HHS Public Access

Author manuscript

Mol Cell. Author manuscript; available in PMC 2019 April 05.

Published in final edited form as:

Mol Cell. 2018 April 05; 70(1): 21–33.e6. doi:10.1016/j.molcel.2018.02.020.

Sequential enhancer sequestration dysregulates recombination center formation at the *IgH* locus

Xiang Qiu[1], **Gita Kumari**[1], **Tatiana Gerasimova**[1], **Hansen Du**[1], **Lawal Labaran**[1], **Amit Singh**[1], **Supriyo De**[2], **William H. Wood III**[2], **Kevin G. Becker**[2], **Weiqiang Zhou**[3], **Hongkai Ji**[3], and **Ranjan Sen**[1,*,4]

[1]Laboratory of Molecular Biology and Immunology, National Institute on Aging, Baltimore, MD 21224, USA

[2]Laboratory of Genetics and Genomics, National Institute on Aging, Baltimore, MD 21224, USA

[3]Department of Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, 21205, USA

SUMMARY

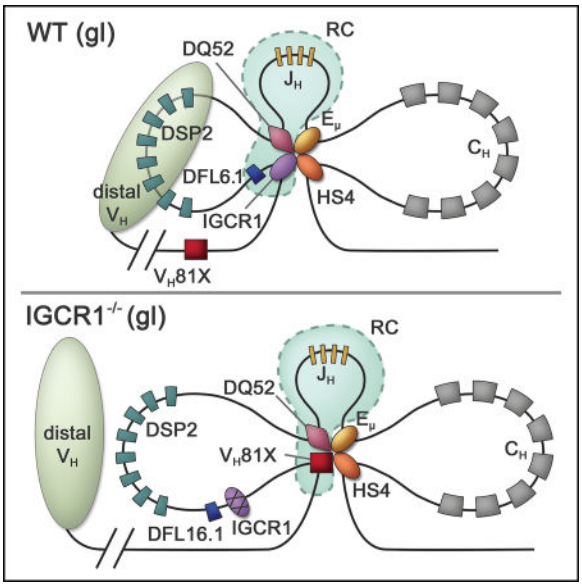
Immunoglobulin heavy chain (*IgH*) genes are assembled by DNA rearrangements that juxtapose a variable (V_H), a diversity (D_H), and a joining (J_H) gene segment. Here we report that in the absence of intergenic control region 1 (IGCR1) the intronic enhancer (Eμ) associates with the next available CTCF binding site located close to V_H81X, via putative heterotypic interactions involving YY1 and CTCF. The alternate Eμ/V_H81X loop leads to formation of a distorted recombination center, altered D_H rearrangements, and disrupts chromosome conformation that favors distal V_H recombination. Cumulatively, these features drive highly skewed, Eμ-dependent recombination of V_H81X. Sequential deletion of CTCF binding sites on IGCR1-deleted allele suggests that they influence recombination of single proximal V_H gene segments. Our observations demonstrate that Eμ interacts differently with IGCR1- or V_H- associated CTCF binding sites and, thereby, identify distinct roles for insulator-like elements in directing enhancer activity.

Graphical abstract

*Correspondence: Senranja@grc.nia.nih.gov.
[4]Lead Contact

AUTHOR CONTRIBUTIONS X.Q. and R.S. conceived and designed the research. X.Q., G.K., T.G., H.D., L.L. and A.S. performed experiments. X.Q., S.D., W.H.W., K.G.B., W.Z. and H.J analyzed the data. X.Q. and R.S. interpreted the data and results, and wrote the manuscript.

DECLARATION OF INTERESTS
The authors declare no competing interests.



INTRODUCTION

Immunoglobulin heavy chain (*IgH*) genes are assembled by DNA rearrangements that juxtapose a variable (V_H), a diversity (D_H), and a joining (J_H) gene segment into an exon that encodes the antigen-binding domain of antibody heavy chains. The order of rearrangements is fixed. D_H to J_H rearrangements occur first to create DJ_H recombined alleles, followed by V_H rearrangements to DJ_H junctions to produce VDJ_H recombined alleles (Jung et al., 2006; Kumari and Sen, 2015). Approximately 194 V_H gene segments, 8–12 D_H gene segments, and 4 J_H gene segments are distributed over 2.8 Mb of the mouse genome (Figure 1A). These gene segments can be assorted randomly during recombination to generate the large number of different antigen-binding specificities required for optimal immunity. To facilitate generation of diversity, the *IgH* locus undergoes developmentally-regulated epigenetic changes and alteration of 3D chromatin conformation (Bossen et al., 2012). Based on live cell imaging studies, a current view is that the variable domain consists of several rosettes that form a ‘cloud’ close to the DJ_H part of the locus to permit comparable utilization of V_H gene segments during recombination (Lucas et al., 2014). We previously proposed that the *IgH* locus conformation is generated by a hierarchy of structural changes that involve different transcription factors (Gerasimova et al., 2015).

Three transcription factors have been linked to *IgH* locus structure. Pax5 and YY1 are required for locus compaction that juxtaposes 5’ and 3’ parts of the locus (Ebert et al., 2011; Fuxa et al., 2004; Liu et al., 2007; Verma-Gaur et al., 2012). The third transcription factor, CTCF, binds to about 80 sites that are mostly in the V_H part of the locus (Degner et al., 2011; Lin et al., 2012). Additionally, a cluster of CTCF binding sites mark the 3’ end of the *IgH* locus (Volpi et al., 2012), and 2 prominent CTCF-binding elements (CBEs) are located within the intergenic control region 1 (IGCR1, Figure 1A). The IGCR1 CBEs have a critical role in regulating VDJ_H recombination as described below. CTCF has also been shown to regulate V(D)J recombination at other antigen receptor loci. Both *TCRβ* and *TCRδ* loci have

CTCF binding sites positioned between V and J gene segments, analogous to the location of IGCR1 in the *IgH* locus (Chen et al., 2015; Majumder et al., 2015; Zhao et al., 2016). Deletion or mutation of these sites changes the frequency of V gene segment utilization at these loci (Chen et al., 2015; Majumder et al., 2015; Zhao et al., 2016). Conversely, introduction of new CTCF binding sites into *TCRβ* locus reduces rearrangements to varying degrees depending on their location (Chen et al., 2016; Rawat et al., 2017).

The tissue-specific endonucleases RAG1 and RAG2 initiate *IgH* gene assembly by introducing DNA breaks at recombination signal sequences (RSSs) associated with each rearranging gene segment (Schatz and Ji, 2011). RAG1/2 load onto antigen receptor loci at spatially restricted regions known as recombination centers (RCs) (Ji et al., 2010). The *IgH* RC forms over a 3.5 kb region between the 3'-most D_H gene segment (DQ52) and Eμ, and includes all four J_H gene segments. Alt and colleagues have recently proposed that RAG1/2 track from the RC to locate complementary RSSs for synapsis and DNA cleavage (Hu et al., 2015). At the *IgH* locus, RAG1/2 tracking is limited to a 60 kb chromatin loop between Eμ and IGCR1 (Hu et al., 2015). Spatial proximity of 5'-most (DFL16.1) and 3'-most (DQ52) gene segments to the RC in this configuration may explain, in part, the high frequency of DFL16.1 and DQ52 rearrangements (Choi et al., 2013; Lin et al., 2016). Finally, absence of V_H gene segments from this configuration provides a plausible explanation for D_H to J_H recombination preceding V_H recombination.

Eμ and IGCR1 cis-DNA elements (Figure 1A) have been implicated in regulating recombination. Both steps of *IgH* gene assembly are reduced on Eμ-deleted alleles, with V_H rearrangements being affected especially strongly (Afshar et al., 2006; Perlot et al., 2005). This is likely mediated by Eμ's effects on epigenetic features of unrearranged and DJ_H rearranged *IgH* alleles (Chakraborty et al., 2009; Subrahmanyam et al., 2012). Deletion or mutation of CBEs within IGCR1, dramatically dysregulates V_H recombination. This is reflected in premature recombination of the 3'-most V_H gene segments to unrearranged DQ52 in a breakdown of the normal sequence of *IgH* rearrangements, sharply reduced use of distal V_HJ558 gene segments in VDJ_H junctions, and suppressed feedback inhibition of proximal V_H recombination (Guo et al., 2011b). Mutation of either CBE partially recapitulates recombination phenotypes of doubly-mutated alleles (Lin et al., 2015).

Our studies identify mechanisms by which CTCF-lacking Eμ and CTCF-binding IGCR1 work together to direct optimal recombination of diverse V_H gene segments. Whereas Eμ and IGCR1 are spatially clustered on WT *IgH* alleles, Eμ loops to an alternate CTCF binding site on IGCR1-mutated alleles. In this configuration V_H81X lies close to the RC whereas DFL16.1 moves away from the RC. The nature of the RC itself changes substantially on IGCR1-mutated alleles as evidenced by considerable RAG1/2 recruitment to the vicinity of V_H81X. We suggest that a dysregulated RC, incorporation of V_H81X into a RAG1/2 tracking domain, and spatial proximity of V_H81X to the J_H-associated RC combine to induce premature recombination on IGCR1-deleted alleles. Our observations lead to model whereby sequestration of Eμ by IGCR1 interferes with its ability to form highly specific loops with V_H region CTCF binding sites, thereby restricting RAG protein to a focused RC and generating a chromatin configuration that attenuates VS recombination while enhancing V_H diversity.

RESULTS

Altered looping on IGCR1-mutated *IgH* alleles

Eμ interacts with IGCR1 on wild-type (WT) *IgH* alleles (Guo et al., 2011a; Guo et al., 2011b). We hypothesized that Eμ may loop to the next available 5' CTCF binding site on IGCR1-mutated *IgH* alleles that lack both CBEs (Figure S1). We tested this hypothesis using chromosome conformation capture (3C) in RAG2-deficient pro-B cells that carry *IgH* alleles lacking both IGCR1 CBEs (Guo et al., 2011b). RAG2-deficiency was essential to restrict the locus to its unrearranged configuration, thereby permitting analysis of structurally homogeneous alleles. Eμ-IGCR1 interactions were evident on WT *IgH* alleles but not on IGCR1-mutated *IgH* alleles (Figure 1B). Instead, Eμ interacted with sequences close to V_H81X on IGCR1-mutated alleles (Figure 1B, compare red and blue bars to green bars). Eμ interaction with HS4, a DNase1-hypersensitive site located at the 3' end of the *IgH* locus, was not affected by the IGCR1-mutation (Figure 1B, labeled HS4). As a negative control we used EOMA cells of endothelial origin derived from the 129 mouse strain (Figure 1B, yellow bars). Conversely, a 3C anchor near V_H81X revealed strong interactions with Eμ as well as HS4 on IGCR1-mutated, but not WT *IgH* alleles in pro-B cell lines. Lower levels of V_H81X to HS4 interactions in non-B cells could reflect cell lineage non-specific chromatin compaction; however, we have not pursued this further in the present study. We conclude that the Eμ/HS4 domain interacts with the next available looping site close to V_H81X in the absence of CBEs within IGCR1, resulting in increased V_H81X transcription and localized histone modification (Guo et al., 2011b) (Figure S2 and 3A).

To further substantiate these results at a single cell level, we carried out high resolution 3D-fluorescent *in situ* hybridization (FISH) using short probes to accurately locate specific regions. We found that spatial proximity of IGCR1 to Eμ on WT *IgH* alleles was disrupted in IGCR1-mutated alleles (Figure 1C, Figure S3B). Conversely, probes close to Eμ and V_H81X were spatially distant on WT alleles but closely juxtaposed on IGCR1-mutated alleles. This new Eμ interaction did not extend to the 5' end of the V_H7183 gene family (see below). The combined 3C and FISH studies indicated that Eμ loops to an alternate site close to the 3'-most functional V_H gene segment, V_H81X, upon disruption of its interaction with IGCR1.

Altered looping requires Eμ, YY1 and CTCF

To test the interpretation that increased V_H81X gene transcription on IGCR1-mutated alleles resulted from association with Eμ, we used CRISPR/Cas9 to delete Eμ sequences from a pro-B cell line with IGCR1-mutated *IgH* alleles. We obtained two independent lines in which Eμ was deleted on both alleles in the context of a mutated IGCR1 (Figure S4A). We found that V_H81X transcription induced by IGCR1-mutation was reduced in both Eμ-deleted lines (Figure 2A, left panel). As controls, we assayed transcription from the Eμ-dependent DQ52 promoter. DQ52 transcripts were reduced by Eμ deletion regardless of the status of IGCR1 (Figure 2A, middle panel). *Ctcf* transcription was comparable in all five cell lines (Figure 2A, right panel). As additional controls, transcription of *Yy1* and PAIR elements in the distal V_H region were not affected by manipulation of Eμ or IGCR1 (Figure S4A). These observations corroborate the idea that proximity to Eμ directs V_H81X transcription and histone modifications in the absence of IGCR1.

