V(D)J Recombination, Somatic Hypermutation and Class Switch Recombination of Immunoglobulins: Mechanism and Regulation

A short title: Genetic mechanism of Ig gene expression

Xiying Chi¹,²#, Yue Li¹,²#, Xiaoyan Qiu¹,²*

1. Department of Immunology, School of Basic Medical Sciences, Peking University, Beijing, 100191, China

2. NHC Key Laboratory of Medical Immunology, Peking University, Beijing, 100191, China

#These authors contributed equally to this work.

*Corresponding Author, Xiaoyan Qiu, Department of Immunology, School of

This article has been accepted for publication and undergone full peer review but has not been
trough the copyediting, typesetting, pagination and proofreading process, which may lead to
differences between this version and the Version of Record. Please cite this article as doi:
10.1111/imm.13176

This article is protected by copyright. All rights reserved
Basic Medical Sciences, Peking University, Beijing, 100191, China

Email: qiuxy@bjmu.edu.cn

**Key words**: B cell, V(D)J recombination, somatic hypermutation, class switch recombination, mechanism

**Abbreviations:**

Ig, Immunoglobulins; DSB, double strand-breaks; CSR, class switching recombination; SHM, somatic hypermutation; IgH, immunoglobulin heavy; IgL, immunoglobulin light; RSS, recombination signal sequence; NBD, nonamer binding domain; HMG, high-mobility group; NHEJ, nonhomologous end joining; C-NHEJ, canonic NHEJ; Alt-NHEJ, alternative NHEJ; DNA-PKcs, DNA-dependent protein kinase; Lig, ligase; pol, polymerases; XRCC4, X-ray cross complementing Group 4; XLF, XRCC4-like factor; DNA-PK, DNA protein kinase; PAXX, parologue of XRCC4 and XLF; PARP1, Poly-ADP ribose polymerase 1; MRN, Mre11-Rad50-Nbs1; ssDNA, single-stranded DNA; RPA, replication protein A; TdT, terminal deoxynucleotidyl transferase; ATM, Ataxia telangiectasia mutated; AMPK, AMP-activated protein kinase; PHD, plant homeodomain; RBP, RNA-binding proteins; FOXO, Forkhead box; Ebf1, early B cell factor 1; Erag, enhancer region of RAG1/2; CTCF, CCCTC-Binding Factor; 3'RR, 3' regulatory region; CBS, CTCF binding sites; HDAC3, histone deacetylase 3; RIC, RSS Information Content; YY1, Yin Yang 1; subTADs, sub topologically associated domains; AID, activation-induced cytidine deaminase; MMR, mismatch repair; BER, base excision repair; Exo1, Exonuclease 1; PCNA, proliferating cell nuclear antigen; UNG, Uracil-DNA glycosylase; APE1/2, apurinic/apyrimidinic endonuclease; PCNA-Ub, PCNA ubiquitination; GLT, germline transcripts; PKA, Protein Kinase A; DDR, DNA damage response; MRN, Mre11-Rad50-Nbs1; MDC1, Mediator of Damage Checkpoint protein; 53BP1, p53-binding protein 1; BRCA1, breast cancer 1; RAP80, receptor-associated protein 80; LSR, Locus suicide recombination; PP2A, protein phosphatase 2; GANP, germinal center-associated nuclear protein; eEF1A, eukaryotic
elongation factor 1 α; NES, nuclear export signal; Bach2, BTB and CNC homology 2; Blimp-1, B lymphocyte-induced maturation protein-1; hs, hypersensitive sites; 3′Eα, 3′ Ca enhancer; G4, G-quadruplexes; DDX1, DEAD-box RNA helicase 1; HMT, histone methyltransferase; MMSET, multiple myeloma SET domain; Mbd4, methyl-CpG binding domain protein 4; MLH1, MutL homologue 1; TSLP, thymic stromal lymphopoietin; BAFF, B cell-activating factor; APRIL, a proliferation-inducing ligand; HKSC, heat-killed Saccharomyces cerevisiae.

**Abstract**

Immunoglobulins (Ig) emerged from B lymphocytes, capable of recognizing almost all kinds of antigens owing to the extreme diversity of their antigen binding portions known as variable (V) regions, plays an important role in immune response. The exons encoding the V regions are known as V (variable), D (diversity), J (joining) genes. V, D, J segments exist as multiple copy arrays on the chromosome. The recombination of V(D)J gene are the key mechanism to produce antibody diversity. The recombinational process including randomly choosing a pair of V, D, J segments, introducing double strand-breaks (DSBs) adjacent to each segment, deleting (or inverting in some cases) the intervening DNA and ligating the segments together is defined as V(D)J recombination, which contributes to surprising Ig diversity in vertebrate immune system. To enhances both the ability of immunoglobulins to recognize and bind to foreign antigens and the effector capacities of the expressed antibodies, naive B cells will undergo class switching recombination (CSR) and somatic hypermutation (SHM). However, the genetics mechanisms of V(D)J recombination, CSR and SHM are not entirely clearly. In this review, we summarize the major progression about mechanism studies of immunoglobulin V(D)J gene recombination and CSR as well as SHM and their regulatory mechanisms.

1.**Immunoglobulin structure and genetic encoding**

Ig is comprised of two immunoglobulin heavy (IgH) chains encoded by the IgH heavy chain locus (chromosome 14 in human and 12 in the mouse) and two immunoglobulin light (IgL) chains.
encoded by either the Igκ (chromosome 2 in human and 6 in the mouse) or Igλ (chromosome 22 in human and 16 in the mouse) light chain loci. IgH chains have five major isotypes (Igμ, Igα, Igγ, Igδ, and Igε), with four subtypes of IgG (Igγ1, Igγ2, Igγ3 and Igγ4) and two subtypes of IgA (Igα1 and Igα2). IgL chain have two types (Igκ and Igλ). IgH chains have four (Igγ, Igα and Igδ) or five (Igμ and Igε) domains while IgL chains have two domains. The N-terminal regions of heavy and light chains, known as variable (V) region, are highly variable in their sequences and are the antigen-binding portions of the antibody (1). While the C-terminal regions of Ig, both IgH and IgL have only a few sequential variations in the same species but different individuals and thus are called constant (C) regions (2). Igγ, Igα and Igδ all have three CH while both Igμ and Igε have four CH (see Figure S1. a) (3).

The V region of IgH chain is encoded by V(variable), D(diversity), J(joining) genes while the one of IgL chain is encoded by V, J segments with the absence of D segments in the light chain loci. The C regions of different Ig isotypes are encoded by different CH exon clusters, which are organized in the order of Cμ, Cδ, Cγ, Cε and Cα in the IgH locus. (4) genes. These genes are assembled in the developing lymphocyte by V(D)J recombination to form a complete Ig (see Figure S1. b)

2. V(D)J recombination

The V regions are assembled through V(D)J recombination of V_H, D_H, and J_H genes on the heavy chain and V_L and J_L genes on the light chain. This recombination is initiated by double strand-breaks (DSBs) produced by the RAG1-RAG2 recombinase at specific recombination signal sequence (RSS).

