joining phase of CSR\(^6\). Absence of 53BP1 leads to a much more profound CSR defect where only 5% of wild type
switching levels is observed accompanied by increased intra-Sμ joining. Various specialized functions have been proposed to explain the role(s) of 53BP1 in CSR, including promoting end-joining, facilitating long-range S-S synapsis and protecting S region DSB ends from resection. Resection is a 5′-3′ nucleolytic degradation process that reveals 3′ single-strand overhangs at DSB ends. Rif1, a recently identified 53BP1-associating effector protein that binds ATM-phosphorylated 53BP1, prevents CtIP-dependent DSB resection; correspondingly, B cells deficient for Rif1 have largely impaired CSR to downstream S regions. Recent studies have revealed that ATM and downstream DSBR factors function as a barrier for end-resection. Our HTGTS assay has recently provided the first evidence at a single nucleotide resolution that deficiency in the ATM-dependent DSBR in CSR-activated B cells leads to increased S region DSBs resection with 53BP1-deficiency showing highest level of S-region DSBs resection, whereas ATM, H2AX and Rif1 deficiency modestly increasing S-region DSBs resection. Resected DSBs with long-exposed 3′ssDNA overhangs are not optimal substrates for classical end joining repair and alternative end joining pathway has been proposed to fuse such DSB ends with the usage of microhomology (MH) sequence revealed in the resected DNA. It is thus conceivable that MH-biased joining of S-region DSBs should occur in DSBR mutants due to increased resection.

CSR represents a unique platform to study S-region DSB resection and its influence on end-joining by studying the structure of S-S joining junctions via DNA sequencing. However previous reports using conventional PCR-Sanger sequencing method with very limited number of CSR junctions showed that in the absence of ATM, H2AX, 53BP1 or Rif1, the MH-usage pattern of S-S junctions did not change compared to wild type B cells; although contradictory findings were also reported from the immunodeficient A-T (ataxia telangiectasia).
human B cells and ATM−/− mouse. On the other hand CSR junctions from B cells-deficient for c-NHEJ factors showed increased usage of MH. Conventional CSR junction analysis method due to technical difficulties in amplifying long and repetitive S sequences, relies on primers at the peripheral ends of chimeric S regions to amplify joining products, thus recovering junctions only mapping to the edges of S regions. In addition to this caveat, DSB resection could possibly delete either primer-binding site, leading to the loss of recovery of CSR junctions from resected DSBs, potentially skewing obtained MH-usage pattern. The number of junction sequences studied with such traditional assay is usually very low (<50) and is also not cost-effective.

Our HTGTS-based end-joining assay differs from the traditional method; it utilizes linear amplification-PCR (LAM-PCR) with a single primer in the bait 5’Sμ region, hence minimizing the effect of resection-induced loss of prey junctions. Furthermore, with only about 100-200bp of the donor and acceptor S sequence amplified during the procedure combined with the power of high-throughput next-generation sequencing, HTGTS readily maps thousands of junctions at single nucleotide resolution throughout the entire S region which was not possible before.

In this study we have analyzed junction structure/MH-usage patterns of CSR junctions identified by high-throughput genomic translocation sequencing (HTGTS)-based end-joining assay in activated DSBR-deficient mouse B cells. HTGTS represents an innovative highly sensitive assay to readily identify tens-of-thousands of unique S-S junctions providing mechanistic insights into the nature of CSR joining. Applying HTGTS-based joining assay in CSR setting provides a powerful resource to directly quantify the level of resection in a desired mutant and study its S-S junction structure. We offer concrete evidence that ATM-dependent
DSBR factors prevent extensive S-region DSBs end resection to promote classical end joining during antibody class switching.

2.3 Results

**HTGTS assay reveals general pattern of CSR junctions between AID-mediated S\(\mu\) and various downstream S region breaks**

To study the joining pattern of AID-initiated S region breaks in activated mature B cells, we *in vitro* stimulated rS\(\mu\) B cells that have most of its S\(\mu\) sequence deleted except retaining truncated ~200bp remnant S\(\mu\) sequence containing 14 AID target RGYW motifs\(^4,25\). HTGTS assay performed using primer to the 5’ of rS\(\mu\) region readily identified thousands of junctions joining AID-mediated rS\(\mu\) breaks to S\(\gamma\)1 and S\(\varepsilon\). Based on the junction structure between rS\(\mu\) bait sequence and downstream prey S region sequence, we categorize CSR junctions into three main groups: those in which bait sequence is followed directly by prey sequence, termed as direct joins; those with ≥1bp overlapping sequence between bait and prey denoted as MH joins, and finally those with ≥1bp gap between bait and prey sequence as insertion joins which account for 20-30% of both rS\(\mu\)-S\(\gamma\)1 and rS\(\mu\)-S\(\varepsilon\) junctions. As insertions are of unknown sources (especially those up to more than 100bp) we decided to separate them out from the total junctions for further analyses\(^12\). When plotted as percentage of junctions with length of MH over the total number of junctions in the core S\(\gamma\)1 or S\(\varepsilon\) region, rS\(\mu\)-S\(\gamma\)1 and rS\(\mu\)-S\(\varepsilon\) junctions both contain about 30% of direct joins and the rest being joined using MH (Fig. 2.1A, 2.1D). The percentage of junctions using ≥5bp MH was very rare (<5%) in both rS\(\mu\)-S\(\gamma\)1 and rS\(\mu\)-S\(\varepsilon\) junctions (Fig.2.1A). When stimulated with LPS and anti-IgD-dextran, rS\(\mu\) exclusively switched
to $S\gamma3/S\gamma2b/S\gamma2a$ and the junctions to these S regions showed a comparable junction structure profile to $S\gamma1$ and $S\varepsilon$ junctions (Fig. 2.1B). Similar direct joining frequency and microhomology usage pattern between r$S\mu$ and $S\alpha$ was also observed in r$S\mu$-CH12F3 cell line stimulated with $\alpha$CD40, IL4 and TGF-β (Fig. 2.1C). We further used this assay for cloning CSR junctions from endogenous full-length $S\mu$. Upon stimulation of wild type B cells with anti-CD40/IL4, we obtained thousands of junctions mapping to $S\gamma1$ and $S\varepsilon$. In this case too, direct joins accounted for about 30% of junctions to downstream $S\gamma1$ and $S\varepsilon$ breaks while the rest being joined using microhomology (Fig. 2.2, Table 2.1).

Our assay corroborates the earlier findings in WT B cells regarding CSR joining being mediated by c-NHEJ wherein, about 30% junctions exhibit direct joining and the rest mostly being joined using MH of up to 3-4 bp$^{20}$. Secondly, our assay reveals that joining of $S\mu$ to all acceptor S region breaks follow similar pattern of junction structure. Unlike prior methods, our HTGTS assay also provides thousands of CSR junctions spreading across the entire length of acceptor S regions.
Figure 2.1. MH pattern of joining from rSμ to downstream S region DSBs in *in vitro* stimulated rSμ cells.

Primary SμΔΔ cells were purified and stimulated by αCD40/IL4 (A) or LPS/αIgD-dextran (B). CH12F3 cells with Sμ on the productive allele deleted were stimulated with αCD40/IL4/TGF-β (C). Cells were harvested 96 hrs (A, B) or 72 hrs (C) after stimulation and genomic DNA were extracted for HTGTS. Junctions with indicated microhomology are plotted as percentage of total junctions excluding insertions. Error bars represent s.e.m. (D) Direct joins (average±s.e.m.) from HTGTS libraries plotted in (A)-(C) are displayed side-by-side. Information regarding replicates and number of junctions analyzed is mentioned in Table 2.3a.
DSBR deficiency promotes MH-biased CSR joining

We have recently shown that in DSBR-deficient B cells AID-mediated S region breaks undergo variable degree of resection with 53BP1-deficient cells exhibiting highest S-region DSBs resection⁴. We asked how the deficiency in DSBR alters microhomology usage in the S region junctions. Preforming HTGTS assay with our 5’Sμ primer, we directly examined 5’Sμ bait junctions mapping to the core Sγ1 or Sε region in WT, ATM⁻/⁻, H2AX⁻/⁻, 53BP1⁻/⁻ and Rif1⁻/⁻ B cells stimulated with anti-CD40/IL4. We identified thousands of junctions in the core Sγ1 and Sε regions from each of the DSBR deficient B cells, compared to WT B cells we observed a marked increase in the percentage of junctions using MH in all the DSBR mutants (Fig. 2.2A, 2.2C). 53BP1-deficient cells showed biggest shift towards MH-usage (Fig. 2.2A, 2.2C), correspondingly, direct joins in core Sγ1 and Sε regions in 53BP1⁻/⁻ cells were significantly decreased at about 10% (Fig. 2.2B, 2.2D). ATM⁻/⁻, H2AX⁻/⁻ and Rif1⁻/⁻ cells displayed roughly 20% direct junctions between Sμ and core Sγ1 and Sε region (Fig. 2.2B, 2.2D). The average length of MH used in Sμ-Sγ1 junctions increased from 1.70bp in WT to 2.47bp, 2.84bp and 3.25bp in ATM-, H2AX- and 53BP1-deficient cells, respectively (Table 2.1); whereas for Sμ-Sε junctions, the average length of MH increased from 1.48bp in WT to 2.05bp, 2.02bp, 1.90bp and 2.74bp in ATM-, H2AX-, Rif1- and 53BP1-deficient cells, respectively (Table 2.1).

Similar results with increased MH usage and decreased direct joining frequency in rSμ-Sγ1 and rSμ-Sε junctions were observed in CSR junctions from ATM- and 53BP1-deficient rSμ B cells (Fig. 2.1.1 A-D, Table 2.2a). Analysis of ATM- and 53BP1- deficient CH12F3 cells stimulated with αCD40, IL4 and TGF-β compared to WT CH12F3 cells also showed similar trend of increased MH usage in the Sμ-Sα junctions (Table 2.1, Figure 2.2.2 A-B). MH analysis
of junctions from HTGTS libraries cloned from I-SceI-generated 3’ΔSγ12xl BE to prey core Sμ DSBs revealed similar trend of increased MH-prone joining in ATM-, H2AX- and 53BP1-deficient B cells (Fig. 2.3.3 A-D, Table 2.2b). Although ΔSγ12xl 3’BE junctions to Sc in ATM-deficient cells did not show significant decrease in direct joins, the average length of MH of these junctions was significantly longer than those from WT cells (Fig. 2.3.3 A-D, Table 2.2b).