Eμ-dependent interactions on WT *IgH* alleles require the transcription factor YY1 (Gerasimova et al., 2015). To determine whether the newly-identified Eμ-V_H81X interaction also had the same requirement, we reduced YY1 expression in IGCR1-mutated pro-B cells by introducing a *Yy1*-specific shRNA, and assayed this interaction by 3D-FISH. We achieved approximately 60% reduction of *Yy1* mRNA and protein with the *Yy1*-specific shRNA compared to a scrambled control (Figure S4B). We found that Eμ-V_H81X interactions were reduced in YY1-knockdown (KD) cells (Figure 2B; Figure S4C). Loop disruption was accompanied by reduced transcription of V_H81X (Figure 2C). As controls, Eμ-dependent transcripts that initiated at the DQ52 promoter (μ0 transcript) and within Eμ itself (5'Eμ and 3'Eμ) were not affected by YY1 KD (Figure 2C, S4B).

The proposed Eμ interaction site close to V_H81X contains a prominent CTCF binding site, but no YY1 binding sites (Figure S1) (Verma-Gaur et al., 2012). We hypothesized that this interaction may arise from heterotypic interactions between YY1 bound at Eμ and CTCF bound at V_H81X, such as those have been proposed to occur during X chromosome inactivation (Donohoe et al., 2007). To test this, we knocked-down CTCF expression with three shRNAs (Figure S4B) and assayed the effects on V_H81X germline transcription on IGCR1-mutated alleles. We found that the V_H81X transcription was reduced by all three CTCF-specific shRNAs but not by a scrambled control (Figure 2C). As seen with YY1-KD, DQ52 and Eμ-associated transcripts were largely unaffected in CTCF-KD cells (Figure 2C, S4B). Additionally, FISH analysis showed that CTCF-KD disrupted the Eμ-V_H81X loop (Figure 2B). We conclude that formation of the new Eμ-V_H81X loop on IGCR1-mutated alleles requires both CTCF and YY1.

Functional consequences of altered looping on IGCR1-mutated alleles

The high recombination frequency of DFL16.1 and DQ52 on WT alleles has been proposed to be due to their spatial proximity to the RAG1/2-rich RC via Eμ-IGCR1 interaction (Kumari and Sen, 2015). A key prediction of this model is that alterations in the WT loop configuration should change the frequency of D_H gene segment usage. In the new configuration on IGCR1-mutated alleles, V_H81X replaces DFL16.1 near the RC. Therefore, DFL16.1 recombination frequency should be reduced compared to other D_H gene segments. We tested this hypothesis by quantifying D_H rearrangements in primary bone marrow pro-B cells from recombinase-sufficient IGCR1-deleted mice.

To amplify DFL16.1, DSP2 and DQ52 rearrangements from pro-B cell genomic DNA, we used 5' primers that recognize either DFL16.1, or the 8 DSP2 gene segments (in 129 mouse strain), or DQ52 and a 3' primer close to J_H4 (Figure 3A). We found that the proportion of DFL16.1 compared to DSP2 and DQ52 rearrangements was reduced on IGCR1-deleted alleles (Figure 3A, quantitated in Figure S5A). We also quantitated D_H to J_H rearrangements in cell lines carrying WT and IGCR1-mutated *IgH* alleles. Because these lines were deficient in RAG2, we induced rearrangements by lentiviral transduction of *Rag2*. Though rearrangement levels were variable between lines, the observed trends were similar to those in primary pro-B cells, with the proportion of DFL16.1 to DSP2 and DQ52 recombination being reduced in both IGCR1-mutated lines (Figure S5B).

We previously proposed that D_H recombination brings the associated 5'-recombination signal sequence (RSS) into the RC for V_H to DJ_H recombination (Subrahmanyam et al., 2012). Thus, D_H recombination frequency should also be reflected in D_H utilization in VDJ_H junctions. To test this we amplified V_H7183-DJ_H1 junctions from bone marrow pro-B cells containing WT and IGCR1-deleted *IgH* alleles followed by cloning and sequencing (Figure 3B, top line; Figure S5C). From two samples each of WT and IGCR1-deleted primary pro-B cells, we obtained sequence information from approximately 90 clones from each genotype (Figure 3B; Table S2). Cumulatively, we found that DFL16.1 was utilized in 28 clones and DSP2 in 61 clones from pro-B cells carrying WT *IgH* alleles. (The small size of DQ52 precluded identification of this gene segment in the VDJ_H clones analyzed). We estimated that any single DSP2 gene segment of 8 was used 7–8 times on average amongst our clones. This distribution reflects the prevailing idea that DFL16.1 rearranges more frequently than DSP2 gene segments on WT alleles (Choi et al., 2013; Lin et al., 2016). In contrast, we obtained only 6 DFL16.1-containing VDJ_H junctions and 85 DSP2-containing VDJ_H junctions from IGCR1-mutated pro-B cells, resulting in 5-fold reduced DFL16.1 utilization in the absence of IGCR1 (Figure 3B). As expected (Guo et al., 2011b), several V_H7183 segments recombined on WT alleles, whereas V_H81X was used most frequently on IGCR1-deleted alleles (Table S2). Our interpretation is that movement of DFL16.1 away from the recombination center leads to its reduced utilization in D_H recombination and, thereby, in reduced availability of DFL16.1 J_H1 junctions as substrates for V_H recombination.

V_H81X recombination in the absence of IGCR1

Proximity of V_H81X close to the RC on IGCR1-mutated alleles also predicts that it should be utilized more efficiently in VDJ_H recombination. To test this, we compared kinetics of V_H recombination on WT and IGCR1-mutated *IgH* alleles 4, 5, or 14 days after *Rag2* transduction into IGCR1-mutated cell lines. In two different cell lines we observed high levels of VDJ_H and VDQ52 rearrangements at d4/5 or d14 after transduction (Figure 4A, B; Figure S6A). Amongst V_H7183 and V_HQ52 genes, V_H81X and $V_HQ52.2.4$ were used almost exclusively on IGCR1-mutated alleles, with no detectable rearrangements of the distal V_HJ558 gene segments (Figure S6B). Both forms of V_H recombination on IGCR1-mutated alleles were Eμ-dependent (Figure 4C). In contrast, WT *IgH* alleles underwent only D_H-J_H recombination in the same time frame. As noted previously Eμ deletion reduced but did not eliminate D_H recombination (Figures S5B, S6A and C). Expression of the introduced *Rag2* gene did not differ more than 2-fold amongst the three cell lines (Figure S6D). We conclude that V_H recombination of germline *IgH* alleles is greatly increased when IGCR1 is inactive, likely due to placement of V_H81X close to the RC. This could also explain the greater efficiency of V_H rearrangements to a pre-existing DJ_H junction on IGCR1-mutated *IgH* alleles (Hu et al., 2015).

Altered RC on IGCR1-mutated alleles

To directly investigate the RC of IGCR1-deleted alleles, we carried out chromatin immunoprecipitation (ChIP) with antibodies directed against RAG1 and RAG2. Earlier studies show that the *IgH* locus is distinct from other antigen receptor loci in recruiting RAG1 only when cells co-express RAG2 (Ji et al., 2010). Because the cells we worked with lacked RAG2, we expressed this protein by lentiviral transduction prior to ChIP. Consistent

with earlier observations, RAG1 did not bind to *IgH* locus in RAG2-deficient control cells with WT *IgH* alleles (Figure 5B, yellow bars). In RAG2-expressing cells we found highest levels of RAG1 at DQ52 and J_H gene segments corresponding to the RC on both WT (Figure 5B, green bars) and IGCR1-mutated alleles (Figure 5B, red and blue bars). Amplicons located near Cγ3 and at the *γ-actin* promoter served as negative controls. Lower levels of RAG1 were detected at DFL16.1, and hardly any was present at the DSP2 gene segments in both genotypes. We did not detect RAG1 binding to remaining *IgH* locus amplicons on WT alleles (Figure 5B, green bars; precise locations of V_H amplicons are shown in Figure S7A). In contrast, IGCR1-mutated cell lines had relatively high levels of RAG1 over a 25 kb region that encompassed 5 proximal V_H gene segments extending from V_H7183.4.6 to V_H7183.1.1 (Figure S7A).

RAG2 binding to WT and IGCR1-mutated alleles showed a similar pattern (Figure 5C), with levels at the proximal V_H genes reaching 25–35% of those seen at the RC (represented by DQ52 and J_H2 amplicons). As expected from earlier studies, the *γ-actin* promoter scored positive in RAG2 ChIP, presumably because of H3K4me3 binding by the RAG2 plant homeodomain. We also noted somewhat higher levels of RAG2 binding at unrearranged DFL16.1 and DSP2 regions than previously detected (Ji et al., 2010). At present, we cannot rule out that this is due to some spreading of RAG2 from the RC due to ectopic expression of this protein. However, close coincidence of our observations with all other aspects of published studies of RAG recruitment to WT *IgH* alleles (Ji et al., 2010) strongly supports the idea that substantial levels of RAG1/2 accumulate over proximal V_H gene segments on IGCR1-mutated alleles.

A caveat to this experiment was that expression of RAG2 that was necessary to evaluate the RC also led to ongoing V(D)J recombination. That is, ChIP signals could have arisen from rearranged V_H7183 and V_HQ52 gene segments present in the cell population. To unequivocally rule out involvement of rearranged *IgH* alleles to RAG1/2 ChIP signals, we deleted the IGCR1 in D345 pro-B cells that express a catalytically inactive RAG1 together with endogenous RAG2 (Ji et al., 2010). In these cells *IgH* alleles are uniformly in unrearranged configuration. To determine the distribution of RAG proteins, we carried out ChIP with anti-RAG1 antibodies in two independently derived D345 clones that lack IGCR1 (Figure S7B). We found increased association of RAG1 over approximately 15 kb encompassing the 3'-most V_H7183 and V_HQ52 gene segments in both lines that lacked IGCR1, but not in parental D345 cells (Figure 5D). The levels of RAG1 in the V_H region in the D345 derivatives were approximately 20% of those seen at the J_H-associated RC. We conclude that an altered RC is generated over proximal V_H gene segments prior to initiation of VDJ_H recombination on *IgH* alleles that lack IGCR1.

Reduced distal V_H looping in the absence of IGCR1

Identification of the Eμ-V_H81X loop on IGCR1-mutated *IgH* alleles provides plausible mechanisms for the hyper-recombinogenic state of V_H81X and V_HQ52.2.4. Generation of this loop *per se* does not readily explain the vastly reduced utilization of distal V_HJ558 gene segments as well as gene segments located in the 5' part of the V_H7183 region on IGCR1-mutated alleles (Guo et al., 2011b) (Figure 6A). A possible explanation proposed for skewed

V_H gene segment usage was that premature V_H81X rearrangements used up substrates to which distal V_H gene segments could otherwise recombine (Guo et al., 2011b).

We have previously proposed that Eμ- and YY1-dependent interactions juxtapose distal V_H gene segments to the 3' *IgH* domain (Gerasimova et al., 2015; Guo et al., 2011a). It was possible, therefore, that altered 3D configuration of IGCR1-mutated *IgH* alleles precluded this association. To test this hypothesis we carried out 3D-FISH with probes located close to Eμ, at regions approximately 400 kb away near the 5' end of the V_H7183 region (5'7183), and approximately 1 Mb away at the 3' end of the V_HJ558 region (3'J558) (Figure 6A). We found that Eμ-5'7183 as well as Eμ-3'J558 interactions were reduced on IGCR1-deleted alleles compared to WT *IgH* alleles in RAG2-deficient pro-B cell lines (Figure 6B; Figure S8A, B). We also used a FISH probe located at IGCR1 together with the 3'J558 probe as another measure of the spatial proximity of distal V_H gene segments to the 3' *IgH* domain. We found that these probes were closely associated on WT *IgH* alleles but not on IGCR1-mutated alleles (Figure 6B; Figure S8A, B). As a control we assayed a CTCF-dependent, but Eμ-independent, interaction that compacts the V_H region (Gerasimova et al., 2015). This 300 kb interaction was not affected by the IGCR1 mutation (Figure S8C), consistent with earlier indications that IGCR1does not regulate V_H structure (Medvedovic et al., 2013). We conclude that the pre-folded distal V_H region does not gain spatial proximity to the D_H part of the locus on IGCR1-deleted alleles. This configuration is very similar to that of Eμ-deficient *IgH* alleles, suggesting that both these regulatory sequences work together to direct distal V_H recombination. Similarity between Eμ- and IGCR1-deficient alleles was also evident in shared transcriptional responses in the distal V_H region in these genotypes (Figure S9).