2.1. RSS

The RSSs (see Figure S2. a) normally consists of a highly conserved heptamer motif (consensus sequence 5'-CACAGTG-3') and a conserved nonamer sequence (consensus sequence 5'-ACAAAAACC-3') separated by a poorly conserved spacer sequence of 12 or 23 nucleotides. (3, 5) The heptamer sequence is considered to be the essential recognition element. The first three nucleotides of the heptamer (closest to the coding flank) show the highest sequence conservation, and
are critical for recombination, whereas the remaining heptamer positions are much less important. The nonamer sequence is dispensable for recombination, with only a few highly conserved positions.

Both the RSS at the 3’ end of the IgH V fragment and the 5’ end of the J fragment carry a 23 bp spacer sequence, while the D fragment has a 12 bp spacer sequence at both the 5’ and 3’ ends of the RSS (see Figure S2. b).(3) According to the 12/23 rule(6), efficient recombination only occurs between RSS with different spacer lengths. Therefore, D can interact with J and V while the V cannot be connected to the J because it does not comply with the 12-23 rule. What’s more, proximal sites undergo recombination substantially more frequently than distal ones.(6) So the order is firstly D-J then V-DJ. For the light chain, the 3’ end of the V fragment and the 5’ end of the J fragment are reverse complementarily complementary, with a 12 or 23 bp spacer sequence, respectively, which can result in a V-J linkage.

However, recently, it has been widely proved that the 12/23 rule of genomic recombination is frequently violated under physiological conditions, resulting in unanticipated hybrid recombinations. Non-12/23 junctions under physiological conditions are thought to be rare, the most common being in VDDJ rearrangements of the IgH locus that occur once per 800 cells. In addition to violating the 12/23 rule, other noncanonical rearrangements form hybrid signal-to-coding junctions are typically generated in artificial systems, but may also be detected at low frequency in vivo. (7)

2.2. RAG1/RAG2

The V(D)J recombinase consists of two lymphoid-specific proteins, RAG1 and RAG2, of 1040 and 527 residues, which carry out DNA cleavage. Both RAG proteins are large multi-domain proteins that consist of core regions and non-core regions. (see Figure S3) The catalytic core and DNA-binding functions are largely contained within RAG-1.(6) The “core” region of RAG1 is required for binding to the nonamer as well as catalysis of cleavage. The non-core domains of the RAG proteins provide important regulatory functions, but the mechanism is not well understood. However, the role for RAG2 is less clear, which may function to activate RAG1 for sequence-specific binding and cleavage, and also provide additional DNA binding capability.(8, 9)
2.3. The mechanism of V(D)J recombination

Efficient cleavage of a DNA substrate requires only RAG1, RAG2, a divalent metal ion, and HMGB1 or HMGB2. The first step in V(D)J recombination is binding of RAG, probably together with HMGB1 or HMGB2. RAG–RSS interactions have been studied. The RAG complex catalyzes two reactions, nicking and hairpin formation, without dissociation. First, RAG complex binds either a 12-RSS or a 23-RSS, resulting in a 12 or 23 signal complex. Then a nicking is introduced precisely at the junction between the coding segment and the RSS. Another type of RSS is captured to form a paired complex, and nicking is followed as well.(10) In vitro, high-mobility group (HMG) proteins such as HMGB1 or HMGB2 have been shown to stimulate RAG’s activity in DNA binding, nicking, and hairpin formation, presumably by inducing RSS bending.(9, 11) Upon PC formation, the free 3′-hydroxyl released from the nicking step attacks the opposing strand by transesterification in the presence of Mg2+ to create a hairpin coding segment and the cleaved RSS ends. Proteins in the nonhomologous end joining (NHEJ) DNA repair pathway are recruited to the coding ends. The coding ends are opened by NHEJ DNA repair factors and then joined, forming an imprecise coding joints that contain added nucleotides.(9, 12, 13) (see Figure 1. a) After V(D)J recombination of IgH and IgL chain genes, the resulting immature naïve B-cells transcribe the IgH and IgL genes and the transcripts undergo alternative splicing to produce IgD and IgM by fusing the μ and δ exon to the J exon.(1)

NHEJ is the major DSB rejoining process and occurs in all cell cycle stages, including two competing non-homologous end joining pathways, canonic NHEJ (C-NHEJ) and alternative NHEJ (Alt-NHEJ) (See Figure 2).

The mechanism of C-NHEJ has been well explained. The main proteins required for C-NHEJ involve the Ku70/Ku80 heterodimer, the catalytic subunit of the DNA-dependent protein kinase (DNA-PKcs), the endonuclease Artemis, the DNA ligase IV (Lig IV), the DNA polymerases μ (pol μ) and λ (pol λ), X-ray cross complementing Group 4 (XRCC4) and XRCC4-like factor (XLF).(14) Ku70/80 is thought to be the first protein to bind to the DSB, based on its abundance and tight affinity.(15) Ku70/80 protects DNA ends from digestion, and the Ku:DNA complex also serves as a
node at which the nuclease, polymerases and ligase of C-NHEJ can attach. There is a Ku:DNA complex at each of the two DNA ends being joined, permitting each DNA end to be modified in preparation for joining. Ku:DNA complex recruits DNA-PKcs and Artemis, a DNA-PKcs-activated endonuclease, generating a DNA protein kinase (DNA-PK). DNA-PKcs undergoes autophosphorylation and activates Artemis, which then gains various nuclease activities including 5’ endonuclease activity, 3’ endonuclease activity, and hairpin opening activity.(15, 16) The nuclease activities ensuring the two ends are compatible by resecting damaged or non-ligatable end groups at the end of the DSB. The gap created after processing is filled by Pol μ/λ recruited by Ku.(17) DNA-PK also facilitates recruitment of a ligation complex, which encompasses LigIV, XRCC4 and XLF. PAXX (parologue of XRCC4 and XLF) is a recently discovered protein with structural similarity to XRCC4 and XLF and promote DSB repair by interacting with Ku.(18, 19)

The role for Alt-NHEJ in V(D)J recombination was originally thought to be a backup pathway for DSB repair.(20) However, emerging evidence has demonstrated that Alt-NHEJ can also occur in C-NHEJ-intact cells.(20, 21) The higher affinity of Ku for DSB compared with Poly-ADP ribose polymerase 1 (PARP1) could explain the predominance of C-NHEJ. The mechanism of Alt-NHEJ involves 5’-3’ resection of the DNA ends, annealing of microhomology, fill-in synthesis and ligation. PARP1 can recognize the broken DNA ends and serves as a scaffold for the recruitment of other DNA damage factors.(22) Firstly, Mre11-Rad50-Nbs1 (MRN) and CtIP complex initiates 5’-3’ DNA resection, creating short single-stranded DNA (ssDNA) overhangs.(20) Resection exposes microhomology internal to break sites, which could facilitate spontaneous annealing of ssDNA. The binding of Replication protein A (RPA) to the resulting short ssDNA overhangs removes secondary structure and prevents annealing of overhangs, which hinders Alt-NHEJ.(23, 24) Polθ–helicase acts as an ATP-dependent annealing helicase that dissociates RPA to promote DNA annealing and stimulate Alt-NHEJ.(25) The paired ssDNA overhangs are subsequently extended by Polθ. Finally, end-rejoining is carried out by the DNA ligase I or III (Lig I or Lig III)/XRCC1 complex in coordination.(22, 26, 27)

The recombination shown above proceeds in a way that intervening DNA is deleted. In fact, the
intervening DNA can also be inverted without deletion. Whether recombination proceeds in a
deletional or inversional manner depends on the relative orientation of the two RSS. V and J segments
located on the same strand are recombined by deletional way, leaving the coding segment on the
chromosome and the signal joint on an excised circle of DNA. If V and J segments are on opposite
strands, they are joined by inversion between the RSSs, generating a signal joint and a coding joint.