As shown previously, treatment of 53BP1−/− deficient B cells with ATM inhibitor KU55933 efficiently blocks ATM-dependent S region DSB resection and rescues CSR4,26. While examining the junction structure profile of Sμ-Sγ1 junctions in ATMi-treated 53BP1-deficient cells, we found that ATM inhibition restored the frequency of direct joins in 53BP1−/− cells to about 20% which was comparable to Sμ-Sγ1 direct junction percentage observed in ATM−/− B cells. (Fig. 2.2B, 2.2D). This was also accompanied by relative decrease in the average length of MH in 53BP1−/−+ATMi cells where for Sγ1 junctions it dropped to 2.69bp from 3.25bp seen in control 53BP1−/−+DMSO cells (Table 2.1). We observe that deficiency in DSBR factors leads to MH-prone joining of S region DSBs during CSR.
Figure 2.2. DSBR-deficient B cells show increased MH usage during S region DSB joining.

Primary spleen B cells with full length S\(\mu\) in wild type, \(ATM^{-/-}\), \(H2AX^{-/-}\), \(53BP1^{-/-}\), \(Rif1^{fl/fl}\) \(Cd19^{cre}\) and \(53BP1^{-/-}\) treated with ATM inhibitor were stimulated with \(\alpha\)CD40/IL4 for 96 hours and assayed as described in Fig.1. MH pattern of junctions 5'\(rS\mu\) to \(S_{\gamma 1}\) (A) and 5'\(rS\mu\) to \(S_{\epsilon}\) (C) breaks are plotted as percentage of total junctions excluding insertions. Error bars represent s.e.m. Percentages of direct joins (average±s.e.m.) of 5'\(rS\mu\) to \(S_{\gamma 1}\) (B) and \(S_{\epsilon}\) (D) in different genetic backgrounds are compared. Information regarding replicates and number of junctions analyzed is mentioned in Table 2.3b. Statistical comparisons (two-tailed t test) for experiments (B and D) are presented in Table 2.4a.
Table 2.1. Average length of MH ± s.e.m. from Sµ-Sγ1, Sµ-Sε and Sµ-Sα junctions

Average length of MH ± s.e.m. in various genotypes as described in figure 2.2 and fig. 2.2.2 are listed. Numbers in parenthesis indicate total number of junctions mapped to the core S regions in the indicated genotypes.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Core Sγ1</th>
<th>Core Sε</th>
<th>Core Sα</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>1.70±0.01 (n=25103)</td>
<td>1.48±0.01 (n=9782)</td>
<td>1.78±0.03 (n=3066)*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.71±0.03 (n=2569)**</td>
</tr>
<tr>
<td>ATM&lt;sup&gt;−−&lt;/sup&gt;</td>
<td>2.47±0.02 (n=17161)</td>
<td>2.05±0.03 (n=4427)</td>
<td>2.52±0.04 (n=2366)***</td>
</tr>
<tr>
<td>H2AX&lt;sup&gt;−−&lt;/sup&gt;</td>
<td>2.84±0.04 (n=3669)</td>
<td>2.02±0.04 (n=2044)</td>
<td>N.A.</td>
</tr>
<tr>
<td>Rif1&lt;sup&gt;−−&lt;/sup&gt;</td>
<td>N.A.</td>
<td>1.90±0.04 (n=1881)</td>
<td>N.A.</td>
</tr>
<tr>
<td>53BP1&lt;sup&gt;−/−&lt;/sup&gt;</td>
<td>3.25±0.02 (n=8657)</td>
<td>2.84±0.04 (n=1836)</td>
<td>2.71±0.07 (n=722)****</td>
</tr>
<tr>
<td>53BP1&lt;sup&gt;−/−&lt;/sup&gt; +ATMi</td>
<td>2.69±0.03 (n=5751)</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
</tbody>
</table>

* WT CH12F3  ** NCΔ-CH12F3  *** ATM<sup>−−</sup> CH12F3  **** 53BP1<sup>−/−</sup> NCΔ-CH12F3
Figure 2.1. MH pattern in S-S junctions from primary Sμα/Δ B cells in WT, ATM or 53BP1-deficient background.

Cells were stimulated with αCD40/IL4 for 96 hours and assayed as described in main figure 2.1. MH pattern of joining from 5’rSμ to Sγ1 (A) and Sε (C) breaks is plotted as percentage of junctions with indicated MH length to the total junctions excluding insertions. Error bars represent s.e.m. Percentages (average±s.e.m.) of direct joins to Sγ1 (B) and Sε (D) in different genetic backgrounds is compared. Information regarding replicates and number of junctions analyzed is mentioned in Table 2.3a. Statistical comparisons (two-tailed t test) for experiments (B and D) are presented in Table 2.4b.
Figure 2.2.2. Linear plot of S region DSB joining and MH pattern in activated WT, ATM- and 53BP1-deficient CH12F3 cells.

CH12F3 cells in WT (n=4) and ATM⁻/⁻ (two independent clones, n=5), NCA-CH12F3 in WT (n=3) and 53BP1⁻/⁻ (n=3) backgrounds were stimulated with αCD40/IL4/TGFβ. HTGTS libraries with genomic DNA harvested 72hrs after stimulation were made with 5'rSμ primer set as described⁴. (A) Linear plot for junctions mapped in the 250kb IgH locus (left) or 20kb Sα-surrounding region (right) were plotted as described previously⁴. (B) MH usage in junctions from 5'rSμ breaks joining to AID-induced Sα breaks in either WT and ATM⁻/⁻ CH12F3 cells (left) or WT and 53BP1⁻/⁻ NCA-CH12F3 cells (right) were plotted as percentage (average±s.e.m) of junctions with indicated length of MH to the total number of junctions in the regions. To map junctions into Sα region in WT and ATM⁻/⁻ CH12F3 cells, junctions falling into chromosomal region with coordinates chr12:114498531-114500087 were excluded because this region also exists on the non-coding Sμ-Sα switched allele thus interfering the assessment of only long-range junctions. Information regarding replicates and number of junctions analyzed is mentioned in Table 2.3d. (C) WT CH12F3, ATM⁻/⁻ (two independent clones) and NCA-53BP1 (two independent clones) CH12F3 cells were stimulated with anti-CD40/IL4/TGFβ and IgA CSR was measuring using FACS on Day 3 post-stimulation.
Figure 2.3.3 MH pattern in joining Δγ12xI to S region DSBs in WT, ATM or 53BP1-deficient B cells.

Δγ12xI splenic B cells of indicated genotypes were stimulated with αCD40/IL4 and infected with I-SceI-expression retrovirus 24 hour later. Cells were harvested 96 hours after stimulation for genomic DNA extraction and HTGTS. MH pattern of joining from I-SceI joining to Sμ (A) and Sε (C) were plotted as percentage of junctions with indicated MH length to the total number of junctions mapped to that core S region. Error bars represent s.e.m. The percentage (average±s.e.m.) of direct joins of ISceI to Sμ (B) and Sε (D) are compared across genotypes. Information regarding replicates and number of junctions analyzed is mentioned in Table 2.3c. Statistical comparisons (two-tailed t test) for experiments (B and D) are presented in Table 2.4c.
Translocation joining to S-region but not genome-wide DSBs uses more MH in absence of DSBR

We next asked whether or not the joining to genome-wide translocation junctions from fixed DSB bait has similar junction pattern to that observed for joining between S region DSBs. We introduced a CRISPR-Cas9-mediated blunt DSB at c-myc locus and assayed its translocation joining to genome-wide breaks versus AID-initiated S region DSBs in IgH locus (Sμ and Sα) in αCD40/IL4/TGF-β stimulated WT, ATM- and 53BP1-deficient CH12F3 cells by HTGTS. The results indicated that similar to control WT-CH12F3 and NCA-CH12F3 cells, genome-wide translocation junctions (removed IgH locus and chr15 junctions) in ATM-deficient cells contained ~30% junctions undergoing direct joining while the remaining being joined using similar range of MH. 53BP1-deficient CH12F3 cells displayed slightly more MH-biased junctions in genome-wide translocations, a small but significant decrease in percentage of direct joins was seen compared to wild type (Figure 2.3A, 2.3B). A striking difference was observed when we looked at translocation junctions joining to AID-initiated S region DSBs (junctions to Sμ and Sα) from CRSPR-Cas9 induced blunt c-myc DSB bait, where unlike wild type cells, ATM-deficient CH12F3 cells had significantly reduced frequency of direct junctions and a concomitant substantial shift towards MH-biased joining. This extensive MH-prone joining pattern for c-myc\textsuperscript{Cas9}-Sμ/Sα translocation junctions was even more pronounced in 53BP1-deficient CH12F3 cells (Figure 2.3C, 2.3D). We suggest that AID-initiated S region DSBs owing to their AID-initiated source, structure (U:G mismatches converted to DSBs) and/or high frequency may potentially be more susceptible to resection than genome-wide DSBs and therefore junctions to S region DSBs are inclined towards being more MH-biased in DSBR deficient cells.
Figure 2.3. Translocation joining to S-region breaks in activated DSB response-deficient B cells are more prone to resection and MH-usage

WT, ATM- or 53BP1-deficient CH12F3 cells were stimulated with αCD40/IL4/TGFβ and a cmyc-Cas9 bait DSB was introduced at 12 hrs after stimulation. HTGTS libraries with genomic DNA harvested 72hrs after stimulation were made and MH usage in junctions from cmyc-Cas9 break joining to genome-wide breaks (A) and S region breaks (Sμ and Sα combined, C) were plotted as percentage (average±s.e.m) of junctions with indicated length of MH to the total number of junctions in the regions. Percentages of direct joins (average±s.e.m.) of cmyc-Cas9 break joining to genome-wide breaks (B) and S region breaks (D) in different genetic backgrounds are shown. Information regarding replicates and number of junctions analyzed is mentioned in Table 2.3e. Statistical significance calculated by two-tailed t test.
MH-mediated repair does not impact deletion orientation-biased CSR

Our previous results from HTGTS assay indicated that CSR is a mechanistically programmed orientation-specific recombination process wherein deletional joining between donor and acceptor S regions is strongly preferred over inversional joining. Highly repetitive S regions provide robust substrates for MH-mediated end joining. Theoretically, the biased nucleotide composition of template versus non-template S region strands could contribute to orientation-specific joining. To test this notion, we examined the MH usage pattern in the junctions mapped to the core Sγ1 and Sε regions in either (-) or (+) orientation from WT and 53BP1-deficient cells. The results showed indistinguishable levels of direct versus short MH junctions in deletional (-) versus inversional (+) orientation joins of 5’Sµ to AID-initiated Sγ1 or Sε BEs in both WT and 53BP1-deficient cells (Fig. 2.4A-D). We have previously shown that deletion-biased joining of S region DSBs is S-sequence orientation-independent. Upon further analyzing junctions from ΔSα/Scel 3’BE joining to Sµ in either physiological or inverted orientation, we found no difference in the pattern of MH usage (Fig. 2.4.4 A-B). Thus, MH-related joining mechanisms are unlikely to contribute to orientation-dependent CSR.
Figure 2.4. MH-mediated repair does not affect deletion orientation-biased CSR.