Sequential utilization of CTCF binding sites

Our observations are consistent with the hypothesis that Eμ interacts with the next available CTCF binding site in the locus upon loss of CTCF binding to IGCR1. To directly test the sequential nature of Eμ interaction with CTCF sites, we used CRISPR/Cas9 to delete one or more CTCF sites in the proximal V_H genes. The first 5 CTCF sites in this region are shown in Figure 6C (C1–C5); as previously noted, each of these lies close to a functional V_H gene segment (Lin et al., 2012). We deleted a 2 kb region containing C1 ($C1^{-/-}$), or a 1 kb region containing C2 ($C2^{-/-}$), or a 12 kb region containing both C1 and C2 ($C1^{-/-}C2^{-/-}$) (Figure S10A), and assessed the consequences on VDJ in the context of WT or IGCR1-mutated *IgH* alleles.

Analysis of V_H gene repertoire by deep sequencing (Hu et al., 2016) revealed substantial differences. As previously noted, IGCR1-mutated alleles recombined V_H81X almost exclusively (Figure 6C, labeled $CBE^{-/-}$(1) and $CBE^{-/-}$(2)). However, loss of C1 (together with a part of V_H81X) resulted in predominant use of V_HQ52.2.4 along with much lower levels of V_H7183.1.1 (a pseudogene with functional RSS) and V_H7183.4.6. That is the gene segment closest to the next available CTCF site (C2) dominated the repertoire. Loss of only C2 by itself recapitulated the phenotype of IGCR1-mutated alleles. We infer that absence of C2 is 'invisible' when C1 is intact. However, when both C1 and C2 were lost, albeit in the context of a larger deletion, we observed dominant use of V_H7183.4.6 which is located close

to C3. As noted with C1 deletion, the adjacent V_HQ52.3.8 gene located approximately 12 kb further 5' also recombined at low levels (Figure 6C). Overall recombination efficiency was comparable among all cell lines (Figure S10B).

We note the following caveats in interpreting these observations. First, because we deleted 1–2 kb of genomic DNA in generating C1 and C2 deficiencies, the observed recombination effects could have arisen by changes in gene spacing rather than loss of specific CTCF binding sites. We consider this is unlikely because individual C1 and C2 mutations removed similar amounts of genomic DNA but resulted in very different recombination phenotypes. Second, because each of the C1 and C2 mutations also removed a part of a nearby V_H gene segment, the observed repertoire shift could be because of the loss of the gene segment rather than CTCF binding. However, loss of C1 greatly accentuated only V_HQ52.2.4 recombination but not V_H7183.1.1 recombination, though both gene segments lie comparable distances from C1. The major difference between them is presence of a CTCF peak near V_HQ52.2.4 but not V_H7183.1.1. Our interpretation is that Eμ sequentially and selectively activates the V_H gene segment closest to the first CTCF binding site on IGCR1-mutated alleles.

DISCUSSION

Consequences of altered Eμ looping

Our observations highlight dynamic interactions of Eμ, an enhancer that does not bind CTCF, with CTCF binding elements in the *IgH* locus to regulate VDJ recombination. Before rearrangements start Eμ interacts with CTCF-binding IGCR1 and with regulatory regions at the 3' end of the locus on WT *IgH* alleles (Figure 7). The Eμ/IGCR1 loop constrains RAG1/2 tracking from the J_H-associated RC to a domain that contains D_H and J_H gene segments but no V_H gene segments. Accordingly, D_H gene segments rearrange first. Loss of Eμ-IGCR1 interactions on IGCR1-mutated *IgH* alleles results in Eμ interacting with the next available CTCF site located approximately 90 kb further 5' near V_H81X (Figures 1, 2 and 7 lower line). In this configuration V_H81X displaces DFL16.1 from its RC-proximal position. We propose that reduced frequency of DFL16.1 rearrangements that we observed on IGCR1-mutated alleles (Figure 3) reflects this spatial displacement, even though the relative disposition of DFL16.1 and DSP2 gene segments on linear DNA remains unchanged. To the best of our knowledge this is the first analysis of recombination efficiency in the context of different chromatin loops.

The altered topological domain of IGCR1-mutated alleles permits RAG1/2 to track to proximal V_H gene segments, enabling greatly increased recombination efficiency and mis-timed V_H recombination. Eμ is essential for these aberrant rearrangements, most of which involve V_H81X. The exquisite specificity of RAG activity on IGCR1-mutated alleles is evident from the much lower levels of recombination of the adjacent functional gene segment, V_HQ52.2.4, that is located only 10 kb 5'of V_H81X ((Guo et al., 2011b), Figure 6). However, deleting the V_H81X-associated CTCF site, C1, redirects recombination to V_HQ52.2.4 (Figure 6). While recognizing the caveats associated with deleting rather than mutating CTCF sites as discussed in the results section, our interpretation is that Eμ loops specifically to the first available CTCF binding site in the absence of IGCR1 and directs

recombination almost exclusively to the associated V_H gene segment. This interpretation is further corroborated by dominant use of V_H7183.4.6, that is located closest to the third CTCF binding site (C3), when the first two sites C1 and C2 are deleted within context of IGCR1-deficiency (Figure 6). Conversely, the recombination phenotype of C2-deleted alleles is indistinguishable from a locus with only the IGCR1 mutation since the first available Eμ looping site C1 remains intact. The specificity of Eμ interactions with CTCF sites in the V_H region has interesting general implications as discussed below.

The nature of Eμ/CTCF interactions

Eμ interaction with IGCR1-associated CBEs or V_H region-associated CBEs has substantially different functional consequences. On WT alleles both Eμ and IGCR1 are necessary to bring a pre-folded V_H region into proximity of the 3' *IgH* domain that contains D_H and J_H gene segments. Mutation of either of these regulatory sequences disrupts distal V_H to D/J_H association as well as increases distal V_H sterile transcription (Figures 6B, S9, (Gerasimova et al., 2015; Verma-Gaur et al., 2012)). By contrast, Eμ interaction with proximal V_H region CBEs (on IGCR1-mutated alleles) leads to highly specific activation of the nearest V_H gene and loss of distal V_H proximity to the D/J_H domain (Figures 4 and 6). We infer that interaction of Eμ with IGCR1 permits flexibility that is essential to generate a diverse repertoire, whereas Eμ interaction with proximal V_H CBEs is functionally precise.

Our observations are consistent with the idea that Eμ and IGCR1 together present a composite interface that serves as a bridge, or linker, to distal V_H genes, whereas proximal V_H CBEs serve only as classical boundaries to limit Eμ activity. One possibility for the different outcomes of Eμ interaction may be the organization of CBEs, with IGCR1 containing two oppositely-orientated CBEs compared to single CBEs associated with each proximal V_H gene segment. The highly specific functional interaction of Eμ with the closest CTCF site on IGCR1-deficient alleles evokes the recently proposed loop-extrusion model for choice of CTCF interacting sites (Goloborodko et al., 2016; Rao et al., 2017). It will be interesting to ascertain whether Eμ orientation plays a role in the site selection process as has been proposed for CTCF sites. The functional complexity of CTCF associations with enhancers revealed here may apply more broadly in specifying tissue-specific gene expression.

Further implications for Eμ-dependent V(D)J recombination

Eμ/IGCR1 interactions affect V_H recombination by at least two mechanisms. First, altered Eμ-dependent looping in the absence of IGCR1 brings specific V_H sequences, and exclude others, from spatial proximity to the J_H-associated RC. Second, distribution of RAG1/2 proteins change markedly on IGCR1-mutated alleles compared to the well-defined localized RC on WT alleles (Figure 5). It is possible that V_H-associated RAG1/2 can initiate recombination on IGCR1-mutated alleles in addition to the normal J_H-associated RAG1/2. For example, RAG1/2 bound near the V_H81X RSS on IGCR1-mutated alleles could readily track to the 5' DQ52 RSS to promote V_H81X to germline DQ52 rearrangements. Conversely, RAG1/2 bound to the 5' DQ52 RSS could track to a spatially proximal V_H81X RSS on IGCR1-mutated alleles, but not on WT alleles where its movements would be constrained by Eμ-IGCR1 association. Our working hypothesis is that availability of a few

V_H gene segments within the RAG1/2 tracking domain, generation of an ectopic RC, and spatial proximity of V_H gene segments to the J_H-associated RC jointly increase recombination efficiency on IGCR1-deficient alleles resulting in premature recombination.

Mechanistic features of Eμ-IGCR1 identified in this study also apply to DJ_H recombined alleles (Figure 7, right). After D_H recombination of WT alleles Eμ would still be sequestered by IGCR1, the RC would be restricted to the DJ_H region, and distal V_H genes would be spatially proximal to the DJ_H part of the locus. This configuration would permit diverse V_H genes to recombine, but with low efficiency. By contrast, in the absence of IGCR1 Eμ would loop to V_H81X, the RC would spread to the proximal V_H gene segments, and distal V_H gene segments would move away from the 3' *IgH* domain. This configuration would yield highly efficient V_H81X recombination to the DJ_H junction, but relatively poor use of distal V_H gene segments. We conjecture that in order to gain diversity the immune system evolved away from efficiency by undercutting the ability of Eμ to loop to very specific sites.

How does IGCR1 enforce V_H diversity?

There is ample evidence that proximal and distal V_H recombination is regulated differently. Most prominently, loss of either transcription factor Pax5 or YY1 significantly reduces distal V_H recombination without affecting proximal V_H recombination (Fuxa et al., 2004; Hesslein et al., 2003; Liu et al., 2007). It is likely, therefore, that mechanisms by which IGCR1 enforces diverse V_H gene segment choice are different for proximal versus distal gene segments. As discussed above, distal V_H gene may achieve regulated access to DJ_H junctions via a composite Eμ/IGCR1 interface. This does not explain how IGCR1 over-rules the tendency of Eμ to loop to the nearest CTCF binding site. Our working model draws upon our earlier observation that sequences 5' of DFL16.1 (that are very likely to be IGCR1) interact in a CTCF-dependent manner with proximal V_H region (Gerasimova et al., 2015; Guo et al., 2011a) and the proposal that CBE1 of IGCR1 interacts with proximal V_H-associated CTCF binding sites (Lin et al., 2015). We propose that IGCR1 contacts one, or a subset, of CTCF sites in individual cells. Eμ is brought to each of these different CTCF sites via Eμ/IGCR1 interaction, thereby activating different V_H gene segments in individual cells. The key distinction between WT and IGCR1-deficient alleles is how Eμ gets close to V_H gene segments. Our hypothesis is that this happens via CTCF/CTCF interactions between IGCR1 and proximal V_H CTCF sites on WT alleles, whereas it happens by loop extrusion from Eμ to the nearest CTCF site on IGCR1-deficient alleles. By sequestering Eμ's tendency to form highly specific loops with proximal CTCF binding sites, IGCR1 ensures generation of a diverse *IgH* repertoire.

STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to and will be fulfilled by the Lead Contact, Ranjan Sen (Senranja@grc.nia.nih.gov).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice—WT 129 strain mice were purchased from Charles River Laboratories International. IGCR1-deleted mice were as described previously (Guo et al., 2011b). All animal experiments were performed under protocols approved by the Institutional Animal Care and Use Committee of Boston Children's Hospital.

Cell lines—Abelson virus transformed pro-B cell lines CBE$^{-/-}$(1) and CBE$^{-/-}$(2) are deficient for RAG2 and homozygous for *IgH* alleles in which both CBEs in IGCR1 are mutated (Guo et al., 2011b). RAG2-deficient pro-B cells contain WT *IgH* alleles, and Eμ$^{-/-}$ pro-B cells have a 220bp deletion of the intronic enhancer on both *IgH* alleles. Both lines lack RAG2 (Chakraborty et al., 2009). EOMA, an endothelial cell line from 129 strain of mice, was purchased from ATCC. CBE$^{-/-}$ (2)/Eμ$^{-/-}$(1), CBE$^{-/-}$(2)/Eμ$^{-/-}$(2), D345/IGCR1$^{-/-}$(1), D345/IGCR1$^{-/-}$(2), CBE$^{-/-}$(1)/C1$^{-/-}$, WT/C1$^{-/-}$, CBE$^{-/-}$(1)/C2$^{-/-}$, WT/C2$^{-/-}$, CBE$^{-/-}$(1)/C1$^{-/-}$C2$^{-/-}$ and WT/C1$^{-/-}$C2$^{-/-}$ were generated with CRISPR/Cas9 system, described below. Cells were cultured in RPMI medium with 10% fetal bovine serum (FBS) and 56μM 2-mercaptoethanol at 37°C in a 5% CO2 humidified atmosphere.

METHOD DETAILS

Isolation of pro-B cells from bone marrow—Bone marrow derived pro-B cells, marked as B220^{+}IgM^{-}CD43^{+}, were purified from WT 129 or IGCR1$^{-/-}$ mice.