Notably, to generate an efficient immune repertoire, V(D)J recombination favors genetic diversity
at two levels. The first level corresponds to a diverse rearrangement of V, D and J genes. The second
level is characterized by a particular joining mechanism of coding segments characterized by the loss
or addition of extra nucleotides. During V(D)J recombination, the terminal deoxynucleotidyl
transferase (TdT) adds random nucleotides (N-segments) at the V-D and D-J junctions in heavy chains
of Ig, thereby significantly contributing to the diversity of the immune repertoire(28).

2.4. Regulatory mechanism of V(D)J recombination

V(D)J recombination undergoes regulation in multiple level, including the regulation of RAG1/2
activity, subnuclear reposition, spatial conformation of chromatin, and Vk recombination has some
specific regulation mechanisms.

2.4.1. The regulation of RAG1/RAG2 activity

Since RAGs are important enzymes in V(D)J recombination, their activity and quantity have an
important influence on this process. The RAG proteins can coordinate their activity through
autoregulatory properties. The RAG1 N-terminus is also associated with ubiquitylation-dependent
regulatory processes by acting as an E3 ligase.(29) RAG2 is a regulatory cofactor, which promotes
RAG1 to bind and cleave DNA. However, with the RAG1 C-terminus (residues 1009–1040)
collaboration, the C-terminus of RAG2 has autoinhibition efficient on RAG activity.(30)

Phosphorylation events are related to the activity of RAGs as well. RAG2 has a conserved
Serine-Glutamine (SQ) phosphorylation site on 365 to 366. This Ataxia telangiectasia mutated

This article is protected by copyright. All rights reserved
(ATM)-mediated phosphorylation site of RAG2 has function of feedback control of cleavage and maintenance of genome stability.\(^{(31)}\) Thr-490 is an essential CDK phosphorylation site in a conserved degradation signal sequence, which links the amount of RAG2 to the cell cycle.\(^{(32)}\) RAGs activity can be also enhanced by the AMP-activated protein kinase (AMPK), which directly phosphorylates RAG1 at serine-528. This phosphorylation event can increase the catalytic activity of the RAGs thus initiates V(D)J recombination.\(^{(33)}\)

Besides, the sensitivity of the gene to RAG cleavage is related to the epigenetics of histone (see Figure S4).

The accessibility hypothesis postulates that the initial step of V(D)J recombination must be chromatin changes, which render the different genes of the Ig loci accessible to the V(D)J recombination machinery. The 3’ proximal region of Ig loci is activated by enhancer-mediated recruitment of histone-modifying enzymes. Consequently, J segments that are adjacent to the Eμ enhancer are characterized by an abundance of active histones like H3K4me2 (trimethylation of histone H3 at lysine 4), H3K4me3, and H3K9ac in pro-B cells. Through a plant homeodomain (PHD) finger, the RAG2 proteins are able to specifically bind the active histone mark H3K4me3 in the RSSs that are associated with J segments.\(^{(34)}\) H3K4me3 subsequently changes the conformation of catalytic and substrate-binding regions of RAG1 and the autoinhibitory domain (residues 352 to 405) in RAG2, increasing substrate affinity and catalytic rate.\(^{(35, 36)}\) Additionally, SUV39H1, a histone lysine N-methyltransferase involved in the formation of heterochromatin, is involved in the establishment of H3K9me3 and H4K20me3, the hallmarks of heterochromatin. The formation of heterochromatin is known to influence RAG2 catalyzing the VDJ recombination. Moreover, SUV39H1 has a role in the methylation of RAG2, which changes RAG2 subnuclear localization, might regulate the chromatin binding of RAG2.\(^{(37)}\)

Apart from those ways to regulate RAGs activity, there are also some mechanisms to regulate the amount of RAGs. Since RAG2 protein is only expressed in the G0-G1 phase, proteins that control the cell cycle can also influence the amount of RAGs. The RNA-binding proteins (RBPs) ZFP36L1 and ZFP36L2 can suppress mRNAs, helping B cells entry and exit the cell cycle. The cells which are lack...
of ZFP36L1 and ZFP36L2 show a delayed V(D)J recombination. (38)

Recent studies have shown that Forkhead box O (FOXO) protein has an important regulatory effect on the activity of RAG proteins. (See Figure S5) Pre-BCR synergize with interleukin-7 receptor (IL-7R) signaling pathways is found to negatively regulate RAG mRNA and protein levels in clonal expansion period of pre-BCR+ cells, but positively regulate the expression of RAGs in differentiation to small pre-B cell period. (39) FOXO protein is the major RAG transcription activators. It can activate the expression of RAG1 and RAG2, as well as upregulate the amount of p27, a stabilizer for RAG2. (40) PI(3)K and Akt kinase, activated by tonic (foreign antigen independent) receptor signaling, can phosphorylate and result in cytoplasmic separation of FOXO protein. (41) Moreover, early B cell factor 1 (Ebf1), a transcription factor, can either directly regulate the accessibility of specific genes at the loci of IgH and Igλ, or indirectly contributes to V(D)J recombination via the activation of Pax5. (42) The expression of Ebf1 is driven by the downstream effector Stat5 of IL-7R. During pre-B cell proliferation, Ebf1 and its downstream target c-Myb have negative influence on RAG transcription, through the negative regulation of FOXO1 binding to the RAG locus. Ebf1 can directly downgrade FOXO1 expression and upgrade zinc finger protein Gfi1b expression as well. (39, 43) Similarly, activation of DNA damage is able to depress RAG expression by ATM-mediated canonical NF-κB transcription factors as well as the loss of ATM-dependent FOXO1 binding to the enhancer region of RAG1/2 locus (Erag). (43, 44)

2.4.2. Subnuclear reposition

Evidence from several model systems have indicated that the location of genes especially their proximity to heterochromatin may be important for initiating or maintaining silence of gene, and modulating locus accessibility. There are two repressive compartments: the nuclear periphery and pericentromeric heterochromatin that are important for propagating the inactive state of genes. It has been revealed that subnuclear location will change during the development of B cell. (45, 46) In progenitor, the IgH locus is anchored at the nuclear periphery via the distal VH genes that is in an accessibility repression stage; (47) meanwhile, the 3’ proximal D and J clusters are free, which
facilitates D–J rearrangements in lymphoid progenitors. Both two IgH alleles will undergo D-J recombination, which leads to Pro-B cell. IgH loci are subsequently relocated from the nuclear periphery to more central positions within the nucleus. This subnuclear repositioning reactives V genes in preparation for V-DJ recombination. The two alleles behave differently at the stage of pre-B cells. Following successful V–DJ rearrangement of one IgH allele which gives rise to pre-BCR, the signaling leads to repositioning of the second, incompletely rearranged IgH allele to repressive pericentromeric heterochromatin,(48) whereas the functionally rearranged IgH allele remains in a central position. This feedback inhibition of V–DJ recombination at the second allele ensures that each B lymphocyte expresses only one monospecific BCR, known as allelic exclusion (see Figure S6).(49) Mono-allelic expression seems to take place preferentially on the associated allele.(45) However, it has also been reported that activated B cells can transcribe both IgH alleles as well.(50) To what extent the two IgH alleles differ remains unclear.