Sμ-Sγ1 (A) and Sμ-Sc (B) junctions from HTGTS libraries made with WT and 53BP1−/− B cells described in figure 2 were separated into junctions joining in (+) or (-) orientation, subsequently MH pattern of these (+) and (-) junctions was plotted as percentage of junctions with indicated MH length to the total number of junctions in that orientation. Error bars represent s.e.m. The percentage (average±s.e.m.) of direct Sμ-Sγ1 (C) and Sμ-Sc (D) joins in both orientation from WT and 53BP1−/− cells are compared. n.s.: not significant (two-tailed t test)
Figure 2.4.4. MH-usage is not affected regardless of S region sequence orientation.

(A) MH pattern of joining of I-SceI-generated 3’BE in CH12F3-ΔSα1xIscel cells with Sμ in either physiological or inverted orientation. Error bars represent s.e.m. (B) The percentage (average±s.e.m.) of direct joins in either condition are plotted as bar graph. Information regarding replicates and number of junctions analyzed is mentioned in Table 2.3f. n.s.: not significant (two-tailed t test)
Table 2.2

(a) Average length of MH ± s.e.m. from rSµ-Sγ1 and rSµ-Sε junctions in various genotypes described in Figure 2.1.1 are listed. Numbers in parenthesis indicate total number of junctions mapped to the core Sγ1 or Sε regions in the indicated genotypes. Bait DSB used for HTGTS is mentioned on the top. (b) Average length of MH ± s.e.m. from ΔSγ1ΔIsceI-Sµ and ΔSγ1ΔIsceI-Sε junctions in various genotypes described in Figure 2.3.3 are listed. (c) Average length of MH ± s.e.m. from c-mycCas9-gRNA - genome-wide general DSBs and c-mycCas9-gRNA - Sµ/Sα junctions in various genotypes described in Figure 2.3 are listed. Numbers in parenthesis indicate total number of junctions mapped to the respective regions in the indicated genotypes. Bait DSB used for HTGTS is mentioned on the top.

<table>
<thead>
<tr>
<th>a</th>
<th>Δ5′rSµ</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Genotype</td>
<td>Core Sγ1</td>
<td>Core Sε</td>
</tr>
<tr>
<td>----</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td></td>
<td>WT</td>
<td>1.59±0.02 (n=10177)</td>
<td>1.39±0.02 (n=8493)</td>
</tr>
<tr>
<td></td>
<td>ATM−</td>
<td>2.32±0.02 (n=8670)</td>
<td>2.05±0.03 (n=3330)</td>
</tr>
<tr>
<td></td>
<td>53BP1−</td>
<td>3.23±0.09 (n=1663)</td>
<td>2.50±0.08 (n=640)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>b</th>
<th>ΔSγ1ΔIsceI 3′ BE</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Genotype</td>
<td>Core Sµ</td>
<td>Core Sε</td>
</tr>
<tr>
<td>----</td>
<td>-----------------</td>
<td>----------</td>
<td>----------------</td>
</tr>
<tr>
<td></td>
<td>ΔSγ1ΔIsceI WT</td>
<td>0.98±0.02 (n=7341)</td>
<td>1.08±0.04 (n=967)</td>
</tr>
<tr>
<td></td>
<td>ΔSγ1ΔIsceI ATM−</td>
<td>1.53±0.02 (n=14829)</td>
<td>1.24±0.03 (n=2016)</td>
</tr>
<tr>
<td></td>
<td>ΔSγ1ΔIsceI H2AX−</td>
<td>1.51±0.03 (n=3021)</td>
<td>1.54±0.06 (n=543)</td>
</tr>
<tr>
<td></td>
<td>ΔSγ1ΔIsceI 53BP1−</td>
<td>1.88±0.02 (n=4145)</td>
<td>2.14±0.05 (n=906)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>c</th>
<th>c-myc Cas9-gRNA 3′ BE</th>
<th>genome-wide junctions (remove chr15 and IgH)</th>
<th>Sµ + Sα junctions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Genotype</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>WT CH12F3</td>
<td>1.68±0.01 (n=15650)</td>
<td>1.45±0.02 (n=3948)</td>
</tr>
<tr>
<td></td>
<td>ATM− CH12F3</td>
<td>1.87±0.008 (n=51811)</td>
<td>2.00±0.03 (n=3774)</td>
</tr>
<tr>
<td></td>
<td>NCΔ CH12F3</td>
<td>1.60±0.01 (n=30065)</td>
<td>1.27±0.01 (n=5564)</td>
</tr>
<tr>
<td></td>
<td>NCΔ-53BP1− CH12F3</td>
<td>2.04±0.008 (n=57390)</td>
<td>2.71±0.03 (n=2149)</td>
</tr>
</tbody>
</table>
Table 2.3

(a-f) Information regarding replicates and number of junctions analyzed in each region is tabulated for various genotypes. Bait DSB used for HTGTS is mentioned on the top of each table.

### a 5′rSµ

<table>
<thead>
<tr>
<th>Genotype</th>
<th>biological replicates (n)</th>
<th># total junctions with direct and MH (1-10 bp) joins</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>core Sγ3</td>
</tr>
<tr>
<td>rSµ WT</td>
<td>3</td>
<td>1488</td>
</tr>
<tr>
<td>NCAΔ-CH12F3-rSµ</td>
<td>3</td>
<td>NA</td>
</tr>
</tbody>
</table>

### b 5′Sµ

<table>
<thead>
<tr>
<th>Genotype</th>
<th>biological replicates (n)</th>
<th># total junctions with direct and MH (1-10bp) joins</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>core Sγ1</td>
</tr>
<tr>
<td>Wild type</td>
<td>5</td>
<td>25103</td>
</tr>
<tr>
<td>ATM⁻/⁻</td>
<td>3</td>
<td>17161</td>
</tr>
<tr>
<td>H2AX⁻/⁻</td>
<td>3</td>
<td>3669</td>
</tr>
<tr>
<td>RIF1⁻/⁻</td>
<td>3</td>
<td>N.A.</td>
</tr>
<tr>
<td>53BP1⁻/⁻</td>
<td>8</td>
<td>8657</td>
</tr>
<tr>
<td>53BP1⁻/⁻ + ATMi</td>
<td>3</td>
<td>5751</td>
</tr>
</tbody>
</table>

### c ΔSγ1²d 3′BE

<table>
<thead>
<tr>
<th>Genotype</th>
<th>biological replicates (n)</th>
<th># total junctions with direct and MH (1-10bp) joins</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>core Sγ1</td>
</tr>
<tr>
<td>ΔSγ1²d WT</td>
<td>4</td>
<td>7341</td>
</tr>
<tr>
<td>ΔSγ1²d ATM⁺/⁺</td>
<td>3</td>
<td>14829</td>
</tr>
<tr>
<td>ΔSγ1²d H2AX⁺/⁺</td>
<td>3</td>
<td>3921</td>
</tr>
<tr>
<td>ΔSγ1²d 53BP1⁻/⁻</td>
<td>4</td>
<td>4145</td>
</tr>
</tbody>
</table>

Table 2.3 (Continued)
### Table 2.3 (Continued)

#### d  5'Sμ

<table>
<thead>
<tr>
<th>Genotype</th>
<th>biological replicates (n)</th>
<th># total junctions with direct and MH (1-10 bp) joins</th>
<th>core Sα</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT CH12F3</td>
<td>4</td>
<td>3066</td>
<td></td>
</tr>
<tr>
<td>ATM⁺ CH12F3</td>
<td>5</td>
<td>2366</td>
<td></td>
</tr>
<tr>
<td>NCA⁻ CH12F3</td>
<td>3</td>
<td>2569</td>
<td></td>
</tr>
<tr>
<td>NCA⁻-53BP1⁺ CH12F3</td>
<td>3</td>
<td>722</td>
<td></td>
</tr>
</tbody>
</table>

#### e  c-myc CRSPR-Cas9 3'BE

<table>
<thead>
<tr>
<th>Genotype</th>
<th>biological replicates (n)</th>
<th># total junctions with direct and MH (1-10bp) joins</th>
<th>genome-wide txns</th>
<th>core Sμ + core Sα</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT CH12F3</td>
<td>4</td>
<td>15650</td>
<td>3948</td>
<td></td>
</tr>
<tr>
<td>ATM⁺ CH12F3</td>
<td>6</td>
<td>51811</td>
<td>3774</td>
<td></td>
</tr>
<tr>
<td>NCA⁻ CH12F3</td>
<td>4</td>
<td>30065</td>
<td>5564</td>
<td></td>
</tr>
<tr>
<td>NCA⁻-53BP1⁺ CH12F3</td>
<td>6</td>
<td>57390</td>
<td>2149</td>
<td></td>
</tr>
</tbody>
</table>

#### f  ΔSα¹xlSce1 3'BE

<table>
<thead>
<tr>
<th>Genotype</th>
<th>biological replicates (n)</th>
<th># total junctions with direct and MH (1-10bp) joins</th>
<th>core Sμ</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔSα¹xl - Sμ (WT)</td>
<td>3</td>
<td>4834</td>
<td></td>
</tr>
<tr>
<td>ΔSα¹xl - Sμ (INV)</td>
<td>3</td>
<td>2738</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.4

(a-c) Statistical comparisons (two-tailed t test) of frequency of direct joins is presented (p value) for denoted junctions between wild type (WT) and different DSBR factor deficient backgrounds.