Chromosome conformation capture (3C)—Chromosome conformation capture (3C) assays were performed as described (Wuerffel et al., 2007) using HindIII to digest crosslinked chromatin. Briefly, cells were crosslinked with 1% (v/v) formaldehyde solution in HBSS buffer. After nuclear extraction, cross-linked DNA was digested overnight with HindIII overnight. Reaction was terminated with SDS, followed by neutralization with Triton™ X-100 solution. Digested DNA was treated with T4 DNA ligase overnight. RNA and protein were digested by RNAse A and proteinase K. 2-Propanolprecipitated DNA was further purified by extraction with phenol and chloroform, followed by 0.3M sodium acetate (ph 5.2) and 65% ethanol. Quant-iT™ PicoGreen® dsDNA Assay Kit was used to quantify DNA. 3C ligation products were measured by Taqman quantitative PCR technology (Hagege et al., 2007). 3C results between experiments were first normalized using the α-amylase gene. Calculation was $X_{\text{specific primer}} = 2^{\text{Ct (α-amylase - specific primer)}}$. To determine primer efficiency, equal moles of bacterial artificial chromosomes covering the genomic region under study, RP23-351J19 (231.6 kb, 5.55 μg), RP23-269D13 (202.8 kb, 4.86 μg), and RP23-363G23 (152.2 kb, 3.65μg) were mixed, digested with HindIII, re-ligated, and measured by Taqman quantitative PCR technology. Calculation of primer efficiency relative to the Eμ-V_H81X was $E_{\text{specific primer}} = 2\text{Ct}^{\text{ (Eμ-VH81X - specific primer)}}$. 3C results between experiments were additionally normalized with respect to primer efficiency. Calculation was Relative Interaction= $X_{\text{specific primer}} / E_{\text{specific primer}}$.

Fluorescent *in situ* hybridization (FISH)—FISH was performed as described (Guo et al., 2011a). Briefly, cells were fixed on poly-L-lysine coated slides at a concentration of 5×10^6 cells/ml using 200,000 cells per slide. Slides were kept for 15 minutes at 37°C, washed in PBS, followed by fixation with 4% paraformaldehyde in PBS. Cells were further

washed with 0.1 M Tris-HCl (pH 7.4), followed by PBS. Fixed cells were treated with 100 μg/ml RNase A in PBS, permeabilized using 0.5% saponin/0.5 % Triton-X 100/PBS for 30 min at room temperature, and rinsed in PBS. DNA probes were denatured at 75°C for 5 min and applied to cells that were denatured in formamide at 73°C followed by incubation in a dark humid chamber.

Position-specific 3-10kb probes were generated by PCR using BAC templates or genomic DNA with the primers listed in Table S3. Images were acquired using a Nikon T200 microscope equipped with a 100× lens and motorized 100μm Piezo Z-axis stage (Applied Scientific Instrumentation). Depending on the size of the nucleus, 30–40 serial optical sections spaced by 0.2μm were acquired. The data sets were deconvolved using NIS-Elements software (Nikon). Statistical analyses of spatial distance measurements were carried out using a two-sample Kolmogorov-Smirnov test in R as previously described (Gerasimova et al., 2015). 100 or more distance measures were analyzed by Kolmogorov-Smirnov statistical analyses (https://www.wessa.net/rwasp_Reddy-Moores%20K-S%20Test.wasp).

YY1 and CTCF knockdown—YY1 and CTCF shRNA lentiviral plasmids were purchased from Sigma-Aldrich. The empty vector pLKO.1 was used as the control. Virus preparation and infection were carried out according to Addgene protocol (https://www.addgene.org/tools/protocols/plko/). Briefly, shRNA and packaging plasmid (psPAX and pMD2.G) were added into 293T with Bio-T. Lentivirus were harvested by high speed centrifugation (25000rpm, 2 hour) after three days of culture. After infection, cells were cultured in puromycin (2μg/ml) for 5 days, and then harvested for RNA, protein and FISH analyses.

***Rag2* expression—**Lentiviral particles expressing *Rag2* were generated as described (Kutner et al., 2009) by transiently transfecting 293T cells with lentiviral plasmid containing *Rag2* and puromycin resistance DNA fragment (pHIV-RAG2-IRES-puro) along with helper plasmids pMD2.G and psPAX2 using Bio-T reagent. pHIV-IRES-puro was used as control. The supernatant containing the virus was collected at 72h after transfection and concentrated by ultracentrifugation for 2h at 25,000 rpm and 20°C over a 20% sucrose cushion. Supernatant was removed after ultracentrifugation, and 200μl PBS were added to the tube. Fresh virus was prepared to transduction. All procedures involving lentiviruses were performed under BSL2 conditions.

DJ_H/VDJ_H recombination assays—Genomic DNA was purified from sorted bone marrow pro-B ($IgM^-B220^+CD43^+$) pro-B cells from WT or $IGCR1^{-/-}$ mice. Fivefold serial dilutions of genomic DNA (200ng, 40ng, 8ng) were used to perform PCR to analyze DJ_H rearrangements. Primers used in this assay are listed in Table S3. Primers flanking the ROSA26 gene were used as a loading control under the same conditions. GeneRuler 100bp Plus, 100 bp DNA Ladder, or 1 Kb Plus DNA Ladder were used to confirm PCR product size. V_H7183-DJ_H1 rearrangements in pro-B cells were amplified with LongAmp® Taq DNA Polymerase, cloned into pGEM®-T vector, transformed into MAX Efficiency® DH5α competent cells and sequenced. The D_H family was identified according to NCBI Genbank (http://www.ncbi.nlm.nih.gov/igblast/showGermline.cgi?

organism=mouse&chainType=DH&seqType=nucleotide&functionClass=1). For quantitation, ethidium bromide-stained agarose gels were scanned and analyzed for gray density with GeneTools software from Syngene.

RAG2-deficient pro-B cell lines carrying IGCR1-mutated alleles ($CBE^{-/-}$(1) and (2)) or WT *IgH* alleles (WT) were infected with RAG2-expressing lentivirus, and were cultured in complete medium with puromycin (2μg/ml). After 4 days ($CBE^{-/-}$(2)), 5 days ($CBE^{-/-}$(1)) or 14 days ($CBE^{-/-}$(1), $CBE^{-/-}$(2), $CBE^{-/-}$(2)/$E\mu^{-/-}$(1), $CBE^{-/-}$(2)/$E\mu^{-/-}$(2), $E\mu^{-/-}$) selection with puromycin, cells were harvested. Genomic DNA was collected with DNeasy Blood & Tissue Kit, and the DNA was used to analyze DJ_H/VDJ_H rearrangements with HotStarTaq DNA Polymerase as described as above.

VDJ_H deep sequencing assays—VDJ_H deep sequencing assays were performed as previously described (Hu et al., 2016). Briefly, 40ng genomic DNA from each sample was sonicated to an average size of 750bp. Sonicated DNA was hybridized with Bio-J_H1 primer, purified with Dynabeads C1 streptavidin beads and used for library generation as described (Lin et al., 2016). Paired-end reads were generated by Illumina HiSeq 2500 sequencer. Samples were separated using barcodes present in Read-1 (Table S4). Adapters, if present were removed by cutadapt and bad quality bases (< Q33) were trimmed from Read-2 keeping a minimum length of 50 bases. The reads (Read-2 only) were aligned to 600 bases of different V-regions using bowtie2. Reads which aligned to different V-regions with an alignment quality > 20 were counted. Genome build utilized mm9 *IgH* V-regions.

Chromatin Immunoprecipitation (ChIP)—ChIP analysis of modified histones and RAG1 and RAG2 was done as described (Chakraborty et al., 2007; Ji et al., 2010). 20×10^6 cells were resuspended in 9ml RPMI containing 2% serum in a 15ml conical flask, followed by 1ml of 10% HCHO and rotated at room temperature for 15 min. Glycine was added to terminate the crosslinking to a final concentration of 0.125M and rocked for 5 min at room temperature. Cells were washed twice with 10 ml cold PBS (2K, 4 min, 4°C). After washes, cell pellet was resuspended in 300 μl RIPA buffer (10 mM Tris pH 7.4, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 0.8M NaCl), and incubated on ice for 10 min. Chromatin was sonicated using a water bath sonicator (Diagenode) (5 × 5 min, high power) resulting in DNA of 300 to 700 bp. Samples were centrifuged (14K, 10 min, 4°C) and the supernatant transferred to a new 1.5ml tube. 40 μg Dynabeads Protein G (Thermofisher cat# 1003D) was washed with 1ml RIPA buffer for preclearing the chromatin. The chromatin was then added to the protein G beads and rotated for 1 hr at 4°C. Beads were removed using magnet and the supernatant was transferred to a new tube containing second batch of protein G beads and rotated for 1 hr at 4°C. Supernatant was transferred to a new tube. 10% chromatin was kept to be used as input. Precleared chromatin was divided equally into two tubes. To each tube heparin was added to final concentration of 500 ng/ml and gelatin from cold water to a final concentration 2%. One tube was incubated with 5μg anti-RAG1/RAG2 and another tube with 5 μg normal rabbit IgG as a control and rocked at 4C overnight. Dynabeads protein G (preblocked with 2% BSA and heparin (500 ng/ml) for 1 hr at RT) were added to the tube and rotated for 3 hr at 4°C, centrifuged (7K, 2 min, 4°C), and the beads were washed sequentially (2×10 min, 4°C) with 1 ml RIPA (0.3 M NaCl), 1

ml RIPA (0.8 M NaCl), LiCl buffer (0.25M LiCl, 0.5% NP-40, 0.5% NaDOC), followed by one wash with 1ml of TE + 0.2 % Triton X-100 and finally once with 1 ml TE. Beads were resuspended in 100 μl TE containing 50 μg Proteinase K and 0.25% SDS and incubated at 65°C for 5h. The "input" chromatin was treated identically. After phenol/chloroform extraction, the aqueous phase was adjusted to 0.2M NaCl and DNA precipitated with 20 μg glycogen and 1 ml ethanol. The DNA pellet was washed with 70% ethanol and resuspended in water. The primers which are used to detect the *IgH* locus and gamma actin promoter are listed in Table S3. qPCR was performed in duplicates using SYBr green in applied biosystem. A standard curve generated for each primer set was used to quantitate DNA recovered in pull down sample and the input chromatin. IP/Inputcorr was calculated as ((IPsp – IPrIg)/Input) × 1000, where IPsp and IPrig are the amount of DNA recovered in IPs with the specific antibody and rabbit IgG as described (Ji et al., 2010).

Histone ChIP (anti-H3K4me3, anti-H3K9ac, anti-H3K9me2), were carried out as described above. After DNA precipitation, real time PCR was carried out with 200pg of immunoprecipitated or input DNA. The abundance of specific sequences in the immunoprecipitate relative to that in input DNA was calculated as previously described (Chakraborty et al., 2007). Calculation is Relative Abundance (specific amplicon) = $2^{Ct\ (input\text{-}IP)}$. Oligonucleotides used for real-time PCR analysis are provided in Table S3.

CTCF ChIP-Seq—CTCF ChIP-Seq was extracted from (Lin et al., 2012) (GEO: GSM987805). Direction of CTCF was analyzed with software designed by Yan Cui at University of Tennessee Health Science Center (http://insulatordb.uthsc.edu/), and determined by using higher score and better match as criteria.

RNA isolation, RT–PCR and RNA-Seq—Total RNA was isolated using RNeasy Plus Mini Kit. 1μg RNA was used to generate cDNA with Superscript III with random hexamers according to manufacturer's protocols. Approximately 1/40 of the reverse-transcription-generated cDNA was analyzed with iTaq™ Universal SYBR. Primers that were used for PCR are provided in Table S3. Briefly, the RNA-Seq library was made with the Illumina TruSeq Stranded mRNA sample preparation kit following the manufacturer's recommended procedure, and the resulting libraries were QCed using Fragment Analyzer and pooled together. The pooled library concentration was further quantified using KAPA library quantification qPCR kit, and sequenced on Illumina NextSeq 500 using a High Output 75 cycle sequencing kit targeting 20–30 million single reads of 75bp. For each RNA-seq sample, the sequencing reads were mapped to mouse genome mm9 using TopHat with default settings. Signal tracks were generated using bedtools and UCSC Genome Browser software. Signals in each sample were normalized to total number of reads and then scaled by multiplying a constant N (N=100,000,000).