2.4.3. Spatial conformation of chromatin

Chromatin interactions can control the gene assembly of IgH chains. In the Pro-B cell, in which D-J recombination has finished while V-DJ recombination is going to take place, contraction will happen in the distal V cluster in a way of looping.(48) The localized activity of the RAG1/2 at the proximal chromatin is strictly dependent on locus contraction for subsequent V(D)J recombination. The long-range contraction of the IgH locus also facilitates V–DJ rearrangements due to the proximity of both elements. In addition, different V genes can recombine with similar frequency, which is essential for the generation of a highly diverse Ig repertoire. While little is known about the molecular mechanisms controlling locus contraction. Only few trans-acting factors like Pax-5(51), YY1(52), CCCTC-Binding Factor (CTCF)(53, 54) have been implicated in the regulation of contraction.(49)

Chromatin interactions control the gene assembly of IgH chains. In developing B lymphocytes, V (D) J recombination is also regulated by remote cis-acting elements. At distant sites of IgH variable region, the 3’ regulatory region (3’RR) conveys a transcriptional silencing activity on transcription of Ig gene, but the repressive activity switched off after V-DJ recombination, promoting transcription of
Ig in mature B cells.\(^{(55)}\)

Besides, the highly conserved 11-zinc finger protein CCCTC-Binding Factor (CTCF) can mediate long-range chromatin interactions that influence organization and function of mammalian genome. CTCF also affects the complex configuration of chromatin loops, that control the functional interactions between elements. CTCF binding sites (CBS) orientations can defining chromatin architecture which supports V (D) J recombination.\(^{(54)}\)

In addition, histone deacetylase 3 (HDAC3), a catalytic component of the NCoR/SMRT co-repressor complexes, is involved in global changes in chromatin structure, which likely have effects on distal V-DJ recombination.\(^{(56)}\)

2.4.4. V\(\kappa\) recombination regulation

The mechanisms of V\(\kappa\) recombination regulation is different from V\(\text{H}\) in some ways. Firstly, the RSS for each V gene is different in quality, which can be quantified as RSS Information Content (RIC) scores. Theoretically, a higher score means a more conducive tendency to recombination. In addition, Ig\(\kappa\) recombination in pre-B cell is activated by the binding of some transcription factors like PU.1, E2A, IRF4 and IRF8 at the iEx and 3’Ex enhancers. PU.1 binding at the RSS is a key feature to determine whether the V\(\kappa\) gene will actively recombine,\(^{(57)}\) and the antisense lncRNA PU.1 is able to inhibit PU.1 mRNA translation by block tRNA recruitment.\(^{(58)}\) In early B-cell progenitors, two DNA methyltransferases, Dnmt3a and Dnmt3b are participated in the methylation of Tcfe2a gene that encodes E2A. The lack of Dnmt3a and Dnmt3b elevate the level of E2A. Moreover, H3K4 methylation and IKAROS binding at the RSS are important to promote higher recombination frequencies.\(^{(59, 60)}\)

Besides, long-range genomic interactions, which distantly located genomic regions on same chromosome undergo, are dynamically controlled at Ig\(\kappa\) light-chain locus. Yin Yang 1 (YY1), a widely expressed TF, contributes to long-range chromatin interactions in IgH chain. Also, CTCF can regulate the expression of developmentally regulated gene. However, their roles in Ig\(\kappa\) locus...
contraction have not been determined. Recent research shows a B cell specific enhancer, E88, is a major hub of long-range chromatin interactions which connects sub topologically associated domains (subTADs) in V gene region with recombination center at the J genes in Igκ. This enhancer ensures a proper folding structure of Igκ locus for V(D)J recombination.

The primary Ig repertoire is initially composed of IgM after the V(D)J recombination. The range of reactivity of the primary Ig repertoire can be further modified by somatic hypermutation (SHM) and class switch recombination (CSR) at the immunoglobulin loci. SHM results in mutations being introduced into the V region of heavy-chain and light-chain, altering the affinity of the Ig for its antigen. While in CSR, the initial heavy-chain C regions are replaced by another isotype, modifying the effector activity of the Ig but not its specificity. Both CSR and SHM are initiated by an enzyme called activation-induced cytidine deaminase (AID), which is expressed specifically in activated B cells. So these two processes do not occur in T-cell receptor genes.

3. SHM and CSR

B cells after going through the V(D)J recombination subsequently undergo two genetic modifications, somatic hypermutation (SHM) and class switch recombination (CSR). The purpose of these alterations, mostly in the germinal center, is to increase the affinity and alter the biological properties of Ig but with a specificity for the antigen. For some species like chickens, Ig gene conversion is used in addition to or instead of SHM(64). The molecular mechanism of SHM, CSR as well as gene conversion, shows many similarities and shares several enzymes. But Activation-induced cytidine deaminase (AID) is thought to be the only factor that is absolutely necessary for all these mechanisms. Next we will discuss the mechanism of SMH and CSR in detail.

AID is a 198-amino acid protein. It is one of 12 members of the APOBEC family of DNA/RNA cytidine deaminases, which shares a conserved catalytic domain with other members. The catalytic domain contains His56, Cys87, and Cys90, which bind Zn$^{2+}$ and are essential for catalytic activity, and Glu58, which serves as a general acid-base catalyst. The C-terminal-domain is essential for AID to
mediate CSR.\(^{(65)}\) AID is a ssDNA-specific cytidine deaminase, which converts cytosine bases (C) to uracils (U) \(\text{in vitro}\) and \(\text{in vivo}\) but with no activity on double-strand DNA, RNA, or RNA:DNA hybrids. The deamination reaction proceeds via a direct nucleophilic attack at position 4 of the pyrimidine ring of cytosine by Zn\(^{2+}\)-coordinated to AID. However, AID is an inefficient enzyme, deaminating only about 3% of the cytidines even at preferred hotspot motifs. How this inefficiency aids in CSR and SHM remains unclear at present.

U caused by AID can trigger several types of DNA repair including the mismatch repair (MMR) and the base excision repair (BER) pathways. In the MMR, U mismatched with G is detected by the heterodimer MutS\(\alpha\), comprised of MSH2 and MSH6\(^{(66)}\). MutS\(\alpha\) recruits MutL\(\alpha\) (MLH1-PMS2) and Exonuclease 1 (Exo1). The endonuclease activity of PMS2 then cleaves the DNA 5’ of the mismatch, creating a DNA nick that serves as point of entry for the Exo1, which initiates resection from the nick going past the mismatch site and creates an extended patch of ssDNA. Proliferating cell nuclear antigen (PCNA) subsequently recruits Pol\(\delta\) to replicate over the gap and Ligase 1 (Lig 1) finalizes repair. While in BER, U is recognized by Uracil-DNA glycosylase (UNG). UNG removes U from DNA\(^{(67)}\), leaving an abasic site that is recognized by apurinic/apyrimidinic endonuclease 1/2 (APE1/2). APE1/2 nick the DNA 5’ of the abasic site. PARP1 is activated and scaffold XRCC1, Pol\(\beta\) and a ligase is recruited. Pol\(\beta\) then remove the 5’ deoxyribose and insert a single nucleotide, followed by a ligation and so complete the repair\(^{(68)}\). Interestingly, both UNG and MutS\(\alpha\) are conserved enzymes that initiate DNA repair with high fidelity through either MMR or BER. Thus, noncanonical MMR or BER pathway is needed for transition or transversion mutations in SHM and DSBs formation required for CSR.\(^{(65)}\) (see Figure 3)

### 3.1. Mechanism of SHM

For error prone MMR in SHM, APE2 rather than MutL\(\alpha\) is recruited to nick the mismatch and provide an entry for the Exo1. Furthermore, low-fidelity polymerases like Pol\(\eta\) instead of Pol\(\delta\) are recruited through PCNA ubiquitination (PCNA-Ub)\(^{(69)}\) to resynthesize the DNA gap and introduce mutation, which is mainly at A:T residues. While in noncanonical BER during SHM, some abasic sites serve as a noninformative template. REV, recruited by PCNA-Ub as well, inserts dCMP into the new
DNA strand opposite the abasic site. After a further round of DNA replication this can result in a stable transversion mutation at the site of the original C:G base pair.