<table>
<thead>
<tr>
<th>a</th>
<th>Junctions</th>
<th>ATM⁺⁻</th>
<th>H2AX⁺⁻</th>
<th>53BP1⁺⁻</th>
<th>53BP1⁺⁻+ATMi</th>
<th>Rif1⁺⁻</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sµ-Sγ1</td>
<td>WT</td>
<td>0.0001</td>
<td>0.0022</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>Sµ-Sɛ</td>
<td>WT</td>
<td>0.0433</td>
<td>0.0026</td>
<td>&lt;0.0001</td>
<td>N/A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>b</th>
<th>Junctions</th>
<th>ATM⁺⁻</th>
<th>53BP1⁺⁻</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rSµ-Sγ1</td>
<td>WT</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>rSµ-Sɛ</td>
<td>WT</td>
<td>0.0023</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>c</th>
<th>Junctions</th>
<th>ATM⁺⁻</th>
<th>H2AX⁺⁻</th>
<th>53BP1⁺⁻</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ΔSγ1⁻²⁺⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻⁻ᵃ</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.4 Discussion

ATM-dependent DSBR factor deficiency promotes MH-biased CSR junctions

As discussed in chapter 1, we report that deficiency in the general ATM-dependent DSBR in CSR-activated B cells leads to increased long resections of S region DSBs with 53BP1-deficiency leading to highest level of long resections, while ATM and H2AX deficiency only modestly increase long resections\(^4\). Our HTGTS based studies reveals that CSR junctions between S\(\mu\) DSBs and DSBs located across the core S\(\gamma1\), S\(\varepsilon\) and S\(\alpha\) regions from ATM, H2AX, Rif1 and 53BP1-deficient cells are all MH-usage biased but to varying degrees.

With regard to ATM-deficiency, our result for mean MH length for S\(\mu\)–S\(\gamma1\) junctions (2.5bp vs. 1.7bp in WT, 129/SVEv strain), is in accord with a prior study with limited number of junctions, that reported a significant increase in MH length in mouse S\(\mu\)–S\(\gamma1\) junctions (2.6 bp vs. 1.2bp in WT, 129/SVEv strain)\(^{19}\), and also similar to findings of increased MH-usage in S\(\mu\)–S\(\gamma\) junctions in B cells isolated from A-T (ataxia telangiectasia) patients\(^{18}\). For H2AX- and Rif1-deficient B cells, we show a modest impact on CSR junctions profile, where similar to ATM\(^{-/-}\), the frequency of direct joins was reduced to ~20% (compared to 30% in wild-type), accompanied by a concomitant increase in proportion of junctions undergoing MH-biased joining. 53BP1-deficient B cells, which underwent highest frequency of S region DSB long resections, had an associated increase in junctions using microhomology (more than 90% compared to ~70% in wild type B cells), and displayed the most substantial decrease in the direct joining frequency of CSR junctions (S\(\mu\)-S\(\gamma1\) and S\(\mu\)-S\(\varepsilon\)). CSR junctions in 53BP1\(^{-/-}\) B cells displayed ~30% junctions using ≥ 4 bp MH, compared to ~10% normally seen in wild type B cells. Given that 53BP1-deficiency results in the most severe CSR defect, where only residual donor S\(\mu\) DSBs undergo “long-range” joining to the acceptor S regions in an orientation-unbiased manner, the small
number of these significantly MH-biased residual S-S junctions in 53BP1\(^{-/}\) B cells may be a subset of normal left-over low frequency junctions observed in wild type cells. We suggest that in DSBR factor-deficiency settings, S region break ends with exposed MH, such as those in repetitive S regions, are more prone to undergo joining MH-mediated end joining.

S region DSB resection and MH-usage in CSR junctions

Given S regions in the absence of DSBR response factors can be very long and AID-initiated DSBs locations within them are diverse, the frequency of short DSB resections contained within S region cannot be estimated by HTGTS (or other current approaches)\(^4\). In DSBR factor deficient backgrounds, like ATM\(^{-/}\) and H2AX\(^{-/}\), even though S region DSB long resections were only modestly increased\(^4\), we still observed a significant shift towards MH-biased joining of switch junctions. In this regard, it is possible that short resections of S region DSBs, undetected by our assay, could potentially cause MH-biased joining of S region DSBs.

We find a rescue of the frequency of direct joins in S\(\mu\)-S\(\gamma\)1 junctions and associated decrease in mean MH length in 53BP1\(^{-/}\) B cells treated with ATM inhibitor, which as shown in chapter 1, also led to inhibition of long S-region DSBs resections\(^4\). The CSR junction (S\(\mu\)-S\(\gamma\)1) profile in 53BP1\(^{-/}\)+ATMi cells, in terms of frequency of direct and MH-based joining, looked very similar to that of ATM-deficient B cells. We suggest a positive correlation between the elevated S region DSB ‘long’ resections and increased MH-usage in the core S region junctions from DSBR-deficient B cells. Detecting MH-usage of CSR junctions by HTGTS is not confounded by inability to map resections within core S-region. We speculate that increased MH-usage frequency of S-S junctions could potentially serve as a substitute indicator of S region DSB resections.
Unique aspects of junctions to S region DSBs

We found extensive genome-wide translocation junctions with HTGTS libraries cloned from a CRISPR-Cas9 blunt bait DSB at c-myc locus. We observed that for c-mycCas-9/gRNA blunt DSB – S region junctions, the frequency of direct joins in wild type CH12F3 cells is relatively higher (~40%), when compared to direct joins observed in wild type S-S junctions (~30%) in the context of CSR. Also, in wild type ΔSγ12xI B cells, I-SceI generated 3’overhangs bearing bait DSB at ΔSγ12xI location joining to S regions showed the frequency of direct joins to be relatively higher (~40%). In this regard, we speculate that unique features associated with AID-initiated S region break ends which may include opposing staggered nicks generated in combination with variable lengths of 5’/3’ overhang bearing break ends, potentially influence processing and/or repair pathway usage during end-joining, hence altering the junction structure. Alternatively, bait DSB structure (blunt end DSB in c-myc locus generated by CRSPR-Cas9, or 3’overhang break end generated by I-SceI in ΔSγ12xI location) might also influence the processing of break ends or repair pathway usage to alter junction structure profile.

Translocation junctions between Cas-9/gRNA-induced blunt DSB in c-myc to AID initiated S region (Sμ and Sα) DSBs in ATM- or 53BP1-deficient CH12F3 cells show substantial biases towards MH-usage (mean MH length for ATM−/− 2.0 bp vs. 1.4 bp in control, p<0.0001; for 53BP1−/− 2.7 bp vs. 1.3 bp in control, p<0.0001), similar to the extent of MH-biased joining observed in the context of S-S CSR junctions (Table 2.1). However, surprisingly translocation junctions to other general genome-wide DSBs showed a very small shift towards MH-usage in ATM-deficient cells (p<0.0001) (mean MH length 1.9 bp vs. 1.7 bp in control) and only a modest increase in MH-biased joining in 53BP1-deficient cells (mean MH length 2.04 bp
vs. 1.7 bp in control, p<0.0001). If increased MH usage indeed reflects resection, our results could imply that S region DSBs are more prone to resection than other genome-wide breaks. Such a potential sensitivity of AID initiated DSBs in S regions to resections could result from many different factors. One possibility is that AID-initiated staggered nicks along a given S region could bypass the requirement of the initiation step of resection. A related possibility is that the high frequency AID-initiated DSBs in S regions could also be a contributing factor in increased resection. Another model for explaining MH-biased joining of S region DSBs in DSBR deficient backgrounds could be that S regions are specialized substrates, which due to their long highly repetitive nature, upon resection could expose robust microhomologies that provide substrates for MH-biased joining.

53BP1-deficiency, while showing minimal general genome-wide instability, displays substantial IgH locus specific break burden and translocations in CSR-activated B cells\textsuperscript{28}. In this regard, the extreme resections of S region DSBs that we find in 53BP1-deficient cells, if reflective of the increased MH usage in S region junctions compared to general DSB junctions, might potentially lead to decreased C-NHEJ and, thereby, underlie IgH-specific DSB and translocation phenotype.
2.5 References


DISCUSSION
This dissertation sheds light on the mechanisms of joining AID-initiated S region DSBs during IgH class switch recombination. The major findings of my thesis are presented in the form of two chapters: 1) Discovery and mechanisms of deletion orientation-biased joining of S region DSBs during CSR. 2) Structure and mechanism of S-S junctions during CSR and its regulation by ATM-dependent DSBR factors. Discussion for our findings in chapter 2 is presented in the discussion section of chapter 2. Below, we highlight major conclusions from chapter 1 and present an extended discussion.

Chapter 1: Discovery and mechanisms of deletion orientation-biased joining of S region DSBs during CSR

B cell selection does not lead to orientation-specific S region DSB joining

We demonstrate that CSR is mechanistically programmed to occur in a productive, deletional orientation [1]. We observed that joining between two I-SceI DSBs in different IgH S-region locations in CSR-activated B cells lacks any notable preference for or against inversional versus deletional joins [1]. Equal recovery of deletion and inversional joining between two I-SceI generated DSBs argues against any junction recovery bias introduced by potential negative selection of B cells which underwent inversional joining to yield a non-productive CSR event, which is incapable of producing a functional BCR. In addition, we also performed time-course experiments with stimulated NCA-rSµ CH12F3 cells harvested at 48 hrs and 72 hrs post-stimulation, HTGTS libraries cloning from bait rSµ DSBs indicated similar joining bias (about 21-fold) to deletional orientation in Sµ-Sα joining at both time-points [1], thus also negating the possibility of B cells selection yielding the orientation-specific joining phenomenon.
Robust IgA CSR in S\(\mu\)(INV) cells

Owing to the specialized strand-biased sequence composition of G-rich sequences on the non-template strand of S regions, transcription through S regions creates R-loops whereby, template strand binds the nascent RNA transcript while looping out the non-template strand as ssDNA [2-5]. However, inversion of endogenous S region experiments show that R-loop formation is not absolutely required for S-region function in CSR [4]. In accordance with previous literature, we now show that inversion of S\(\mu\) sequence in IgH locus on the productive allele in NCA-CH12F3 cells, yielded substantial IgA CSR, albeit at modestly reduced levels of \(~60\%\) of the control NCA-CH12F3 cells [1]. Recently, a provocative study showed that spliced intronic S region RNA forms G-quadruplex structure which binds AID and recruits it to S region in the IgH locus [6]. However, in our experimental system, S\(\mu\)(INV) sequence derived spliced intronic RNA would be C-rich and thus would be not be expected to form G-quadruplex structure capable of targeting AID to S\(\mu\) region \textit{in trans}. In addition, in our system, we have deleted the switched chimeric S\(\mu\)-S\(\alpha\) sequence on the non-productive IgH allele, thus ruling out any contribution of potential G-quadruplex structure forming spliced intronic RNA from non-productive allele IgH locus. Given our finding of reduced, yet robust levels of IgA CSR observed in NCA S\(\mu\)(INV) CH12F3 cells, we suggest presence of intronic switch RNA G-quadruplex structure independent mechanisms for AID-targeting to S region during CSR.