CRISPR/Cas9 System—pHIV-U6-gRNA-Cas9-Blast plasmid was generated by insertion of U6 promoter and gRNA DNA fragment into lentiCas9-Blast. Two different gRNA sequences, targeting one DNA locus, were separately cloned into plasmid lentiCRISPR v2 and pHIV-U6-gRNA-Cas9-Blast as previously described (Sanjana et al., 2014) with appropriate primer (Table S3), followed by transformation into One Shot® Stbl3™ competent cells. Lentiviral particles were generated and harvested according to protocol

(Sanjana et al., 2014). CBE$^{-/-}$(1), CBE$^{-/-}$(2) or D345 cell lines were infected with freshly prepared gRNA/Cas9 expressing lentivirus. Cells were cultured with the presence of both puromycin (2μg/ml) and blasticidin (15μg/ml) for 4 days, followed by single cell cloning. Single cell clones were collected, and genomic DNA was purified with Quick-DNA™ Universal 96 Kit. PCR was employed for detection of deletion of Eμ (CBE$^{-/-}$(2)) or IGCR1 (D345) with appropriate primer (Table S3). CBE$^{-/-}$(2)/Eμ$^{-/-}$(1), CBE$^{-/-}$(2)/Eμ$^{-/-}$(2), D345/IGCR1$^{-/-}$(1) and D345/IGCR1$^{-/-}$(2), were harvested and used for analysis of RNA and or RAG1 ChIP. CBE$^{-/-}$(1)/C1$^{-/-}$, CBE$^{-/-}$(1)/C2$^{-/-}$ and CBE$^{-/-}$(1)/C1$^{-/-}$C2$^{-/-}$ were used for VDJ$_H$ deep sequencing assays.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis, including D-statistics and P-values, were carried out for FISH data (Figure 1C, 2B, 6B). Detailed statistical analysis result for FISH is listed on Table S1.

DATA AND SOFTWARE AVAILABILITY

Raw data of images are deposited in Mendeley (http://dx.doi.org/10.17632/5f23hhtpnj.1). Raw data for RNA-Seq and VDJ-Seq are deposited at NCBI (GSE109631 and GSE110090). We also provide five supplemental data files (Tables S1–5). Software for Kolmogorov-Smirnov test can be used online (https://www.wessa.net/rwasp_Reddy-Moores%20K-S%20Test.wasp). RNA-Seq analyses were carried out using TopHat (https://ccb.jhu.edu/software/tophat/index.shtml), Bedtools (http://bedtools.readthedocs.io/en/latest/) and UCSC Genome Browser software (http://hgdownload.soe.ucsc.edu/admin/exe/linux.x86_64/). VDJ-Seq analyses were carried out using Cutadapt (http://cutadapt.readthedocs.io/en/stable/guide.html) and Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Frederick W. Alt, Sherry Lin and Suvi Jain (Harvard Medical School) for generously providing IGCR1-mutated cell lines and sorted bone marrow pro-B cells from IGCR1-deleted mice, and for critiquing the manuscript. Antibodies against RAG1 and RAG2 were generously provided by Elizabeth Corbett and David G. Schatz (Yale University), and we thank Amy L. Kenter (University of Illinois College of Medicine) for comments on the manuscript. Marc Raley (NIDA) contributed to preparation of the figures. This research was supported by the Intramural Research Program of the NIH, National Institute on Aging.

References

Afshar R, Pierce S, Bolland DJ, Corcoran A, Oltz EM. Regulation of IgH gene assembly: role of the intronic enhancer and 5'DQ52 region in targeting DHJH recombination. J Immunol. 2006; 176:2439–2447. [PubMed: 16456003]

Bossen C, Mansson R, Murre C. Chromatin topology and the regulation of antigen receptor assembly. Annu Rev Immunol. 2012; 30:337–356. [PubMed: 22224771]

Chakraborty T, Chowdhury D, Keyes A, Jani A, Subrahmanyam R, Ivanova I, Sen R. Repeat organization and epigenetic regulation of the DH-Cmu domain of the immunoglobulin heavy-chain gene locus. Mol Cell. 2007; 27:842–850. [PubMed: 17803947]

Chakraborty T, Perlot T, Subrahmanyam R, Jani A, Goff PH, Zhang Y, Ivanova I, Alt FW, Sen R. A 220-nucleotide deletion of the intronic enhancer reveals an epigenetic hierarchy in immunoglobulin heavy chain locus activation. J Exp Med. 2009; 206:1019–1027. [PubMed: 19414554]

Chen L, Carico Z, Shih HY, Krangel MS. A discrete chromatin loop in the mouse Tcra-Tcrd locus shapes the TCRdelta and TCRalpha repertoires. Nat Immunol. 2015; 16:1085–1093. [PubMed: 26258942]

Chen L, Zhao L, Alt FW, Krangel MS. An Ectopic CTCF Binding Element Inhibits Tcrd Rearrangement by Limiting Contact between Vdelta and Ddelta Gene Segments. J Immunol. 2016; 197:3188–3197. [PubMed: 27613698]

Choi NM, Loguercio S, Verma-Gaur J, Degner SC, Torkamani A, Su AI, Oltz EM, Artyomov M, Feeney AJ. Deep sequencing of the murine IgH repertoire reveals complex regulation of nonrandom V gene rearrangement frequencies. J Immunol. 2013; 191:2393–2402. [PubMed: 23898036]

Degner SC, Verma-Gaur J, Wong TP, Bossen C, Iverson GM, Torkamani A, Vettermann C, Lin YC, Ju Z, Schulz D, et al. CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. Proc Natl Acad Sci U S A. 2011; 108:9566–9571. [PubMed: 21606361]

Donohoe ME, Zhang LF, Xu N, Shi Y, Lee JT. Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch. Mol Cell. 2007; 25:43–56. [PubMed: 17218270]

Ebert A, McManus S, Tagoh H, Medvedovic J, Salvagiotto G, Novatchkova M, Tamir I, Sommer A, Jaritz M, Busslinger M. The distal V(H) gene cluster of the Igh locus contains distinct regulatory elements with Pax5 transcription factor-dependent activity in pro-B cells. Immunity. 2011; 34:175–187. [PubMed: 21349430]

Fuxa M, Skok J, Souabni A, Salvagiotto G, Roldan E, Busslinger M. Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. Genes Dev. 2004; 18:411–422. [PubMed: 15004008]

Gerasimova T, Guo C, Ghosh A, Qiu X, Montefiori L, Verma-Gaur J, Choi NM, Feeney AJ, Sen R. A structural hierarchy mediated by multiple nuclear factors establishes IgH locus conformation. Genes Dev. 2015; 29:1683–1695. [PubMed: 26302788]

Goloborodko A, Imakaev MV, Marko JF, Mirny L. Compaction and segregation of sister chromatids via active loop extrusion. Elife. 2016; 5

Guo C, Gerasimova T, Hao H, Ivanova I, Chakraborty T, Selimyan R, Oltz EM, Sen R. Two forms of loops generate the chromatin conformation of the immunoglobulin heavy-chain gene locus. Cell. 2011a; 147:332–343. [PubMed: 21982154]

Guo C, Yoon HS, Franklin A, Jain S, Ebert A, Cheng HL, Hansen E, Despo O, Bossen C, Vettermann C, et al. CTCF-binding elements mediate control of V(D)J recombination. Nature. 2011b; 477:424–430. [PubMed: 21909113]

Hagege H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, de Laat W, Forne T. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat Protoc. 2007; 2:1722–1733. [PubMed: 17641637]

Hesslein DG, Pflugh DL, Chowdhury D, Bothwell AL, Sen R, Schatz DG. Pax5 is required for recombination of transcribed, acetylated, 5' IgH V gene segments. Genes Dev. 2003; 17:37–42. [PubMed: 12514097]

Hu J, Meyers RM, Dong J, Panchakshari RA, Alt FW, Frock RL. Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. Nat Protoc. 2016; 11:853–871. [PubMed: 27031497]

Hu J, Zhang Y, Zhao L, Frock RL, Du Z, Meyers RM, Meng FL, Schatz DG, Alt FW. Chromosomal Loop Domains Direct the Recombination of Antigen Receptor Genes. Cell. 2015; 163:947–959. [PubMed: 26593423]

Ji Y, Resch W, Corbett E, Yamane A, Casellas R, Schatz DG. The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. Cell. 2010; 141:419–431. [PubMed: 20398922]

Jung D, Giallourakis C, Mostoslavsky R, Alt FW. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. Annu Rev Immunol. 2006; 24:541–570. [PubMed: 16551259]

Kumari G, Sen R. Chromatin Interactions in the Control of Immunoglobulin Heavy Chain Gene Assembly. Adv Immunol. 2015; 128:41–92. [PubMed: 26477365]

Kutner RH, Zhang XY, Reiser J. Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors. Nat Protoc. 2009; 4:495–505. [PubMed: 19300443]

Lin SG, Ba Z, Du Z, Zhang Y, Hu J, Alt FW. Highly sensitive and unbiased approach for elucidating antibody repertoires. Proc Natl Acad Sci U S A. 2016; 113:7846–7851. [PubMed: 27354528]

Lin SG, Guo C, Su A, Zhang Y, Alt FW. CTCF-binding elements 1 and 2 in the Igh intergenic control region cooperatively regulate V(D)J recombination. Proc Natl Acad Sci U S A. 2015; 112:1815–1820. [PubMed: 25624508]

Lin YC, Benner C, Mansson R, Heinz S, Miyazaki K, Miyazaki M, Chandra V, Bossen C, Glass CK, Murre C. Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. Nat Immunol. 2012; 13:1196–1204. [PubMed: 23064439]

Liu H, Schmidt-Supprian M, Shi Y, Hobeika E, Barteneva N, Jumaa H, Pelanda R, Reth M, Skok J, Rajewsky K, et al. Yin Yang 1 is a critical regulator of B-cell development. Genes Dev. 2007; 21:1179–1189. [PubMed: 17504937]

Lucas JS, Zhang Y, Dudko OK, Murre C. 3D trajectories adopted by coding and regulatory DNA elements: first-passage times for genomic interactions. Cell. 2014; 158:339–352. [PubMed: 24998931]

Majumder K, Koues OI, Chan EA, Kyle KE, Horowitz JE, Yang-Iott K, Bassing CH, Taniuchi I, Krangel MS, Oltz EM. Lineage-specific compaction of Tcrb requires a chromatin barrier to protect the function of a long-range tethering element. J Exp Med. 2015; 212:107–120. [PubMed: 25512470]

Medvedovic J, Ebert A, Tagoh H, Tamir IM, Schwickert TA, Novatchkova M, Sun Q, Huis In 't Veld PJ, Guo C, Yoon HS, et al. Flexible long-range loops in the VH gene region of the Igh locus facilitate the generation of a diverse antibody repertoire. Immunity. 2013; 39:229–244. [PubMed: 23973221]

Perlot T, Alt FW, Bassing CH, Suh H, Pinaud E. Elucidation of IgH intronic enhancer functions via germ-line deletion. Proc Natl Acad Sci U S A. 2005; 102:14362–14367. [PubMed: 16186486]

Rao SSP, Huang SC, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon KR, Sanborn AL, Johnstone SE, Bascom GD, Bochkov ID, et al. Cohesin Loss Eliminates All Loop Domains. Cell. 2017; 171:305–320 e324. [PubMed: 28985562]

Rawat P, Jalan M, Sadhu A, Kanaujia A, Srivastava M. Chromatin Domain Organization of the TCRb Locus and Its Perturbation by Ectopic CTCF Binding. Mol Cell Biol. 2017; 37

Retter I, Chevillard C, Scharfe M, Conrad A, Hafner M, Im TH, Ludewig M, Nordsiek G, Severitt S, Thies S, et al. Sequence and characterization of the Ig heavy chain constant and partial variable region of the mouse strain 129S1. J Immunol. 2007; 179:2419–2427. [PubMed: 17675503]

Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. Nat Methods. 2014; 11:783–784. [PubMed: 25075903]

Schatz DG, Ji Y. Recombination centres and the orchestration of V(D)J recombination. Nat Rev Immunol. 2011; 11:251–263. [PubMed: 21394103]

Subrahmanyam R, Du H, Ivanova I, Chakraborty T, Ji Y, Zhang Y, Alt FW, Schatz DG, Sen R. Localized epigenetic changes induced by DH recombination restricts recombinase to DJH junctions. Nat Immunol. 2012; 13:1205–1212. [PubMed: 23104096]

Verma-Gaur J, Torkamani A, Schaffer L, Head SR, Schork NJ, Feeney AJ. Noncoding transcription within the Igh distal V(H) region at PAIR elements affects the 3D structure of the Igh locus in pro-B cells. Proc Natl Acad Sci U S A. 2012; 109:17004–17009. [PubMed: 23027941]

Volpi SA, Verma-Gaur J, Hassan R, Ju Z, Roa S, Chatterjee S, Werling U, Hou H Jr, Will B, Steidl U, et al. Germline deletion of Igh 3' regulatory region elements hs 5, 6, 7 (hs5-7) affects B cell-specific regulation, rearrangement, and insulation of the Igh locus. J Immunol. 2012; 188:2556–2566. [PubMed: 22345664]