3.2. Mechanism of CSR

CSR is a DNA deletional-recombination reaction that proceeds through the generation of DSBs in switch (S) region preceding each C\(\text{H}\) gene and is completed by NHEJ between donor and acceptor S regions, resulting in the replacement of the expressed C\(\text{H}\) exon cluster.

The S regions are repetitive DNA elements that locate upstream of all the C\(\text{H}\) genes except C\(\delta\). Mouse S regions are composed of 1–10 kb of tandem repeats that are G-rich on the non-template strand. Both the sequence and number of tandem repeats vary among mouse S regions. S\(\mu\), for example, is approximately 3.2 kb long and is comprised of GAGCT pentameric motifs, with the AGCT palindromic sequence representing a canonical RGYW/WRCY sequence. S\(\gamma1\) is approximately 10 kb long and has multiple RGYW/WRCY sequences embedded within 49 bp repeat units. Other repeat motifs found in S regions include GGGGA/T in S\(\mu\), S\(\gamma1\), S\(\gamma2b\), and S\(\gamma3\) and GGGCT in S\(\varepsilon\) and S\(\alpha\). S regions are the targets for AID that initiates CSR.(65, 70)

There are a transcriptional intervening (I) promoter and an I exon upstream of all C genes except C\(\delta\). At the onset of CSR, I promoters are selectively activated and produces non-coding germline transcripts (GLTs), which initiates at the I exon, proceeds through the S region and terminates downstream of the corresponding C\(\text{H}\) gene.(71) Since AID is a ssDNA-specific cytidine deaminase, GLTs is thought to separate the two double-strand DNA and provide access to AID.(72) AID binds to the S region by 14-3-3 adaptors(73) and recruit Protein Kinase A (PKA) as well. AID is phosphorylated at Ser38 of the N-terminal region by PKA, generating a binding site for RPA.(74) RPA enhances the deamination activity of AID and transfers C into U. If two U are sufficiently near in opposite strands, a DSB is resulted through BER pathway. While conversion of two more distal U into a blunt DSB that is necessary for CSR requires MMR.(75)

AID-initiated DSBs in two S regions activate the ataxia telangiectasia-mutated (ATM)-dependent
DNA damage response (DDR). The Mre11-Rad50-Nbs1 (MRN) complex is firstly recruited to a DSB, where it recruits the protein kinase ATM. ATM phosphorylates and recruited numerous DDR proteins including H2AX, Mediator of Damage Checkpoint protein (MDC1)(76) ubiquitin ligases RNF8 and RNF168, p53-binding protein 1 (53BP1), breast cancer 1 (BRCA1) and receptor-associated protein 80 (RAP80)(77). H2AX and 53BP1 mediate synapsis of upstream and downstream DSBs,(78) and 53BP1 can also prevent the rejoining of intra S-region DSB.(79) DSBs in different S regions are eventually rejoined by C-NHEJ or Alt-NHEJ. Both C-NHEJ and Alt-NHEJ pathway are used in CSR.(80) DSBs generated by the MMR pathway seem to favor Alt-NHEJ possibly by exposing microhomologies after Exo processing.(75)

IgD CSR is restricted to a few B-cell subsets in specific lymphoid tissues such as mesenteric lymph nodes, peritoneal cavity and mucosa-associated tissue in both mice and humans.(81-83) A recent study suggested that IgD CSR is initiated by microbiota,(83) implying a role of IgD in the regulation of the microbial homeostatic.

3.3. Locus suicide recombination

Ig production rely on the selection of Ag-specific B cell. That implies not only proliferation and differentiation of those B cells specially binding Ag but also elimination of the less efficient or inappropriately activated cells.(84) It’s important for B cells to undergo such a choice since inappropriate production of useless might trigger autoimmunity and inflammation. Locus suicide recombination (LSR) is reported as a variant form of recombination mechanism to eliminate inappropriate B cells. LSR is characterized by recombination between Sμ and the 3’ regulatory region (3’RR) downstream of the C region, deleting the whole IgH constant gene cluster and thereby terminating B cell function. Apart from direct junctions from Sμ to 3’RR, sequence analysis of LSR both in vivo and in vitro reveal the presence of junctions of Sγ or Sα regions rather than Sμ with 3’RR, indicating that LSR has taken place in previous CSR in at least 14% of cases.(85) Murine 3’ RR includes repetitive DNA sequences resembling S-regions. In humans, there are two 3’RRs due to the interval duplication of Cα gene on the IgH locus.
(3’RR1 and 3’RR2 respectively downstream of Cα1 and Cα2).(86) Analyzing the structure of human
3’RR shows their high similarity with the mouse 3’RR and both of them share multiple features with S
regions. These “Like switch” (LS) regions promote LSR through a process that is similar to CSR:
noncoding germline transcription of Sμ and LS within the 3’RR, DSB caused by AID, and finally DNA
repair, mainly by the NHEJ pathway as in CSR.(87) However, it’s recently reported the major
involvement of the A-NHEJ rather than C-NHEJ during DSB repair during LSR compared with CSR,(88)
suggesting a mechanism to control the balance the CSR and LSR. LSR is mostly detected during B-cell
activation but remain undetectable in long-lived memory cells or plasma cells.(85) Interestingly, Since
LSR is a process towards cell death, it remains difficult to evaluate its regulation.(85)

3.4. Regulation of AID

Given that AID plays a significant role in both CSR and SHM, it is a crucial mechanism to regulate
CSR and SHM by modulating AID expression, stability and localization.