“Long-range” S region synapsis: role of IgH locus structure and S region DSBs

Based on our findings, we have proposed a working model for orientation-specific CSR, in which a key component is the organization of S-regions within topologically-associated domains (TADs) that promote their frequent S-region synapsis [7-9] via Langevin motion [8-10].
We further implicate additional IgH-specific organizational features, not yet fully elucidated, in playing a fundamental role in mediating synapsis in an orientation that promotes deletional joining. This proposal is in accordance with our lab’s previous results where, I-SceI–generated DSBs at a pair of I-SceI target sites replacing endogenous Sμ and Sγ1, respectively, supported I-SceI meganuclease–dependent IgM-to-IgG1 recombinational class switching at robust levels [7, 9]. This suggested that sufficiently frequent DSBs at donor and acceptor S regions can drive potential physiological CSR levels by leveraging general aspects of three-dimensional chromatin organization and general DNA DSB response and repair mechanisms [7, 9].

HTGTS analysis of joining of the ISce-I generated bait ΔSμ2xI-broken ends to an array of DSBs at 28x I-SceI sites replacing Sγ1 region, indicated orientation-independent joining [1]. Thus in addition to IgH locus organizational features, our current study highlights the requisite role of S regions in mediating orientation-specific CSR joining, potentially associated with their ability to promote AID-initiated DSBs, multiple frequent DSBs, or both. We do not rule out any potential specialized “long-range” synapsis functions of S regions or AID and/or associated complexes in mediating orientation-specific CSR joining. However, we show that AID-initiated DSBs per se are not sufficient to promote joining orientation specificity, as demonstrated by orientation-independent joining of ISce-I induced DSB in c-myc gene on chr15 to S region DSBs on chr12 in trans [1]. Thus, beyond S-region sequences and/or high frequency AID initiated DSBs within them, aspects of IgH locus organization in cis also play a critical role in promoting orientation-dependent CSR joining.
Role of ATM-dependent DSBR factors

Our studies implicate DSBR factors in enforcing deletion orientation-specific CSR joining. Previous work has established a role of ATM-dependent DSBR factors including ATM, H2AX and 53BP1 in promoting end-joining of S region DSBs in activated B cells, where, similar to C-NHEJ factors deficiency, absence of ATM-dependent DSBR factors led to accumulation of unrepaired IgH chromosome breaks along with IgH locus derived translocations [11-13]. We propose that the broader DSBR contributes to orientation-specific CSR by tethering S region DSBs for efficient re-joining, thus keeping them from separating into chromosomal breaks that could frequently translocate and join with orientation independence to S-region broken ends within the TAD [8, 11]. Such tethering function of general DSBR factors would also help promote re-joining of unsynapsed S region DSBs via C-NHEJ to allow subsequent AID-initiated breakage and joining to a synapsed S region leading to deletional orientation-specific productive CSR joining.

Our HTGTS based assay using 5'Sµ bait DSBs joining to AID-initiated DSBs generated along the entire length of downstream S regions, offers a unique capability of assaying tens-of-thousands of S region DSBs joining events and their distribution pattern along the long repetitive S regions. Application of 5'Sµ HTGTS assay in various ATM-dependent DSBR factor-deficient backgrounds showed that frequency of prey junctions in various acceptor S regions cloned from 5'Sµ bait DSBs roughly reflects the IgH class switching frequency observed by FACS staining.

Deficiency in DSBR factors leads to varying reductions in CSR, and resection of AID-initiated S region DSBs has been proposed to be at least partly accountable for associated CSR defects [14-17]. In addition, ATM-dependent DSBR factors including H2AX and 53BP1 by antagonizing 5’-3’ resection of DNA ends, have been implicated in influencing repair pathway
choice by promoting repair of DSBs by the C-NHEJ pathway versus HR or A-EJ [14-18]. However, given technical limitations, directly measuring resection within endogenous S region DSBs is not feasible as S regions are long and AID-initiated DSBs locations within them are diverse; hence we estimated relative resection by quantifying bait broken end to prey broken end junctions downstream of acceptor S-region positions where the incidence of wild-type junctions decreases to background. This provides us a single nucleotide resolution ‘long’ S region DSBs resection assay, which also has the potential to reveal the functional impact of resection on the junctional nature of S region DSB joining.

We found that deficiency for various general DSBR factors in CSR-activated B cells leads to increased S region DSB resections with 53BP1-deficiency leading to the highest level of long S region DSBs resections, followed by Rif1-deficiency causing second highest long S region DSBs resection, while ATM and H2AX deficiency only modestly increasing S region DSBs long resections [1]. We extended application of our 5’Sμ HTGTS assay in Rif1+/− B cells, where Rif1-deficiency in B cells results in a substantial decrease in CSR, however it is not as severe as that associated with 53BP1-deficiency. Correspondingly, the distribution of Sμ-Se junction pattern from Rif1+/− (CD19cre Rif1fte) B cells had a relatively higher proportion of deletional junctions mapping to the Se region, compared to “flattened” (described in chapter 1, Fig. 1.3) Se junctions pattern observed in 53BP1+/− B cells [1]. These profiles imply that S region DSBs in absence of Rif1, relatively undergo higher frequency of comparatively normal joining than 53BP1-deficient B cells, potentially reflecting intact S region synapsis promoted by the presence of intact 53BP1 in Rif1+/− B cells. Thus, the main similarity with 53BP1+/− B cell junctions is the small subset of junctions extending beyond S regions, probably reflecting extensive resection of the smaller subset of broken ends that are not rapidly fused.
We propose that DSBR factors prevent long end-resections that could cause S-region broken ends to linger in resection complexes, preventing synapsis with other S-region broken ends and/or diminishing ability to be joined by C-NHEJ. Different DSBR factors have a differential impact in tethering versus resection inhibition and, thus, may impact orientation dependence via different routes. Previous studies have shown that inhibition of ATM or its substrate CtIP, partially inhibited resection and partly rescued class-switching defects in H2AX- or 53BP1-deficient B cells [14, 15, 17]. In this regard, treatment of 53BP1−/− cells with ATMi (ATM kinase inhibitor) led to significant reduction in S region DSB ‘long’ resections, where the long resection level was reduced to that observed in ATM-deficient B cells. Although treatment of 53BP1-deficient activated B cells with ATM inhibitor substantially diminished long Sγ1 region DSBs resections, it did not restore orientation-dependent joining in Sμ-Sγ1 junctions. This finding may reflect potential increased shorter resections within Sγ1 region in ATMi-treated 53BP1-deficient versus ATM-deficient B cells that are not revealed by our long resection assay. We suggest antagonistic functions of ATM in regulating resection in CSR settings, where ATM deficiency inhibits resection by impairing CtIP activation [19], but promotes resection via other nucleases by impairing inhibitory activities of H2AX, 53BP1 and, indirectly, Rif1 [20, 21].

53BP1-deficiency is unique in that it both impairs tethering for rejoining and activates resection of unjoined ends by failure to activate Rif1, leading to extreme resections and the greatest impairment of CSR and orientation-dependent joining. In addition, 53BP1 could also have a putative specialized role for 53BP1 in stabilizing synapsed S regions [22]. As common and unique impacts of 53BP1 deficiency markedly affect both donor and acceptor S-regions, they would be multiplicative and, thereby, explain the profound impact of 53BP1-deficiency on CSR.
References


METHODS
METHODS

M.1 Chapter 1

M.1.1 Mice

$Igh^{1-96k} AID^{-/-}$ (ref. 1), $\Delta S\mu^{2d}/\Delta S\gamma^{2d}$ chimaera, $\Delta S\gamma^{2d}$-SceI (ref. 2), $S\mu^{4A}$ (ref. 3), $c-myce^{25d}$-SceI (ref. 4), $ATM^{-/-}$ (ref. 5), $H2AX^{-/-}$ (ref. 6), $53BP1^{-/-}$ (ref. 7), $Rif1^{+/+}CD19^{cre}$ (ref. 8) lines have been reported previously. Mouse work was performed under protocols approved by the Boston Children’s Hospital and the Rockefeller University Institutional Animal Care and Use Committees.

M.1.2 Plasmids and oligonucleotides

Oligonucleotides for gRNAs for CRISPR/Cas9-mediated targeting of various $Igh$ regions have been cloned into pX330 vector (Addgene plasmid ID 42230) as described. Exchange vector (pLH28) with heterologous loxP sites was obtained from Kefei Yu. A 200-bp GFP-derived sequence was amplified and ligated to an I-SceI recognition sequence and subsequently introduced into the pLH28 vector to make the pLH-1× I-SceI exchange vector. To obtain the I-SceI expression plasmid for transducing CH12 cell lines, I-SceI-IRES-GFP fragment was shuttled from a retroviral construct (pMX-I-SceI-IRES-GFP) into pCDNA3.0 (Invitrogen) vector.

M.1.3 B cell culture, transduction and FACS analysis

Mature splenic B cells isolated by using CD43-negative selection kit (MACS), were cultured in lymphocyte medium R15 (RPMI1640, 15% FBS, L-glutamate, 1× penicillin and streptomycin). B cell stimulation was performed with anti-CD40 (1 $\mu$g ml$^{-1}$, eBioscience) plus IL4 (20 ng ml$^{-1}$,
PeproTech) or LPS (25 ng ml$^{-1}$, Sigma) plus anti-IgD-dextran (3 ng ml$^{-1}$, a gift from R. Casellas) for 96 h. Infection with I-SceI expression or control retrovirus was carried out at day 1 post-stimulation by the standard spinning method with the presence of 4 μg ml$^{-1}$ polybrene as previously described$^{10}$. Efficiency of retrovirus infection and switching levels were evaluated by flow cytometry as previously described$^{10}$. Where indicated, ATM inhibitor KU-55933 (Tocris) was added to stimulated cells at day 1 post-stimulation to a final concentration of 10 μM and was maintained during the course of the experiment until harvesting of the cells for FACS and HTGTS libraries.