Wuerffel R, Wang L, Grigera F, Manis J, Selsing E, Perlot T, Alt FW, Cogne M, Pinaud E, Kenter AL. S-S synapsis during class switch recombination is promoted by distantly located transcriptional elements and activation-induced deaminase. Immunity. 2007; 27:711–722. [PubMed: 17980632]

Zhao L, Frock RL, Du Z, Hu J, Chen L, Krangel MS, Alt FW. Orientation-specific RAG activity in chromosomal loop domains contributes to Tcrd V(D)J recombination during T cell development. J Exp Med. 2016; 213:1921–1936. [PubMed: 27526713]

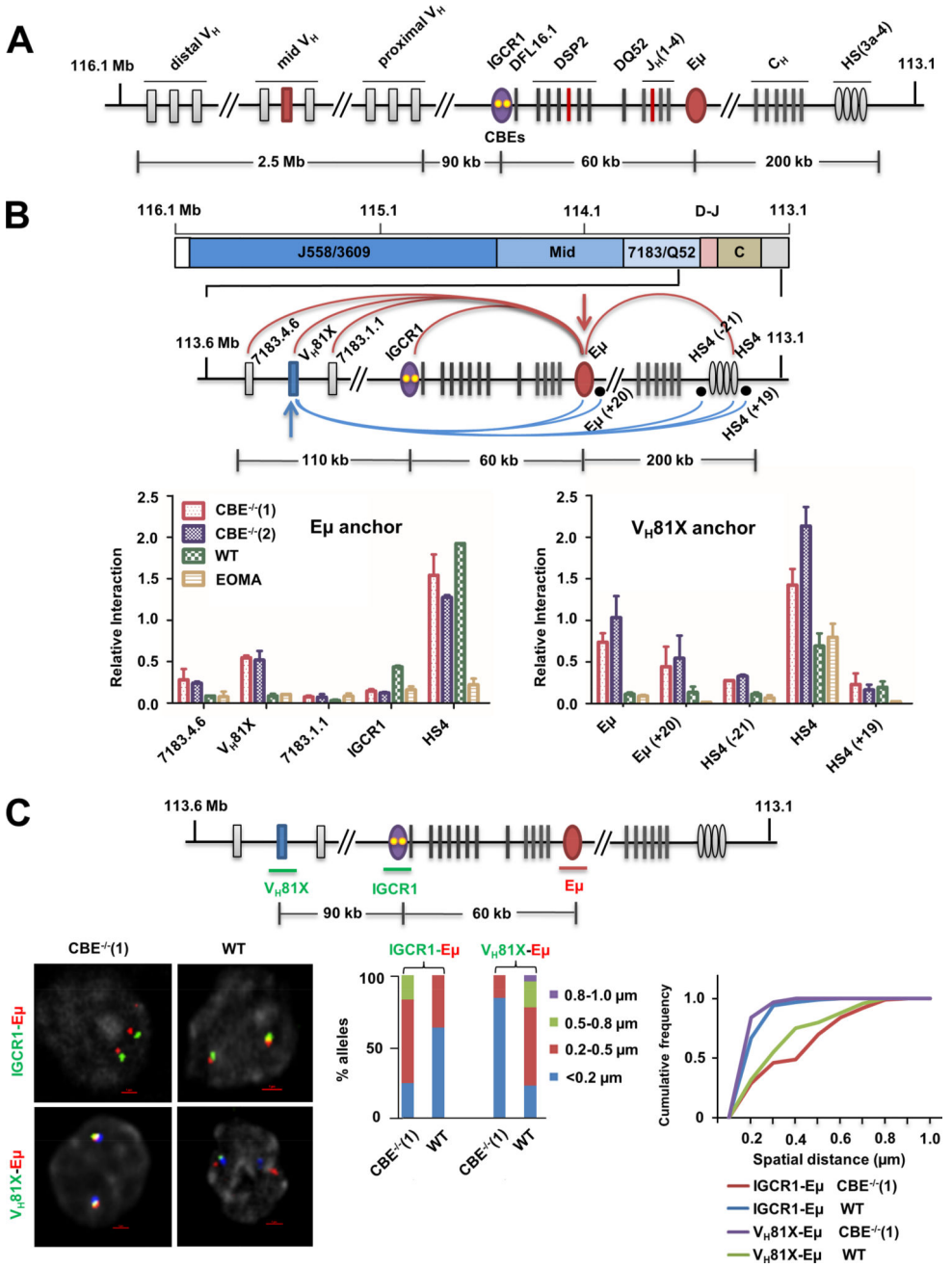


Figure 1. Altered looping of IGCR1-mutated *IgH* alleles

A. Schematic representation of the murine germline (unrearranged) *IgH* locus. Variable gene segments (V_H) occupy much of the locus starting at the 5' end (left). V_H gene segments are broadly categorized into three gene clusters: distal (comprising V_HJ558 and V_H3609 genes), mid (comprising several gene families) and proximal (comprising V_H7183 and V_HQ52 genes). A gap of 90 kb separates the 3'-most V_H gene segments from the 5'-most diversity (D_H) gene segment, DFL16.1. A family of closely-related DSP2 gene segments and a unique 3' DQ52 gene segment spread over 60 kb complete the D_H genes. Four joining gene segments (J_H) are located within 1 kb 3' of DQ52, followed by exons that encode the

constant parts of antibody heavy chains (labeled C_H). *IgH* gene assembly by recombination puts together one V_H, one D_H and one J_H gene segment (highlighted in red) to encode the variable part of antibody heavy chains. Three key regulatory sequences are indicated as ovals. The IGCR1 is located just 5' of DFL16.1 (purple; yellow dots represent 2 functional CTCF binding elements (CBEs)), the enhancer Eμ is located in the intron between J_H gene segments and Cμ exons (red), and the 3' regulatory region (HS3a-4) is a cluster of DNase1-hypersensitive sites located beyond the last C_H exons (gray).

B. Chromosome conformation capture (3C) analyses of wild-type (WT) or IGCR1-mutated alleles in which both CBEs have been mutated. Pro-B cell lines carrying normal (WT) or IGCR1-mutated *IgH* alleles ($CBE^{-/-}$(1) and $CBE^{-/-}$(2)), and a non-B cell line of endothelial origin (EOMA), were subject to 3C analysis using anchor primers located at Eμ (left) or close to the 3'-most functional V_H gene, V_H81X (right). All pro-B cells were recombinase (RAG2) deficient to maintain *IgH* alleles in germline configuration. The region of interest is expanded below a scale view of the *IgH* locus (in which V_H gene families are shown, the part marked D-J contains all D_H and J_H gene segments, and the part marked C contains all C_H exons). Locations of anchor primers (arrows) and sites queried for interactions are indicated. Relative interaction frequency (Y axis) was calculated as described (Guo et al., 2011a) after normalization for ligation efficiency using α-amylase-specific primers. Amplification efficiency for different primer pairs was calculated based on restriction and re-ligation of 3 bacterial artificial chromosomes (BACs) that span the region of interest. Positions and sequences of primers are provided in Table S3. Data are shown as mean ± standard error of measurement (SEM) of two biological replicate 3C assays.

C. Fluorescent *in situ* hybridization (FISH) analyses of WT and IGCR1-mutated alleles. FISH was carried out as previously described (Guo et al., 2011a) using short probes located as shown in the schematic. Representative nuclei containing WT or IGCR1-mutated ($CBE^{-/-}$(1)) alleles are shown; probes were color-coded as indicated to the left of the micrograph. BAC RP23-201H14 was used to mark *IgH* alleles (blue) in selected FISH analyses. It is located 200 kb 3' of HS4. Spatial distances between probes were measured after deconvolution of images from 100 nuclei. Bar graphs show the percentage of *IgH* alleles in which the distance between indicated probe pairs fell in the ranges shown in different colors. The cumulative frequency distribution of spatial distances for each color-coded probe combination is shown to the right of the bar graph. D-statistics and P-values for differences between WT and IGCR1-mutated alleles were calculated using a two-sample Kolmogorov-Smirnov test (Table S1). Data from biological replicate experiments are shown in Figure S3B.

See also Figure S1–3 and Table S1, 3.

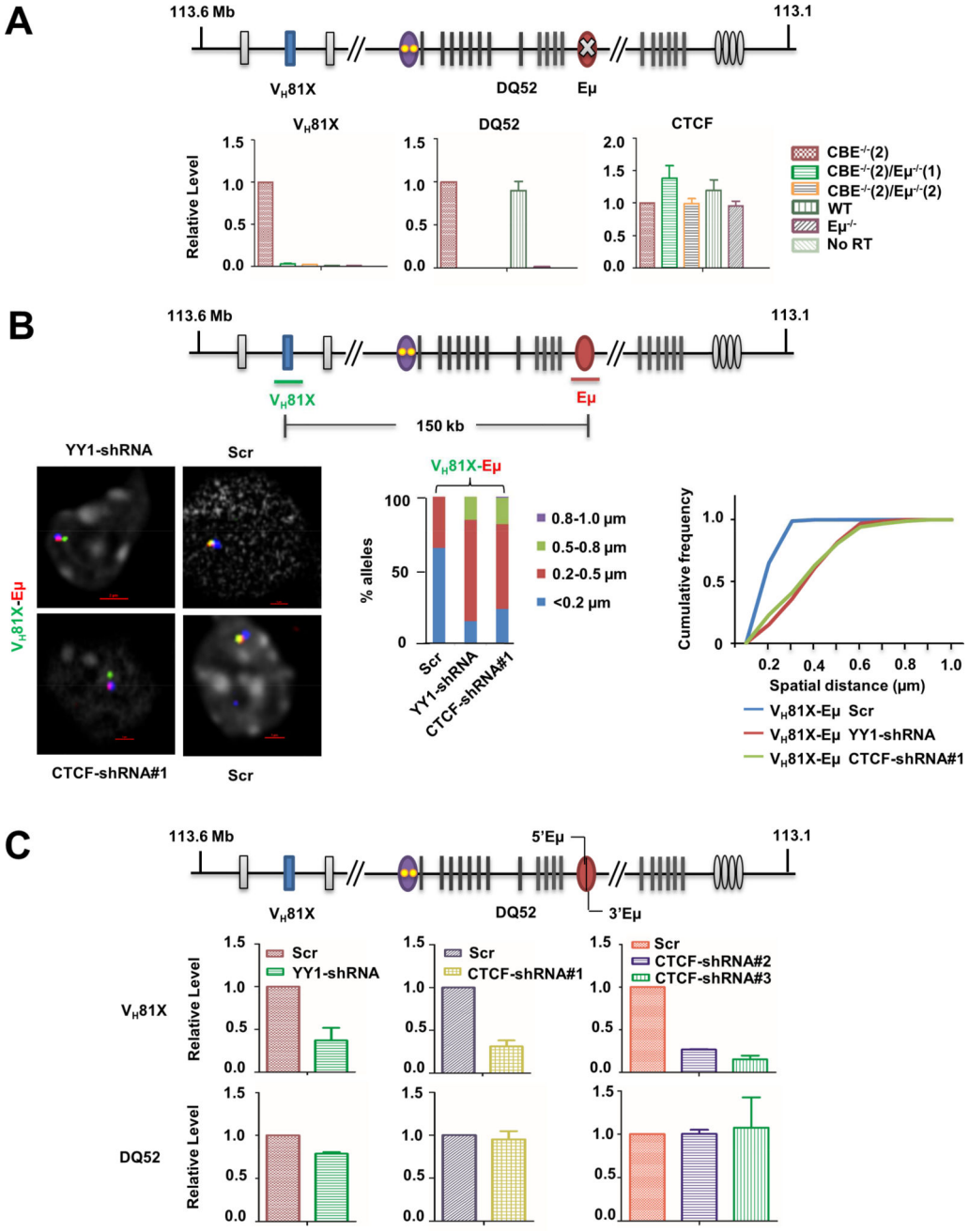


Figure 2. Features of the Eμ-V_H81X loop on IGCR1-mutated alleles

A. Eμ-dependent activation of V_H81X transcription on IGCR1-mutated alleles. $CBE^{-/-}$(2)/$Eμ^{-/-}$ (1) and $CBE^{-/-}$(2)/$Eμ^{-/-}$(2), are two cell clones derived from the $CBE^{-/-}$(2) pro-B cell line in which Eμ was deleted by CRISPR/Cas9 technology. Pro-B cells with wild-type and Eμ-deficient *IgH* alleles are indicated as WT and $Eμ^{-/-}$, respectively. Total RNA was assayed by quantitative RT-PCR using amplicons indicated on the top line. *IgH* locus transcript levels were normalized to *γ-actin* mRNA and are represented relative to levels in $CBE^{-/-}$(2) cells (Y axis). *Ctcf* mRNA levels served as an additional out-of-locus control. Data are shown as mean ± SEM of three independent RNA preparations and analyses.