3.4.1. Epigenetic regulation of AID

Histone modification and miRNA are two significant epigenetic mechanism that regulates gene
expression. In B-cell activation, stimulatory signals induce several histone-modifying enzymes to active
H3K4me3, H3K9ac and H3K14ac in the promoter regions of AID and remove the repressive H3K27me3
and H3K9me3, changing the chromatin structure or recruiting other modifying factors thus upregulating
AID. These signal also activate miRNA like mir-16, mir-155 and mir-181b that decrease the expression
of AID by binding to and degrading complementary sequences of the mRNA.(89)

Phosphorylation also modulates AID activity at a post-translation level. 5-15% of the AIDs in
activated B cells are phosphorylated at serine 38 by PKA, and it has a role in targeting AID to DNA and
increasing its activity.(90) Moreover, threonine 140 can be also phosphorylate but PKC family members,
enhancing AID relatively. Threonine 140 phosphorylation preferentially affects SHM, suggesting that
post-translational modifications may contribute to the choice between CSR and SHM.(91) Serine 3 is
another phosphorylation point controlled by protein phosphatase 2 (PP2A). Increasing S3
phosphorylation leads to a reduced activity of AID.(92)

3.4.2. The localization and stability of AID

It is contradictory to figure out that AID is predominantly located in cytoplasm while both CSR and SHM happen in nucleus. Further studies have revealed that AID is transported between nucleus and cytoplasm continuously, of which the mechanism is not fully clear though. Germinal center-associated nuclear protein (GANP), an RNA-binding protein, is assumed to transport AID into the nucleus in mice.(93) Cytoplasmic AID interacts with eukaryotic elongation factor 1 α (eEF1A), a translation elongation factor involved in the AID cytoplasmic retention.(94) Export of AID is mediated by the Ran-dependent nuclear exportin CRM1 since AID has a powerful nuclear export signal (NES) at the C terminus.(95)

AID localization is also associated with its stability. In the nucleus, AID undergoes a ubiquitin-dependent protein degradation and as well as a Reg-γ-mediated ubiquitin-independent degradation.(96) While in the cytoplasm, molecular chaperones like Hsp90 and DnaJα1 can inhibit the polyubiquitination and proteasomal degradation of AID thus increase the stability of AID.(97) The complex regulatory mechanism of AID subcellular localization and stability allows proper amounts of AID to access into the genome, protects the genome from continuous mutations, and minimizes the detrimental off-target effects.

3.4.3. Bach2

BTB and CNC homology 2 (Bach2) is a B-cell-specific transcription repressor and is thought to be the common pathway in controlling the expression, protein stability and subcellular location of AID.(98) Bach2 upregulates AID mainly by suppressing B lymphocyte-induced maturation protein-1 (Blimp-1), a factor that drives plasma differentiation by suppressing mature B-cell-associated genes including AID.(99) Bcl6 can co-bind to the cis-regulatory sequence of the Blimp-1 gene with Bach2, upregulating AID and maintaining the stability of the Bach2.(100) However, Bcl6 as well as other suppressors of Blimp-1 like Pax5 and Spib, can be in turn blocked by Blimp-1 at a transcriptional level, forming a
negative feedback regulatory pathway. Moreover, Bach2 can also upgrade the expression of AID through a Blimp-1-independent pathway. Apart from decreased amount of AID in B cells, Bach2 deficiency also profoundly decreased AID level in nucleus. Bach2-deficient cells also had reduced expression of molecular chaperones known to stabilize cytoplasmic AID.

3.5. Other Regulatory mechanism of CSR

3.5.1. 3’ RR

The 3’RR is located at the most distal 3’ region of the Ig H chain locus, is comprised of four hypersensitive sites (hs) hs3a, hs1,2, hs3b and hs4. Each hs is a weak enhancer but they form a strong enhancer synergistically. The 3’RR is crucial for conventional CSR, but its mechanism remains unclear. Conclusions drawn from knock-out mice are that the 3’RR plays a role in promoting germline transcription and synopsis between S regions in CSR. In addition, it has also been reported that 3’RR mainly controls the S acceptor region rather than the S µ donor region in multiple aspects including alterations of epigenetic marks, germline transcription, R loop formation, paused RNA Pol II, AID targeting, and generation of DSBs. And once DSBs are generated, the S µ - S γ 1 junctions are not affected by the 3’ RR. The complete 3’RR deletion dramatically affects CSR and Ig secretion for all isotypes. But the IgG1 class displays a special status since only Cγ1 transcription and CSR are partly preserved after alteration or even complete deletion of the 3’RR, while all other class-switched Ig was nearly abolished, suggesting that other elements may support the γ1 transcription. Furthermore, it seems that IgD CSR and IgA CSR in B1 cells are totally independent on the 3’RR. And Sµ-σδ junction is mainly produced by A-NHEJ rather than C-NHEJ, which markedly differs from IgG, IgA and IgE CSR.

3.5.2. Chromosome looping

CSR’s activation likely needs the signal from BCR, and the secondary signals from CD40 and TLR. Next, the transcription of unrearranged C_H genes regulate naïve B cells to specific isotype. The type of
cytokines can regulate the direction of CSR by chromosome looping, as different transcription promoters, located at upstream of each acceptor S region, are activated by different cytokines. Then specific S-C loci (like Sε-Cε loci) is closed to the Eμ-Sμ-Cμ and 3′ Ca enhancer (3′Eα) region. Finally, the formation of chromosome looping can induce the cell switching to that specific isotype (Figure 4). (111)

3.5.3. G-quadruplex

G-quadruplexes (G4) are secondary structures which are formed in nucleic acids by guanine-rich sequences. They are constructed around continuous G-tetrads of Hoogsteen hydrogen bond guanine bases.(112) G4 DNA structures have been previously reported to be present on the non-template DNA strand of Sμ and Sγ regions in vitro. Recent studies show that AID deaminates ssDNA more robustly in the context of structured substrates, such as G4. Some studies have shown that G4 formation plays a role in CSR. G4 mediates recruitment and oligomerization of AID, which helps create mutation clusters and DSB on IgH chain.(73) Conversely, G-quadruplex(G4)-stabilizing agents, like RHPS4, can inhibit CSR and decrease immunoglobulin secretion. (113) Moreover, two NME [NM23/NDPK (nucleoside diphosphate kinase)] isoforms are novel players in CSR process. Before CSR activating, NME1 binds to the S region and suppresses CSR. Upon stimulation, NME1 dissociates from activated S region, while NME2 binds to G4 in the S region to promote CSR. NME1 and NME2 act as coordinated inhibitory stimulus pairs to modulate CSR.(114)

3.5.4. R-loop

R-loop is a stable intermediate of CSR, which is composed of a DNA:RNA hybrid and the associated non-template ssDNA. R-loop forms behind the extended RNA polymerase II (Pol II) in the S-region and results in ssDNA that may act as a substrate for AID, the ssDNA-specific cytidine deaminase. Recent years, many studies have showed that compared to the original model, R-loop is unlikely to play a role in initial AID-dependent mutagenesis phase, and it must have other AID-independent roles in CSR. For example, R-loop can regulate DNA replication environment at IgH Locus.(115, 116)
There are also many other factors participate in the regulation of CSR through the R-loop pathway. DEAD-box RNA helicase 1 (DDX1) is a DEAD-box RNA helicase required at the IgH locus for CSR in vivo. DDX1 binds to G4 structures, converting them into S-region R-loop. This process targets AID to the S regions, resulting in the promoting of CSR (Figure 4. d). (116) shows how DDX1 promotes R-loop formation.