M.1.4 Cell lines and nucleofection

CH12F3 cell line stimulation to IgA was done as described$^{11}$. CH12F3 cells with recombinase mediated cassette exchange (RMCE) in place of the endogenous Sα region, referred to as 1F7 cells$^{11}$ were maintained at 37 °C, 5% CO$\_2$ and cultured in RPMI media with 10% FCS, 0.5% penicillin/streptomycin, 50 μM β-mercaptoethanol. Exchange vector with heterologous loxP sites containing 1× I-SceI site embedded in 200 bp of GFP-derived sequence was cloned. RMCE was performed as previously described$^{11}$. Exchanged $\Delta S\alpha^{\mu}$ clones were verified by PCR, Sanger sequencing and Southern blotting. $\Delta S\alpha^{\mu}$ cells were then stimulated with anti-CD40, IL4 and TGF-β for 15 h followed by nucleofection with pcDNA-I-SceI-IRES-GFP expression vector using 4D-nucleofector X (Lonza, solution SF, protocol CA-137) and re-plated in stimulation-conditioned media. On day 3 post-stimulation cells were harvested and gDNA was isolated for HTGTS library preparation.

To obtain CH12F3 (productive allele $S\mu(INV)$, non-productive allele $\Delta S\mu-S\alpha$) cells, wild-type CH12F3 cells were first nucleofected using the 4D-nucleofector X (Lonza, solution SF)
protocol CA-137) with the gRNA vectors to excise the sequences between JH4 intron and ~130 bp downstream of Cα polyadenylation on the non-coding allele that has already switched to Sα. Single-cell subclones were seeded into 96-well plates 12 h post-nucleofection, and the resulting clones were screened by PCR and Southern blot. One confirmed positive clone was further modified by gRNA vectors targeted at the 5’Sμ_1 and 3’Sμ regions to invert the Sμ (~4 kb) sequence. Initial screening for positive clones was done by PCR, followed by Southern blotting and Sanger sequencing for the inversion junction. The resultant cells were stimulated with anti-CD40, IL4 and TGF-β, IgA CSR was measured by FACS on days 2 and 3 post-stimulation. ΔSα<sup>↓↓</sup> Sμ<sup>(INV)</sup> cells were obtained by targeting the aforementioned 1 × I-Sce1 RMCE-positive cells with gRNA targeting 5’Sμ_2 and 3’Sμ for inverting the Sμ sequence same as above. The resultant positive clones were verified by PCR, Southern blotting and Sanger sequencing for the inversion junction. To make rSμ-CH12F3 cells, the aforementioned CH12F3 (non-productive allele ΔSμ-Sα) cells were used to further truncate Sμ sequences on the coding allele with gRNA targeting 5’Sμ_2 and 3’Sμ. Single-cell deletion subclones were screened and confirmed by PCR and Southern blot. The resultant rSμ-CH12F3 cells were stimulated with anti-CD40, IL4 and TGF-β and harvested on days 2 and 3 for gDNA isolation for HTGTS library preparation.

**M.1.5 DC–PCR**

The DC–PCR assay was performed as described previously<sup>12</sup>. In brief, genomic DNA was isolated and subsequently purified by phenol chloroform extraction from day 4 anti-CD40/IL4 stimulated B cells. Five micrograms of genomic DNA was digested overnight with 20 U of EcoRI (Roche). Ligations were performed under diluted conditions to promote circularization.
Digested DNA was ligated overnight at 16 °C with a concentration of 1.8–9 ng μl⁻¹ in a total volume of 100 μl per reaction. Three to four ligation reactions were pooled, column purified, concentrated and serially diluted at a 1:5 ratio. PCR was then performed in 50 μl per reaction using 2.5 U Taq (Qiagen) with serially diluted DNA starting from ~50–150 ng. Primers were designed to amplify the Sµ-Sγ1 rearrangements that occur during CSR to IgG1 in direct chromosomal joining of Sµ-Sγ1 with excision of circular DNA or inversion of sequences between broken ends of Sµ and Sγ1. As a control for EcoRI digestion and circularization of input DNA, amplification of an EcoRI fragment of nicotinic acetylcholine receptor B subunit gene (CHRNB1) was performed, which, after EcoRI digestion and circularization, generates a 753-bp DC–PCR product. To quantify the amount of direct or inversion joins amplified by PCR, DC–PCR products of direct or inversion joins were cloned into the pcR2.1 Topo TA vector. Precise plasmid concentrations were determined and a standard curve was generated ranging from 4 to 10,000 copies per reaction. After running on 1% agarose gel, PCR fragments were transferred to nitrocellulose membrane and hybridized to a 3′Sγ1 probe according to standard Southern blotting procedures. Primers for direct joining PCR: forward, 5′-
CATGAGAGCTGGAGCTAGTATGAAGGTG -3′; reverse, 5′-
ACTGACTGACTGATGTGCTCTCAAC-3′. Primers for inversional joining PCR: forward, 5′-
CAGTCACAGAGAAACTGATCCAGGTGAG -3′; reverse, 5′-
CCATAGCAGTTGGTCAATCCTTGTCTCC-3′. Primers for control CHRNB1 DC–PCR: forward, 5′-GCGCCATCGATGGACTGCTGTGGGTTTCACCCAG-3′; reverse, 5′-
GGCCGGTCGACAGGCGCGACTGACACCCTAAG-3′. Oligonucleotide probe for the detection of both deletional and inversional CSR joining products: Sγ1-
CCTGGGTAGGTTACAGGTCAGGCT.
M.1.6 High-throughput genome-wide translocation sequencing (HTGTS)

HTGTS libraries were generated by emulsion-mediated PCR (EM–PCR) and linear-amplification-mediated PCR (LAM–PCR) methods as described in chapter 1 ref. 5 In brief, sonicated (Bioruptor, Diagenode) gDNA was subjected to LAM–PCR using 1 U Taq polymerase (Qiagen) per reaction with a single biotinylated primer for 50 cycles of 94 °C for 180 s; 94 °C for 30 s; 58 °C for 30 s; 72 °C for 90 s. One more unit of Taq polymerase was added to the reaction mixture to execute PCR for additional 50 cycles. Biotinylated DNA fragments were captured with Dynabeads MyOne streptavidin C1 beads (Invitrogen) at room temperature for 1 h, followed by on-bead ligation at 25 °C for 2 h with bridge adapters in the presence of 15% PEG-8000 (Sigma) and 1 mM hexammine cobalt chloride (Sigma). After washing beads with B&W buffer as described by the manufacturer, ligated products were subjected to 15 cycles of on-bead PCR with Phusion polymerase (Fisher), locus-specific and adaptor primer followed by blocking digestion with appropriate restriction enzymes to remove uncut germline gDNA. A third round of tagging PCR to add Illumina Miseq-compatible adapters at 5′ and 3′ ends of the second-round PCR product was carried out for another 10 cycles with Phusion polymerase. PCR products were size-fractionated for DNA fragments between 300–1000 bp on a 1% agrose gel, column purified (Qiagen) before loading onto Illumina Miseq machine for sequencing.

M.1.7 Data analyses

Data analysis of MiSeq sequencing reads has been described in chapter 1 ref. 5. In brief, de-multiplexing for the MiSeq reads was performed using the fastq-multx tool from ea-utils (https://code.google.com/p/ea-utils/) and adaptor sequence trimming was done using the SeqPrep utility (https://github.com/jstjohn/SeqPrep). Reads were mapped using Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml) to either mm9 (for libraries generated with Rif1
knockout cells and CH12F3-derived cells) or modified mm9 reference genome (for all other genotypes) containing the 176-kb *Igh* constant region of 129S genome, in which the region between chr12:114493849-114665808 of mm9 was replaced with DNA sequence ranging 1416975-1593283 on the 129S *Igh* reference sequences AJ851868.3. In cases where necessary, for instance when aligning reads to the *Sμ*<sup>1-3μ</sup> locus on the *Igh<sup>1-96k</sup>* allele and other circumstances, we further modified the custom 129S_IgHC genome to insert the cassette sequences to accurately reflect the changes of genomic information before aligning MiSeq reads by Bowtie2. CH12F3 clone was derived from CH12.LX lymphoma cell line<sup>13</sup>. CH12.LX cells were subcloned from the original CH12 lymphoma cell line<sup>14</sup>, which originated from a C57BL/10 mouse substrain double congenic for H-2<sup>a</sup> H-4<sup>b</sup> (ref. 15). C57BL/10 and C57BL/6 are both substrains of C57BL and thus we use BL/6 (mm9) as reference genome when running our HTGTS data analyses pipeline on libraries made with CH12F3 cells. To reflect additional genome modifications (for example, *Sμ*(INV) shown in Fig. 1.2b), the mm9 genome sequence was modified accordingly.

A best-path searching algorithm (based on YAHA read aligner and break point detector<sup>16</sup>) was used to select optimal sequence alignments from Bowtie2-reported top alignments with an alignment score above 50, which represents a perfect 25-nt local alignment. To avoid detecting possible mis-priming events, we set a bait alignment threshold of at least ten perfectly aligned nucleotides extending from the end of cloning primer. Aligned reads were subsequently filtered on following criteria: (1) reads must include both a bait alignment and a prey alignment; and (2) the bait alignment cannot extend more than 10 nt beyond the targeted site. For reads mapped to the repetitive low-mappability regions, multiple competing alignments with identical or similar scores exist and the coordinates for best alignment are randomly chosen.
among the competing ones. For junctions mapped to each individual repetitive S-region, there are no competing alignments from outside of that region as shown by simulation (see details below), although the exact junction coordinate within the region could not be identified. We also applied filter to remove duplicates (referred to as ‘de-dup’ hereafter) wherein the coordinates of the end of the bait alignment were compared to the start of the prey alignment across all reads. A read is marked as a duplicate if it has bait and prey alignment coordinates within 2 nt of another read’s bait and prey alignments. To plot all the S-region junctions, we took the ones filtered by a mappability filter but unequivocally mapped to S-regions and removed the repeats through the de-dup program mentioned above, before combining with ‘good’ reads passing both the mappability and de-dup filters. A grey box over S-regions (for example, Sμ and Sγ1) in the figures is used to denote the repetitive regions in these S sequences wherein the randomly assigned mappability-filtered reads were included. Additionally we applied post-filtering stringencies to remove junctions mapped to simple sequence repeats, telomere repeats and reads with excessive microhomology >20 nt and insertions >30 nt before further analysis. In the end, the combined and cleaned junctions were then plotted genome-wide or onto desired S-regions by using the PlotRegion tool (for details see section below).