B. CTCF- and YY1-dependence of Eμ-V_H81X loop. CTCF and YY1 expression was reduced in $CBE^{-/-}$(2) cells by lentiviral expression of shRNA directed against the respective mRNAs (Figure S4B). Prevalence of the Eμ-V_H81X loop was assessed by 3D-FISH using probes close to these sites. Cells expressing a scrambled shRNA (Scr) served as controls. Representative nuclei from YY1-shRNA, CTCF-shRNA#1 and control (Scr) cells are shown. Spatial distance between probes was measured after deconvolution of images from 100 nuclei. Bar graph shows the percentage of *IgH* alleles in which the distance between probes fell in the ranges shown in different colors. The cumulative frequency distribution of spatial distances is shown to the right. D-statistics and P-values for differences between KD and control cells were calculated as described in the Figure 1 legend (Table S1). Data from biological replicate experiments are shown in Figure S4C.

C. Transcriptional analyses in YY1- and CTCF-knockdown cells. Total RNA isolated from cells with indicated shRNA expression was assayed by quantitative RT-PCR using amplicons indicated on the top line. *IgH* locus transcripts (V_H81X and DQ52) were normalized to *γ-actin* mRNA and are represented relative to levels in control cells (Y axis). Additional *IgH* locus transcripts from Eμ (5'Eμ and 3'Eμ) levels are shown in Figure S4B. Data are shown as mean ± SEM of two independent knock-down experiments followed by RNA analyses.

See also Figure S4 and Table S1, 3.

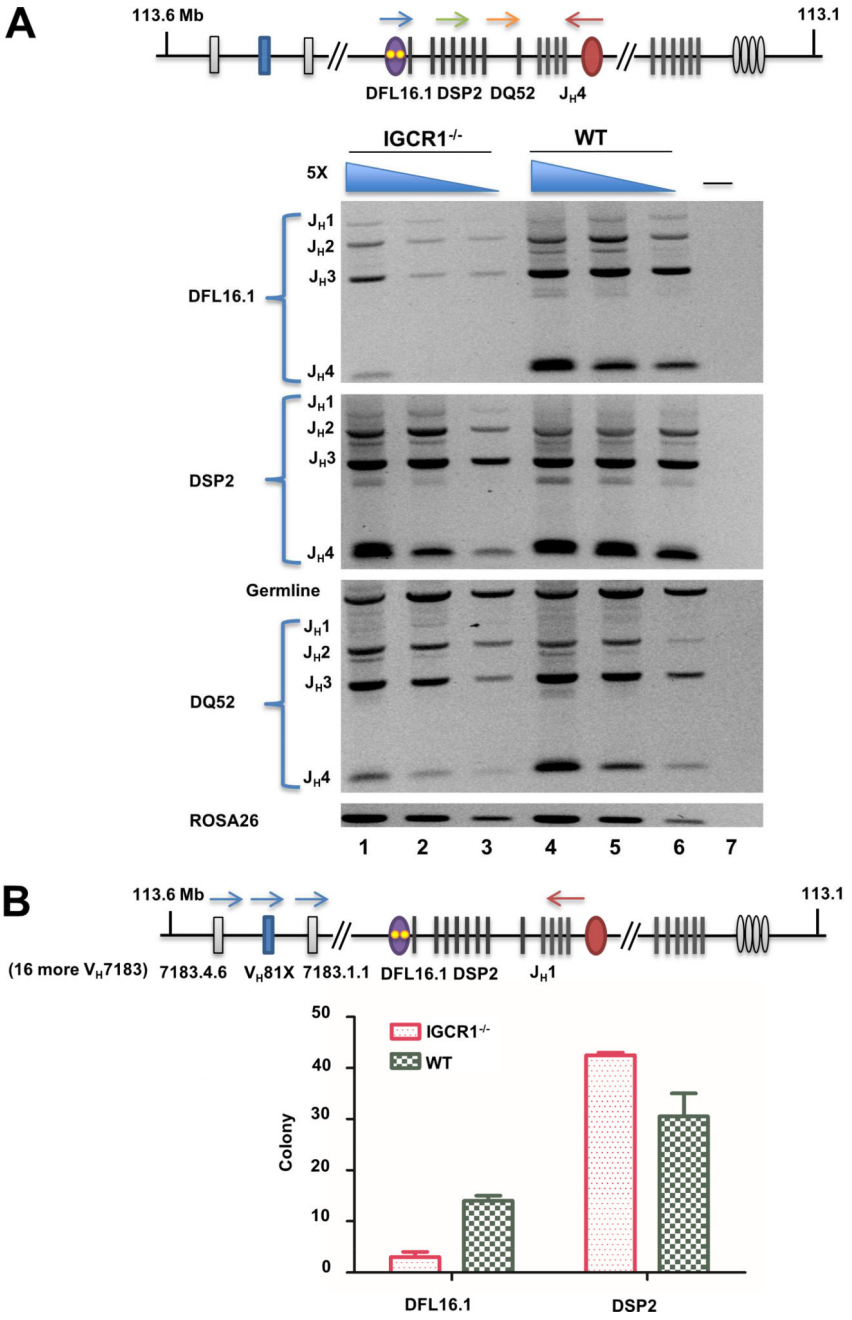


Figure 3. D_H gene segment utilization on IGCR1-deleted *IgH* alleles in bone marrow pro-B cells

A. Analysis of DJ_H junctions. Genomic DNA purified from pro-B cells ($B220^+IgM^-CD43^+$) from WT and IGCR1-deleted *IgH* alleles(Guo et al., 2011b) was used in amplification reactions with 5' primers specific for either DFL16.1 (blue arrow), 8 DSP2 (green arrow) or DQ52 gene segments (orange arrow) together with a 3' primer located beyond J_H4 (red arrow). Amplification products were analyzed by agarose gel electrophoresis. 5-fold increasing amounts of genomic DNA starting at 4ng (lanes 3, 6) are shown for each genotype. The non-rearranging ROSA26 locus served as a loading control. Rearrangement of each D_H gene segment to the four J_H gene segments is indicated to the left of the gel.

Lane 7 is an amplification reaction with no genomic DNA. Two biological replicate purifications of pro-B cells were performed. Data shown is a representative rearrangement analysis from one experiment. Averaged data from both experiments is provided in Figure S5A.

B. D_H utilization in VDJ_H junctions in pro-B cells from mice carrying WT and IGCR1-deleted *IgH* alleles. Pro-B cell genomic DNA was amplified using a 5' primer specific for 19 out of 49 rearrangeable V_H7183 gene segments (blue arrows and detailed in Figure S5C), including the 3'-most V_H7183 gene segment indicated in the schematic, and a 3' primer located just after J_H1 (red arrow). Amplification products were sub-cloned into pGEM®-T, and 55 clones from each of two pro-B cell preparations from WT and IGCR1-deleted mice were sequenced to estimate V_H and D_H usage (Table S2). DFL16.1 and DSP2 utilization in each genotype ranged between 39 and 50 subclones per experiment; D_H gene segment identity could not be established in the remaining clones due to end processing during VDJ_H recombination. Numbers of clones with DFL16.1 and DSP2 rearrangements from each genotype are shown in the graph. Data are shown as mean ± SEM of two experiments carried out with independently sorted pro-B cell preparations.

See also Figure S5 and Table S2, 3.

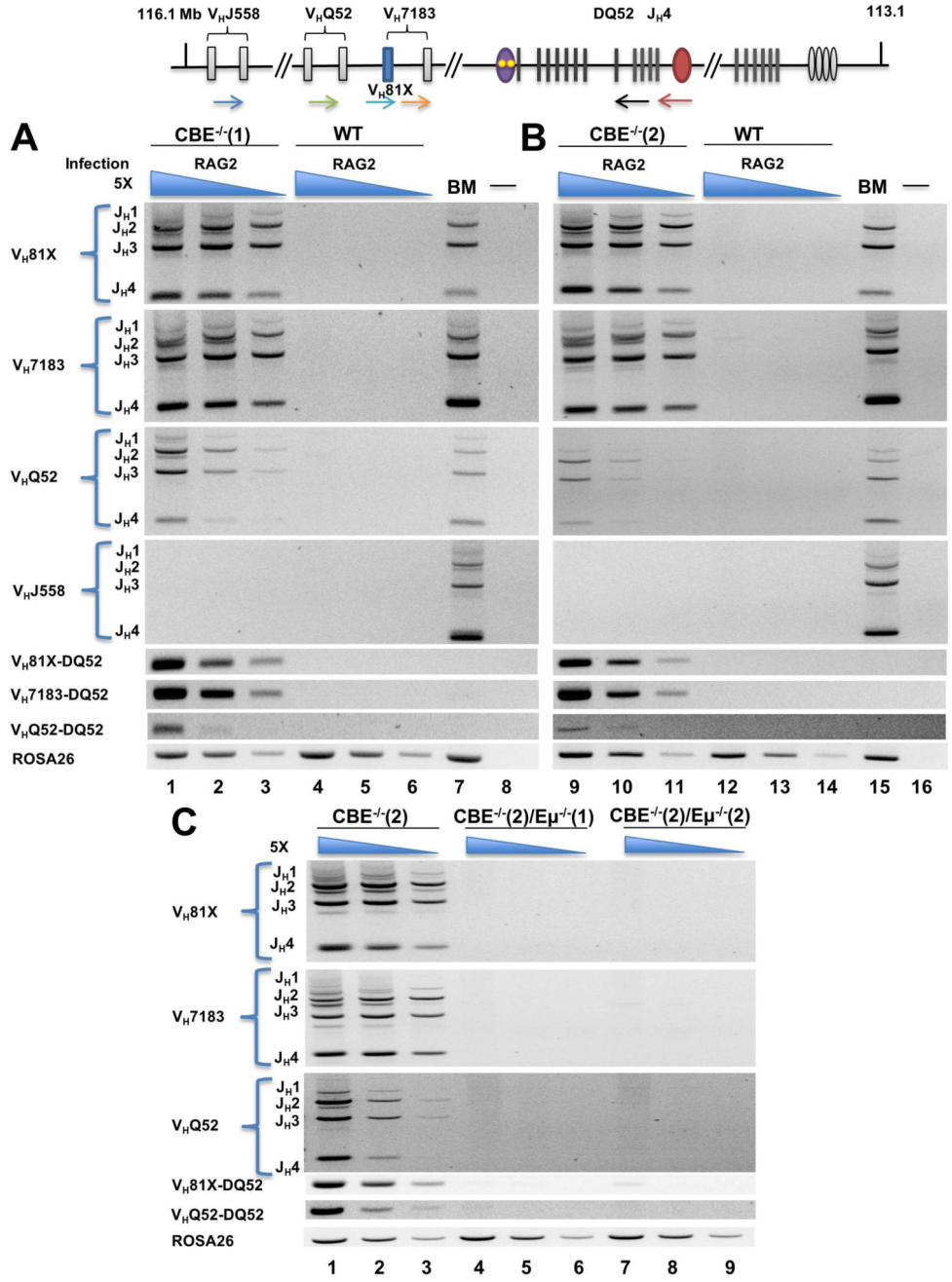


Figure 4. Eμ-dependent recombination efficiency of IGCR1-mutated *IgH* alleles

Top line shows a schematic of the *IgH* locus. 5' primers that score for V_HJ558 (blue), V_H7183 (orange), V_HQ52 (green) or specifically V_H81X (light blue) gene segments were used with a 3' primer located just after DQ52 (black) or J_H4 (red) to quantitate VD_H or VDJ_H recombination, respectively, on WT and IGCR1-mutated *IgH* alleles. Pro-B cell lines with wild-type (WT), IGCR1-mutated alleles ($CBE^{-/-}$(1) and (2)), Eμ and IGCR1 double-mutated alleles ($CBE^{-/-}$(2)/$Eμ^{-/-}$(1) and $CBE^{-/-}$(2)/$Eμ^{-/-}$(2)) and were infected with *Rag2*-expressing lentivirus. After 4–5 days (A and B) or 14 days (C) selection with puromycin, genomic DNA was prepared to assay rearrangements.

A, B. Representative rearrangement profiles for each IGCR1-mutated line assayed using primers indicated to the left of the gel. For the top 4 panels, the 3' primer was located beyond J_H4, thereby scoring for VDJ_H rearrangements to each J_H gene segment. Lanes 7 and 15 are positive controls using total bone marrow genomic DNA from WT mice; lanes 8 and 16 contain no genomic DNA. The next 3 panels use V_H primers as indicated, with a 3' primer located just after DQ52; these products score for V_H to DQ52 rearrangements. An amplicon from the ROSA26 locus was used as a DNA loading control. The three lanes associated with each cell type use 5-fold increasing genomic DNA concentration starting at 8ng (lanes 3, 6, 11, 14).