3.5.5. Other regulatory pathways

Many other molecules have been discovered to regulate CSR in recent years. For example, histone methyltransferase (HMT) and multiple myeloma SET domain (MMSET) can promote AID-mediated DNA breaks during CSR. (117) CD11b, an integrin expressed on the surface of activated B2 cells induces AID’s expression by NF-κB signaling pathway. (118) Estrogen receptors (ER) in activated B cells may control CSR by the direct binding of ER to key regulatory elements in the IgH chain locus, such as Sμ, Eμ, and the 3′RR. (119) BATF promote the production of class switched Ig by positively regulating the expression of Nfil3 and miR155hg, but negatively regulating Wnt10a. (120) The BER enzyme methyl-CpG binding domain protein 4 (Mbd4) participates in MMR-directed DNA end processing by interacting with MutL homologue 1 (MLH1). (121) The nuclear structural protein NuMA interacts with 53BP1, controlling 53BP1 diffusion and negatively regulating 53BP1 in DSB repair. (122) The serine/threonine phosphatase PP4 shows a promoting component of CSR by involved in sequential recruitment of RPA and NBS1. (123)

3.5.6. Ig specific isoform induction

In addition to the mechanisms above, certain conditions can also induce Ig specific isoform producing. Isotype switch recombination is controlled by cytokine receptors and their downstream signals. It is widely believed that IgG1 and IgG3 are mainly produced in response to protein antigens, but chronic stimulation with these antigens leads to an increase in the proportion of IgG4. IL-4 can induces IgE and IgG secretion while IL-10 can induce the production of IgG1 or IgG3. TGF-β is involved in the conversion to IgA, and IFN-γ can induce the production of IgG2a or IgG3. (124, 125)
More conditions have been discovered in recent years that can induce the production of Ig specific isoform. First, there are many newly discovered inducing factors for IgE has a lower activation threshold, so compared to IgG1, IgE has higher transcription probability in the early choice. (126) Thymic stromal lymphopoietin (TSLP) can activate dendritic cells, so the cells express the OX40 ligand (OX40L; CD252), triggering naive CD4+ T cells to differentiate into inflammatory Th2 effector cells. Th2 effector cells secrete the cytokines of IL-4, IL-9 and IL-13, which enhance IgE production. In allergen-specific Th2 memory cells, immunoglobulin class switch from IgM to IgE with allergen stimulation as well as CD40 and CD40L interactions. (127) In addition, microRNA-146a upregulates 14-3-3σ expression, enhancing class switch and secretion of IgE in B cells. (128)

There have also been some recent discoveries about IgG and IgA CSR. Specific inhibitors of elastase cause cells to secrete higher levels of IgG and IgA in mice, and increase transcription of factors involved in murine B cell differentiation, like mRNA for AID, IL-10, B cell-activating factor of the TNF family (BAFF) and a proliferation-inducing ligand (APRIL). (129) Dectin-1 agonists, like heat-killed Saccharomyces cerevisiae (HKSC), can selectively induce IgG1 class switching in LPS-activated B cells through reinforcing surface expression and direct stimulation of Dectin-1. (130)

**Conclusion**

Given the importance of the diverse repertoire of Ig in immune responses, it is reasonable to suppose that sophisticated genetic mechanisms have evolved to generate and regulate these proteins. However, these mechanisms are still not fully elucidated. After more than one-century research, it has been discovered that the generation of Ig undergoes a series of specific genetic events including V(D)J recombination, CSR and SHM. V(D)J recombination contributes to the generation of surprising Ig diversity. CSR and SHM can increase the affinity of Ig to recognize and bind with various antigens. Even though the molecular mechanisms of V(D)J recombination, CSR as well as SHM have been deeply studied, there are still many things remains unclear. For example, how inefficient AID initiates CSR and SHM, how CSR and SHM are well regulated and competed with error-free DNA repair pathways, and how B cells regulate LSR are still not well understood. And identification of regulatory factors and their
function in V(D)J recombination, SHM and CSR, especially those associated with RAGs and AID, do require further work to elucidate.

**Conflicts of Interest**

The authors declare no financial or commercial conflicts of interest.

**Acknowledgements**

This work was supported by research grants from the Key Support Projects of the National Natural Science Foundation’s Major Research Programme (91642206), the Strategic Priority Research Programme of the Chinese Academy of Sciences, China (XDA16010500), Major International Cooperation Projects of the National Natural Science Foundation (81320108020), the National Natural Science Fund for Distinguished Young Scholars (31625016), the Research Institute Fund of the NHC Key Laboratory of Medical Immunology, Peking University (BMU2018JDJS010), the Non-profit Central Research Institute Fund of the Chinese Academy of Medical Sciences (2017PT31037), Shenzhen Science and Technology Program (JCYJ20170413141047772 and JCYJ20180507181659781), CAS Key Research Projects of Frontier Science (QYZDY-SSW-027), the CAS Youth Innovation Promotion Association (2018133 and 2016097), and the Shanghai Municipal Science and Technology Major Project (2017SHZDZX01).

**Reference**


This article is protected by copyright. All rights reserved
short-homology regions is suppressed under terminal deoxynucleotidyltransferase expression, as revealed by a novel vector system for analysing DNA repair by nonhomologous end joining. FEBS Open Bio. 2016;6(1):16-23.


23. Deng SK, Chen H, Symington LS. Replication protein A prevents promiscuous annealing

This article is protected by copyright. All rights reserved


43. Ochodnicka-Mackovicova K, Bahjat M, Maas C, van der Veen A, Bloedjes TA, de Bruin AM, et al. The DNA Damage Response Regulates RAG1/2 Expression in Pre-B Cells through ATM-FOXO1


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved


111. Stavnezer J, Schrader CE. IgH chain class switch recombination: mechanism and regulation. J


121. Grigera F, Wuerffel R, Kenter AL. MBD4 Facilitates Immunoglobulin Class Switch


This article is protected by copyright. All rights reserved
Figure Legends

Fig. 1. Overview of V(D)J recombination. (a) The RAG1-RAG2 complex is shown as light blue and light green trapezoids. Briefly, RAG binds a single RSS in the presence of HMGB1 to form a 12 or 23 signal complex. Another RSS is captured to form a paired complex. The DNA is cleaved, generating hairpin coding ends and RSS ends with a 3′ hydroxyl (OH) group in the presence of Mg^{2+}. The hairpin is released and opened and rejoined by NHEJ, resulting in imprecise coding joints that contain added nucleotides (blue bars). RSS ends is processed through NHEJ pathway as well, creating a signal joint. (b) B cell ontogeny during V(D)J recombination. This figure shows the chronological order of B cells in different stages of development in the bone marrow. B cells progress from stem cells to pro-B cells, pre-B cells and immature pre-B cell stages. During this differentiation, V(D)J rearrangement of heavy chains occurs in pro-B cell periods, resulting in the expression of pre-B cell receptors (pre- BCR), which are composed of Igμ and surrogate light chains (composed of VpreB and λ-5). After receiving the pre-BCR signaling, the light chains rearrange in small pre-B cell, resulting in the expression of a mature BCR (composed by rearranged heavy and light chains).

Fig. 2. A model for C-NHEJ and Alt-NHEJ. C-NHEJ (left) and Alt-NHEJ (right) are shown. For C-NHEJ, Ku70/80 binds to the DSBs firstly and protects DNA ends from digestion. The Ku:DNA complex recruits DNA-PKcs in complex with Artemis, generating DNA-PK. DNA-PKcs undergoes autophosphorylation and activates Artemis, which then gains various nuclease activities, ensuring the DSBs are compatible by resecting damaged DNA or non-ligatable end groups. The gap created is filled by Pol μ/λ recreated by Ku. DNA-PK also facilitates recruitment DNA ligase IV, XRCC4 and XLF to complete the ligation of DSBs. For Alt-NHEJ, PARP1 recognizes the broken DNA ends and recruits MRN and CtIP complex to initiate 5′-3′ DNA resection, creating ssDNA overhangs. Resection exposes microhomology internal to break sites, facilitating spontaneous annealing of ssDNA. The binding of RPA to the ssDNA overhangs removes secondary structure and prevents annealing of overhangs, which hinders Alt-NHEJ. Polθ–helicase acts as an ATP-dependent annealing helicase that dissociates RPA to promote DNA annealing and stimulate Alt-NHEJ. The paired ssDNA overhangs are subsequently extended by Polθ–polymerase. Finally, endrejoining is carried out by the DNA ligase I or III (Lig I or Lig III)/XRCC1 complex.