Scripts and details of pipeline parameters are available upon request.

### M.1.8 S-region junction plotting

As described above, junctions filtered by the mappability filter are retrieved and de-duped before combining with normal junctions. To plot junction coordinates onto individual S-regions or the entire *Igh* constant region, combined junctions are binned using the PlotRegion tool into 100 bins (bin size varies depending on the length of target region that libraries are plotted to) on the basis of the junction coordinates and orientation of joining. The bincount file (histogram information
for junction distribution in both joining orientation) generated by the PlotRegion tool is used to calculate the percentage of junctions in each bin in either + or – orientation of the total number of junctions mapped to the region of interest. The results were then plotted as linear graph by the Prism software. Note that the scale on top of each graph is to indicate the size of region plotted and it is fixed as 1/10 of the size of the plotted region, thus is always 10× bin size.

M.1.9 Calculation of joining orientation bias and acceptor S-region resection

For simplicity, joining from 5′Sµ to downstream Sγ1 and Sε breaks are used for the explanation of orientation bias and resection of acceptor S-region DSBs. Junctions mapped to Sγ1 and Sε can be divided into six regions (denoted by a–f) in either + or – orientation:

\[ \begin{align*}
    & a \mid b \mid c \text{ (})+\text{ (}) \\
    & d \mid e \mid f \text{ (})–\text{ (})
\end{align*} \]

Junctions encompassing core Sγ1/Sε are illustrated as b and e regions for – and + junctions respectively, c region (deletional joining, – orientation) or d region (inversional joining, + orientation) represent joining of bait DSB broken ends to resected acceptor Sγ1/Sε DSBs. Junctions falling into regions a or f represent joining to non-AID-generated de novo breaks of unknown source and are often very small in number, and thus were omitted from the calculation of both resection and orientation bias. Since in most genetic backgrounds other than 53BP1−/− inversion joins are much rarer than deletions, the level of resection junctions into the d region fluctuates much more than resection junctions into the c region. We thus chose the c region for calculating resection in all genotypes as follows:

\[ \text{resection rate} = \frac{c}{b + c} \times 100 \]
The degree of orientation bias, for the purpose of positively correlating with the level of resection, is calculated as the ratio of inversionsal joins versus deletional joins as below:

$$\text{bias ratio} = \frac{d + e}{b + c} \times 100$$

To make a bar graph for comparison of orientation bias degree and resection levels in the CSR junctions obtained from libraries with different genetic backgrounds, individual replicate HTGTS libraries were first size-normalized to the one with smallest junction number in the region of interest among the replicates; resection and bias ratio values from individual experiments were calculated separately and averages were used for statistical analysis with unpaired two-tailed $t$-tests. Experiments for each genotype were performed for at least three times.

**M.1.10 Pipeline simulation for S region mappability**

To evaluate the mappability of the switch regions and the effect of junctions mapped to repetitive region on orientation-specific joining, simulated sequencing reads containing 5’Sμ CSR junctions cloning from the 5’BE primer were generated. For this purpose, 20,000 reads from each 129s acceptor S region were simulated with each assigned with a 5’Sμ bait part starting from the first nucleotide of the primer and a random length ranging 40-100bp; the prey junctions within each individual switch region were generated by randomly picking a nucleotide position in the region and a strand (in either + or – orientation) with a read length of 40-200bp. Mutations were randomly introduced at a rate up to 0.002 to account for physiological mutation, PCR amplification-derived mutation and sequencing errors. Simulated read libraries were then run through our HTGTS pipeline to identify joining junctions, which are then plotted against the whole genome and along the entire IgH constant region to evaluate the ability of the pipeline to
correctly map them. In theory the randomly simulated junctions should be only mapped to the specific S region where they are generated from and distribute evenly across the entirety of that region with equal densities in (+) and (-) orientations. As shown in the table below, >99% of all randomly simulated HTGTS reads from each individual S region can be correctly mapped to the specific S regions that they are generated from. The very rare mis-mapped reads are often of short length and with “mutations”; the genomic regions where these reads are mis-mapped to are usually simple repeats that resemble short S region repeat sequences.

**Accuracy of mapping simulated S region junctions:**

<table>
<thead>
<tr>
<th>Sγ3</th>
<th>Mapped correct (%)</th>
<th>Mis-mapped (%)</th>
<th>Genome-wide</th>
</tr>
</thead>
<tbody>
<tr>
<td>98.8</td>
<td>0.1</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>Sγ1</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sγ2b</td>
<td>99.9</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>Sγ2a</td>
<td>99.8</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>Sε</td>
<td>99.9</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>Sα</td>
<td>99.3</td>
<td>0</td>
<td>0.7</td>
</tr>
</tbody>
</table>

**Figure M.1:** Mapping of simulated reads from various switch regions.

To assess the mappability of each S region, simulated reads from each individual S region that were filtered out by the mappability filter are shown in the plots above as grey areas in both joining orientation and compared with the “good” reads that pass all filters, shown in the...
plots above as deletion (- orientation) or inversive (+ orientation) joining. Across all S regions, junctions from simulated reads evenly distribute in (+) and (-) orientation with “valleys” indicating low mappability regions, the junctions mapped to which adopt similar even distribution pattern. Different S regions have various degrees of repetitive sequences of lower mappability as reflected by the relative amounts of junctions filtered out from mapping. Among these, Sγ1 contains the longest repetitive region in the center and thus highest percentage of junctions filtered out by the mappability filter. Whereas Sγ3, Sγ2b and Sα regions have intermediate level of repetitive sequences, Sγ2a and Sε sequences are least repetitive and thus very mappable in general. When adding junctions filtered by mappability filter back, the total junctions evenly distribute along the whole regions including the repetitive, less mappable cores. Taken together, our HTGTS pipeline is capable of correctly mapping CSR junctions to the repetitive sequences inside S regions without built-in bias against or for either orientation; by adding back low-mappability junctions in we can reliably measure junction distribution in the repetitive S regions.

M.1.11 Normalization of deletional versus inversive joining between a fixed single I-SceI break to AID-initiated Sμ breaks by “symmetrical joins”

To assess whether or not the orientation-unbiased Sμ junctions in 5’-BE libraries from ΔScel CH12F3 cells is due to the less recovery of excision circles, we compared the deletional junctions from 3’-BE libraries and excision circle junctions from 5’-BE libraries by normalizing them with “symmetrical joins” that are “shared” between two libraries. The rationale is that in a certain cell population that repairs a single fixed DSB, there exists a population of junctions capable of being captured by primers amplifying from either end of the DSB. In other words,
there are junctions in one library that contain prey sequences that are suitable bait in the same library cloning from the opposite direction. We designate this group of junctions as “symmetrical junctions” (diagramed in the left panel below) which should by definition be mapped to the region restrained by the two cloning primers from either end in each library. By normalizing either a 5’-BE or 3’-BE library with the number of symmetrical joins it contains, frequency of joining from a single fixed I-SceI break to certain selected regions in this library can be directly compared to that of its paired library generated from the other broken end of the same DSB. Based on the normalization as shown in the right panel below, we discovered that for each symmetrical junction identified from the 3’-BE libraries, we identified about ~13 junctions joining to deletional orientation and ~3 junctions joining in inversion orientation (thus a biased degree of Del:Inv=4:1), while for each symmetrical junction identified, we detected about ~4 excision circle junctions and ~4 inversion junctions from the 5’-BE libraries. We thus conclude that the low recovery for excision circle may account for the unbiased broken end usage in 5’-BE libraries from ΔSα<sup>1-ISceI</sup> CH12F3 cells.

Figure M.2: Schematic illustration for HTGTS normalization based on ‘symmetric’ junctions. Normalization of HTGTS libraries from ΔSα<sup>1-ISceI</sup> CH12F3 cells.
### M.1.12 DNA oligos