Fig. 3. overview of error-free BER and MMR, SMH and CSR. (a) Error-free repair pathways for base excision repair (BER, left) and mismatch repair (MMR, right). Canonical BER is initiated by UNG, creating an abasic site that is recognized by APE1/2. APE1/2 nick the DNA 5′ of the abasic site. PARP1 and XRCC1 are activated and recruit PCNA and Polβ. Polβ remove the remaining 5′ deoxyribose and insert a single nucleotide, followed by ligation with Lig3. In MMR, MutSα recognizes the U:G mispair and recruits MutLα. MutLα nick the DNA 5′ of the mismatch via PMS2. Exo1 creates an ssDNA from the nick going past the mismatch site. PCNA subsequently recruits Polα to fill over the gap, following ligation by Lig1. (b) During SHM, uracil can act as a template for replication leading to a C–T transition mutation. Alternatively, noncanonical BER or MMR recruit low-fidelity polymerases Polη through PCNA ubiquitination (PCNA-Ub) leads to transition or transversion mutations. (c) For CSR, noncanonical BER can lead to DSBs when two uracils in opposite strands are closely spaced. MMR can process distantly spaced uracils, leading to staggered DSBs. Blunt DSBs are joined by C-NHEJ, whereas staggered breaks are repaired.
Fig. 4. Chromosome looping and R loop. (a) This picture shows the structure of the chromosome looping, and each colorful circle representing the specific S-C loci. Eμ and 3′Eα interact, causing Sμ and the downstream S regions located in a chromosomal loop. (b) After the Igε transcription promoter (located at the upstream of Igε acceptor S region) activated, Sε-Cε loci are close to 3′Eα segment and Eμ-Sμ-Cμ, inducing the B cell to switch to IgE. (c) Structures which recruit AID: The switch sequence has a high G-richness on the non-template chain. During transcription, these G-rich regions are formed secondary structures such as the R-loop, G4 and G-loop, which help recruit AIDs in S-region. (d) DDX1 promotes R-loop formation: RNA polymerase II (Pol II) produces switch transcripts, whose intron contains the G4. After splicing step, the intron lariat intermediate is debranched by RNA lariat (intron) debranching (DBR1) enzyme. Bound G4-AID complex, DDX1 targets to the S region DNA. At last G4 is resolved, and R-loop is formed in the S-region.

Fig. S1. Ig gene locus and IgG structure. (a) The basic structure of Ig (IgG as an example) is shown. IgG consists of two IgH chains and two IgL chains. the disulfide bonds that link the IgH chains are shown. The rectangles represent Ig domains that constitute the IgH chains (blue) and IgL chains (yellow) including V regions (shaded in light blue), which are encoded by VH, DH, JH in IgH or VL, JL in IgL, and C regions, which are encoded by CH or CL. Complementarity determining regions (CDRs) are indicated as regions in red boxes. (b) Organization of Ig genes in human is shown. Since Ig chains are extracellular proteins, the V gene is preceded by an exon encoding a leader peptide (L), which directs the protein into the cell’s secretory pathways. The chains are encoded by V (blue), D (brown), J (yellow), and C (orange) genes. The dotted line represents omitted genes.

Fig. S2. RSS sequence and location. (a) The consensus RSS sequence is shown. Both signal sequences consist of a conserved heptamer and conserved AT-rich nonamer; they are separated by nonconserved spacers of 12 or 23 bp. The most highly conserved positions of the heptamer and nonamer are shaded in gray. Two different RSSs, with either 12 or 23 spacer, is represented by red or orange triangle. (b) The two types of RSS have characteristic locations within Ig DNA. During DNA rearrangement, genes adjacent to the 12-bp RSS can join only with segments adjacent to the 23-bp RSS.

Fig. S3. RAG1/2 structure and function. (a) Domains of RAG1/2 are demonstrated. The core region of RAG1 is responsible for the most binding and catalysis activity. While the non-core N or C-terminal domains of the RAG proteins work to regulate proteins function. RAG1 N-terminus is associated with ubiquitylation-dependent regulation. RAG1 C-terminus collaborates with RAG2 C-terminus and leads to an autoinhibition. Phosphorylation sites on RAG proteins also regulate RAG protein activity. S528 of RAG1 increases the catalysis activity. S365 phosphorylation site of RAG2 can control the cleavage of DNA and maintain the genome stability. T490 links the amount of RAG2. It also shows nonamer binding domain (NBD) of RAG1. (b) This schematic shows the spatial structure of the RAG complex. RAG1 (blue and green) and RAG2 (yellow) are respectively represented as cartoons. Among them, the two NBDs of RAG1 are flexibly attached.

Fig. S4. Histone Modification and Contraction of V cluster. The Eμ enhancer recruit enzymes that contribute to histone modification and thus activates the proximal J gene region. The
RAG1/2 complex binds to the active histone H3K4me3 in the RSS and leads to D–J recombination (shaded in red) in progenitors. Then the complex capture one of the many V genes upon IgH locus contraction in pro-B cells for V(D)J recombination. The dotted line represents omitted V segments.

Fig. S5. Molecules that regulate the amount of RAGs by FOXO. This figure shows the mechanism of how the FOXO protein and its related factors affect RAG activity. As the major RAG transcriptional activator, FOXO can activate the expression of RAG, and upregulate the amount of stabilizer p27 of RAG2. PI(3)K and Akt kinase are activated by upstream tonic receptor signaling, phosphorylating and leading to cytoplasmic separation of FOXO protein. Moreover, during pre-B cell proliferation, Ebf1, a transcription factor, and its downstream target c-Myb have a negative effect on RAG transcription. Expression of Ebf1 is driven by the downstream effector Stat5 of IL-7R. Ebf1 and c-Myb negatively regulate FOXO1 binding to the RAG locus, as well as directly downgrade FOXO1 expression and upgrade Gfi1b expression. Gfi1b represses RAG transcription through biding near the Erag and Irag2 elements, and recruiting chromatin modifiers which deposit transcriptional repression chromatin mark H3K9me2. Similarly, activation of DNA damage can inhibit RAG expression by ATM-mediated classical NF-κB transcription factors, as well as the loss of binding of ATM-dependent FOXO1 to the Erag.

Fig. S6. Subnuclear location of the IgH locus. The subnuclear location of the IgH locus is shown at different stages of early B cell development. The distal V region (red) and proximal D-J region (green) of the IgH locus as well as the repressive compartments of the nuclear periphery (gray) and pericentromeric heterochromatin (blue) are demonstrated.
KU70/86 binding → DNA-PKcs → Artemis

End processing → polymerases μ/λ

Gap filling → DNA ligase IV/XRCC4/XLF

Ligation → C-NHEJ

End processing → RPA binding → Polymerase 0

Annealing and extending → XRCC4

Ligation → Alt-NHEJ
(a) [Diagram of gene structure with various color-coded elements]

(b) [Diagram of gene structure with various color-coded elements]

(c) A R-loop

B G4

C G4 loop

(d) [Diagram showing RNA G4 resolution & R-loop formation with DBX1 and AID]