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence 5' -&gt; 3'</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bio-TEL-ΔSγ12dT</td>
<td>/5BiosG/TAGAAGGCCGCTCTTTTGC</td>
<td>ΔSγ12dT 3'-BE, first round bio-PCR</td>
</tr>
<tr>
<td>RED-TEL-ΔSγ12dT</td>
<td>GCAGGAAATTGATATACGAAGCT</td>
<td>ΔSγ12dT 3'-BE, second round PCR</td>
</tr>
<tr>
<td>Bio-CEN-ΔSγ12dT</td>
<td>/5BiosG/GGAATATACGAGAAGCTGAGG</td>
<td>ΔSγ12dT 5'-BE, first round bio-PCR</td>
</tr>
<tr>
<td>RED-CEN-ΔSγ12dT</td>
<td>GCCCTGAGGGACCTAATAAC</td>
<td>ΔSγ12dT 5'-BE, second round PCR</td>
</tr>
<tr>
<td>3'-Bio-96k</td>
<td>/5BiosG/TAGAAGGCCGCTCTTTTGC</td>
<td>Sγ1 3'-BE, first round bio-PCR</td>
</tr>
<tr>
<td>3'-RED-96k</td>
<td>TAACGGTCTAAGGGACGAGC</td>
<td>Sγ1 3'-BE, second round PCR</td>
</tr>
<tr>
<td>5'-Bio-96k</td>
<td>/5BiosG/GCTTCAATTGTCAGATTG</td>
<td>Sγ1 5'-BE, first round bio-PCR</td>
</tr>
<tr>
<td>5'-RED-96k</td>
<td>CCAGGTATAGGTTAAACGTG</td>
<td>Sγ1 5'-BE, second round PCR</td>
</tr>
<tr>
<td>3'-Bio-ΔSα12d</td>
<td>/5BiosG/TGACGATTGCTTTTATCAATTCTCG</td>
<td>CH12F3-ΔSδ12d 3'-BE, first round bio-PCR</td>
</tr>
<tr>
<td>3'-RED-ΔSα12d</td>
<td>GCAGGAAATTGATATACGAAGCT</td>
<td>CH12F3-ΔSδ12d 3'-BE, second round PCR</td>
</tr>
<tr>
<td>5'-Bio-λα</td>
<td>/5BiosG/CCTGCCGTGGTCGGGGATG</td>
<td>CH12F3-ΔSδ12d 5'-BE, first round bio-PCR</td>
</tr>
<tr>
<td>5'-RED-λα</td>
<td>CCATCTGGAGTCTCGGTCG</td>
<td>CH12F3-ΔSδ12d 5'-BE, second round PCR</td>
</tr>
<tr>
<td>5'-Bio-μμ</td>
<td>/5BiosG/CAGACCTGGAATGTATG</td>
<td>CH12F3/rSμ-5'BE, first round bio-PCR</td>
</tr>
<tr>
<td>5'-RED-μμ</td>
<td>CACACAAAGACTCGGACTG</td>
<td>CH12F3/rSμ-5'BE, second round PCR</td>
</tr>
<tr>
<td>3'-Bio-μμ_rSμ</td>
<td>/5BiosG/GTCGATGTCTAAATTCTCAAGGA</td>
<td>rSμ-3'BE, first round bio-PCR</td>
</tr>
<tr>
<td>3'-RED-μμ_rSμ</td>
<td>AATGTGAGTGACCAAGAA</td>
<td>rSμ-3'BE, second round PCR</td>
</tr>
<tr>
<td>3'-Bio-μμ_ΔSμ</td>
<td>/5BiosG/GCCACCAATGCGGATACGAGAG</td>
<td>CH12F3-rSμ-3'BE, first round bio-PCR</td>
</tr>
<tr>
<td>5'-RED-μμ_ΔSμ</td>
<td>GAAGACATTTGGGAAGGACTGACT</td>
<td>CH12F3-rSμ-3'BE, second round PCR</td>
</tr>
<tr>
<td>AP2</td>
<td>ACTATAAGCAGCGGTGTT</td>
<td>Second round PCR, adaptor side</td>
</tr>
<tr>
<td>ssA-upper-T_5N</td>
<td>GCCGACTATAGGCCACACGCTGTTNNNNN-3AmMO/</td>
<td>Common adaptor oligo</td>
</tr>
<tr>
<td>ssA-lower</td>
<td>/5Phos/CCACGCGTGGCCCTATAGTCGC/-3AmMO/</td>
<td>Adaptor oligo, to ligate with A-tailed first round PCR generated DNA</td>
</tr>
<tr>
<td>ssA-lower-A</td>
<td>/5Phos/ACACGCGTGGCCCTATAGTCGC/-3AmMO/</td>
<td>Adaptor oligo, to ligate with G,C,T-tailed first round PCR product</td>
</tr>
<tr>
<td>5'-Sμ_1-Cas9</td>
<td>CCCATCCACCTGGCTGCTCA</td>
<td>CRISPR-guide RNA for Sμ inversion in CH12F3 cells, 5'</td>
</tr>
<tr>
<td>5'-Sμ_2-Cas9</td>
<td>GAGTCTTCTAGTCAGTATTG</td>
<td>CRISPR-guide RNA for Sμ deletion / inversion in CH12F3 cells, 5'</td>
</tr>
<tr>
<td>3'-Sμ-Cas9</td>
<td>GTAGAACAAGGAAATAGCA</td>
<td>CRISPR-guide RNA for Sμ deletion / inversion in CH12F3 cells, 3'</td>
</tr>
<tr>
<td>JH4_intron_Cas9</td>
<td>GGAGCCCGCTGAGAGAGTT</td>
<td>CRISPR-guide RNA for non-productive Sμ-Sα deletion in CH12F3 cells</td>
</tr>
<tr>
<td>3'-Cα_PAS_Cas9</td>
<td>GGAACCTAGGACTCTGAGT</td>
<td>CRISPR-guide RNA for non-productive Sμ-Sα deletion in CH12F3 cells</td>
</tr>
</tbody>
</table>
M.1.13 Coordinates used for plotting

<table>
<thead>
<tr>
<th>CH12F3 ΔSα1xl libraries</th>
<th>Start</th>
<th>End</th>
<th>Core S region</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgH (200kb)</td>
<td>114484000 - 114684000</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Sμ (10kb)</td>
<td>114654000 - 114664000</td>
<td>114657000 - 11461000</td>
<td></td>
</tr>
<tr>
<td>Isce1-1x break-site (10kb)</td>
<td>114494000 - 114504000</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>3’BE &amp; 5’BE symmetrical jxns</td>
<td>114498726 - 114498960</td>
<td>n/a</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CH12F3 Sμ(INV) ΔSα1xl libraries</th>
<th>Start</th>
<th>End</th>
<th>Core S region</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgH (200kb)</td>
<td>114484000 - 114684000</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Sμ (10kb)</td>
<td>114653813 - 114663813</td>
<td>114657000 - 11461000</td>
<td></td>
</tr>
<tr>
<td>Isce1-1x break-site (10kb)</td>
<td>114494000 - 114504000</td>
<td>n/a</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>rSμ-CH12F3 libraries</th>
<th>Start</th>
<th>End</th>
<th>Core S region</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgH (200kb)</td>
<td>114480000 - 114680000</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>rSμ (10kb)</td>
<td>114656000 - 114666000</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Sα (10kb)</td>
<td>114495500 - 114505500</td>
<td>114498531 - 114502837</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ΔSγ1^2xl B cell libraries</th>
<th>Start</th>
<th>End</th>
<th>Core S region</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgH (200kb)</td>
<td>114500000 - 114700000</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Sμ (10kb)</td>
<td>114662000 - 114672000</td>
<td>114663618 - 114669313</td>
<td></td>
</tr>
<tr>
<td>Sε (10kb)</td>
<td>114509000 - 114519000</td>
<td>114511933 - 114516114</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>rSμ B cells libraries</th>
<th>Start</th>
<th>End</th>
<th>Core S region</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgH (200kb)</td>
<td>114500000 - 114700000</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>rSμ (20kb)</td>
<td>114654000 - 114674000</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Sε (20kb)</td>
<td>114504000 - 114524000</td>
<td>114511933 - 114516114</td>
<td></td>
</tr>
<tr>
<td>Sγ1 (20kb)</td>
<td>114563000 - 114583000</td>
<td>114565429 - 114579090</td>
<td></td>
</tr>
<tr>
<td>Sγ3 (20kb)</td>
<td>114594000 - 114614000</td>
<td>114601770 - 114606687</td>
<td></td>
</tr>
<tr>
<td>Sγ2b (20kb)</td>
<td>114537000 - 114557000</td>
<td>114544100 - 114549283</td>
<td></td>
</tr>
<tr>
<td>Sγ2a (20kb)</td>
<td>114521000 - 114541000</td>
<td>114527344 - 114533964</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5’Sμ B cells libraries</th>
<th>Start</th>
<th>End</th>
<th>Core S region</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgH (200kb)</td>
<td>114500000 - 114700000</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Sμ (20kb)</td>
<td>114657000 - 114677000</td>
<td>114663618 - 114669313</td>
<td></td>
</tr>
<tr>
<td>Sε (20kb)</td>
<td>114504000 - 114524000</td>
<td>114511933 - 114516114</td>
<td></td>
</tr>
<tr>
<td>Sγ1(20kb)</td>
<td>114563000 - 114583000</td>
<td>114565429 – 114579090</td>
<td></td>
</tr>
</tbody>
</table>

149
M.1.14 Sequencing data

HTGTS sequencing data has been deposited in the GEO database under the accession number GSE71005.


10. Gostissa M, et al., IgH class switching exploits a general property of two DNA breaks to be joined in cis over long chromosomal distances. Proc Natl Acad Sci U S A. 2014 Feb 18;111(7):2644-9


M.2 Chapter 2

M.2.1 Materials and Methods

Mice, cells and HTGTS experiments have been described previously [1].

M.2.2 Cell lines

Generation of ΔSα1xI-Sce1-CH12F3 and NCA-CH12F3 cell lines is described previously [1]. To obtain ATM−/− cells, wild type CH12F3 cells were first nucleofected using the 4D-nucleofector X (Lonza, solution P3, protocol CA-137) with the two pX330 vectors with gRNA sequence GTCCTCAGTCGATTATCACT and TATCTTGATAAACGAGCAGT targeting the ATM gene. Single cell subclones were seeded into 96-well plates 12 hours post nucleofection and the resulting clones were screened by PCR and Southern blot. Two independent clones used in this study were confirmed by western blot with α-ATM antibody (SantaCruz Biotechnology, (2C1) sc23921). To obtain 53BP1−/− cells, NCA-CH12F3 cells were first nucleofected using the 4D-nucleofector X (Lonza, solution P3, protocol CA-137) with the two pX330 vectors with gRNA sequence targeting the 53BP1 gene. Single cell subclones were seeded into 96-well plates 12 hours post nucleofection and the resulting clones were screened by PCR. Two independent clones used in this study were confirmed by western blot with α-53BP1 antibody (Cell Signalling Technology, #4908).

For generating HTGTS libraries cloning from c-myc-Cas9 mediated bait DSB, two million CH12F3 cells were stimulated with anti-CD40, IL4 and TGF-β for 15hrs followed by nucleofection with pX330 vector with gRNA (GACGAGCGTCACTGATAGTA) targeting c-myc using 4D-nucleofector X (Lonza, solution P3, protocol CA-137), cells were replated in
stimulation conditioned media. On day 3 post-stimulation, cells were harvested and genomic DNA was isolated for HTGTS library preparation. For each clone of a given genotype, three independent experiments with cytokine stimulations / nucleofection were performed and cells were harvested for extracting genomic DNA to perform HTGTS.

**M.2.3 Data analyses**

HTGTS junctions from B cells in various genetic backgrounds were described previously and raw data have been submitted to the NCBI GEO. Microhomology is defined as the longest region at the switch junction of perfect uninterrupted donor/acceptor identity. Sequences with gap between donor and acceptor sequences are considered as insertions and were excluded from calculations. Direct joins were defined as junctions with no MH and no insertions (MH=0). To plot the MH pattern, direct joins and junctions with MH were pooled and sorted by length, the number of junctions with indicated length of MH were counted and calculated as percentage to the total number of junctions mapped to the region of interest. The coordinates of each S region for the calculation are described previously⁴.

**M.2.4 Statistical analysis**

Unpaired two-tailed $t$ test was used to examine the statistical significance between samples. A p value $<$0.05 was considered as significant.

**M.2.5 Sequencing Data**

HTGTS sequencing data analysed here has been deposited in the GEO database under the accession number GSE71005.
M.2.6 References