C. Representative rearrangement profiles of IGCR1- plus Eμ- deficient *IgH* alleles. Data shown is representative of 2 independent *Rag2* transduction experiments into each cell line. Analyses of D_H recombination after 4–5 days selection and analyses of V_H and D_H recombination after 14 days selection are shown in Figures S5B, S6A and S6C, respectively. See also Figure S6 and Table S3.

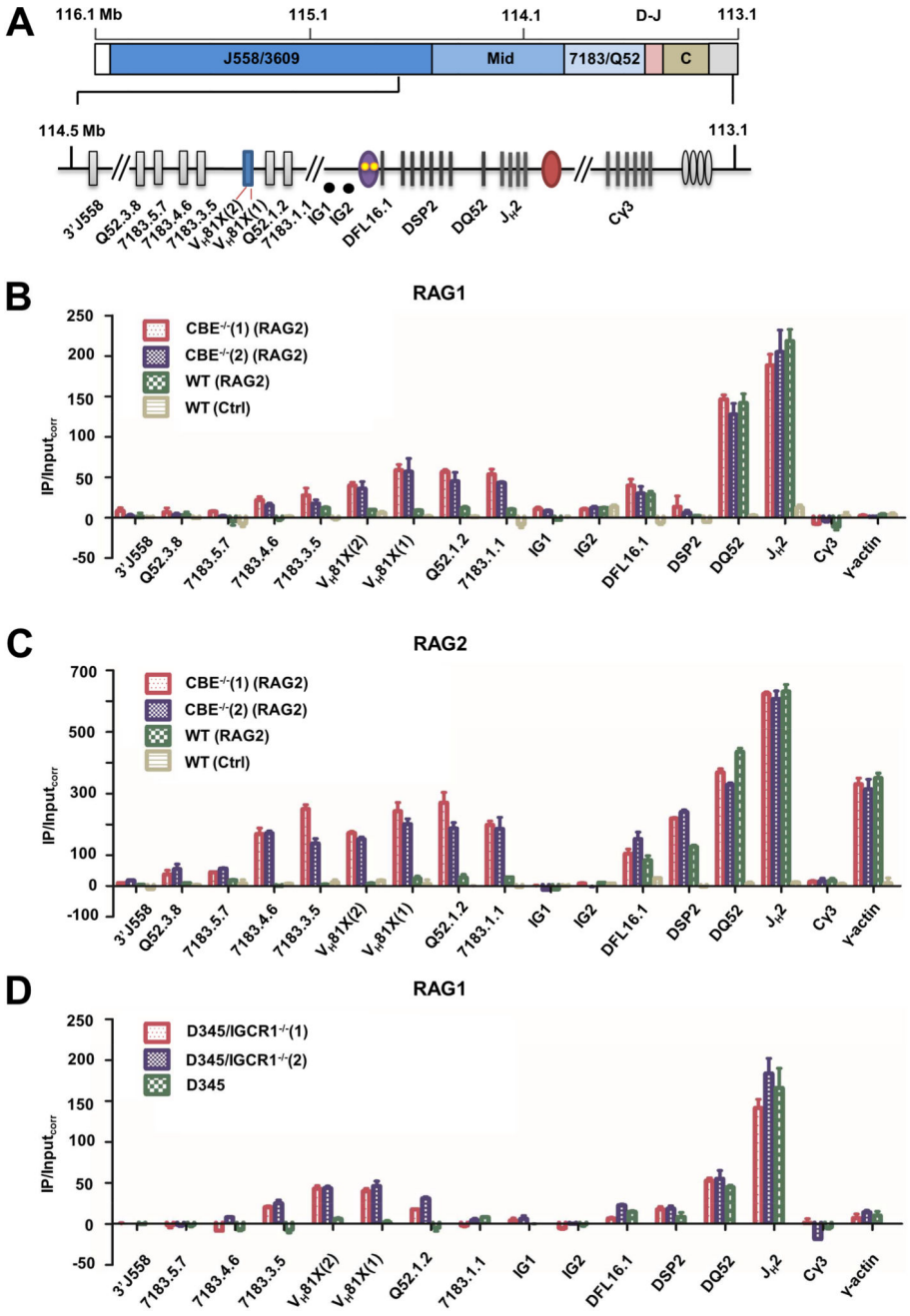


Figure 5. RAG1 and RAG2 recruitment to WT and IGCR1-mutated alleles

A. Scale schematic of the *IgH* locus indicating V_H gene families (J558/3609, mid, and 7183/Q52), and the 3' part of the locus containing D_H and J_H gene segments (D–J), and C_H exons (C_H). An expanded view of the portion of the locus assayed for RAG1/2 binding by ChIP. ChIP amplicon V_H81X (2) scores for unrearranged V_H81X only; intergenic amplicons IG1 and IG2 are located approximately 30 kb and 60 kb from V_H81X in the 90 kb region between DFL6.1 and V_H81X. DSP2 ChIP primers identify 8 DSP2 gene fragments in the unrearranged configuration.

B. RAG1 ChIP. Anti-RAG1 antibody was used to co-precipitate RAG1-associated genomic DNA as described by (Ji et al., 2010) from RAG2-deficient pro-B cell lines that contain IGCR1-mutated ($CBE^{-/-}$(1) and $CBE^{-/-}$(2)) or WT *IgH* alleles (red, purple and green bars, respectively). RAG2 was expressed in these lines by lentiviral transduction. The pro-B cell line with WT *IgH* alleles was transduced with a puromycin-only lentivirus serving as a control (yellow bars). Co-precipitated genomic DNA and input DNA were assayed by quantitative (real time) PCR and RAG1 enrichment calculated as described by Ji et al. (Y axis). Cγ3 and *γ-actin* amplicons served as negative controls for RAG1 binding. Two independent RAG2 transductions were carried out, data are shown as mean ± SEM of three independent ChIP experiments analyzed in duplicate for each amplicon.

C. RAG2 ChIP. Anti-RAG2 antibody was used to co-precipitate associated genomic DNA. Cell lines and amplicons are as described in part B. In this case only Cγ3 amplicon served as a negative control, since RAG2 is known to be recruited to the *γ-actin* promoter in the absence of RAG1. Two independent RAG2 transductions were carried out, data are shown as mean ± SEM of two independent ChIP experiments.

D. RAG1 ChIP in IGCR1-deleted D345 cells. CRISPR/Cas9 was used to delete IGCR1 in D345 pro-B cell line that expresses a catalytically inactive RAG1 and endogenous RAG2. Two cell clones, D345/$IGCR1^{-/-}$(1) and D345/$IGCR1^{-/-}$(2), were used for ChIP with anti-RAG1 antibodies. Co-precipitated DNA was analyzed as described in parts B/C with indicated primers. Cγ3 and *γ-actin* amplicons served as negative controls for RAG1 binding. Data are shown as mean ± SEM of two independent ChIP experiments from each cell line.

See also Figure S7 and Table S3.

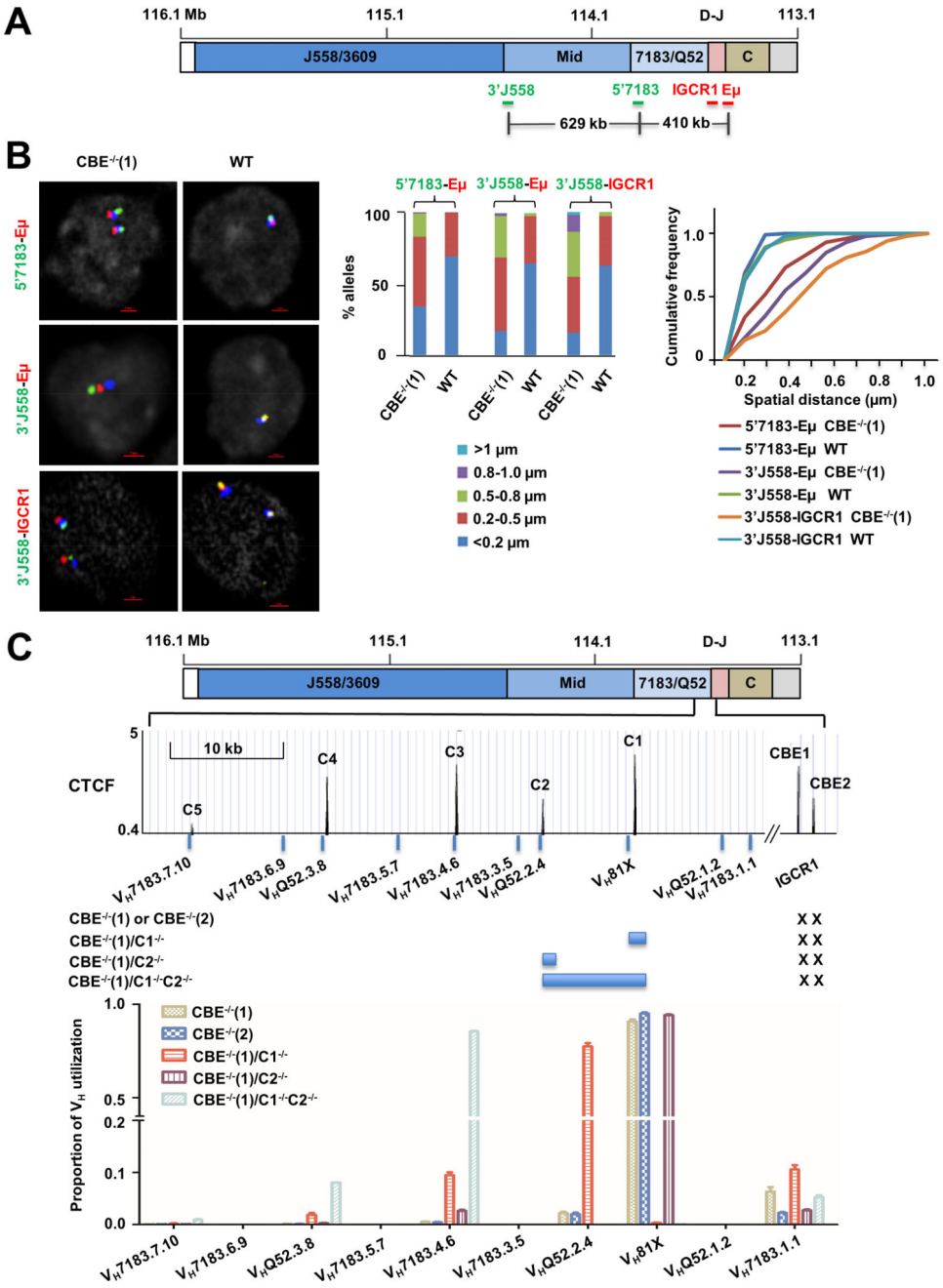


Figure 6. Re-directing VDJ$_H$ recombination on IGCR1-mutated *IgH* alleles

A. Scale schematic of the *IgH* locus indicating locations of FISH probes used to assay conformation of IGCR1-mutated alleles. Probes labeled 3'J558, 5'7183, IGCR1 and Eμ range in length from 4–10 kb, and were generated by amplification of corresponding regions (Table S3) from BACs or genomic DNA. Distances between selected probes are indicated based on mm10.

B. Analysis of V_H-DJ_H interactions by FISH. Probes located in the V_H part of the locus (5'7183, 3'J558) and in the D_H-J_H part of the locus (Eμ and IGCR1) were hybridized to pro-B cell lines containing WT or IGCR1-mutated (CBE$^{-/-}$(1)) *IgH* alleles. BAC RP23-201H14

(blue) was used to mark *IgH* alleles. Representative nuclei are shown, with probe combinations indicated on the left. Spatial distances between probes were measured after image deconvolution from 100 nuclei. Bar graphs show the percentage of *IgH* alleles in which inter-probe distance fell in the ranges shown by different colors. Cumulative frequency distribution of spatial distances for each color-coded probe combination is shown to the right of the bar graph. D-statistics and P-values for differences between WT and IGCR1-mutated alleles were calculated using the two-sample Kolmogorov-Smirnov test (Table S1). At least 2 independent FISH experiments were carried out with each probe combination (Figure S8B).

C. Sequential activation of CTCF-proximal V_H gene segments on IGCR1-mutated alleles. The 3' part of the V_H region is expanded below the schematic to show the location of several proximal V_H gene segments and CTCF binding sites. Sites C1–C5 represent the first to fifth CTCF binding sites beyond IGCR1 (labeled CBE1 and CBE2). Sequences indicated by blue bars around C1 and C2 were deleted using CRISPR/Cas9 in the context of the IGCR1 mutated cell line $CBE^{-/-}(1)$ to generate cell lines $CBE^{-/-}(1)/C1^{-/-}$, $CBE^{-/-}(1)/C2^{-/-}$, and $CBE^{-/-}(1)/C1^{-/-}C2^{-/-}$. All cell lines were infected with RAG2-expressing lentivirus to assay recombination. After 10 days selection with puromycin, genomic DNA was prepared to analyze V_H utilization by deep sequencing (Hu et al., 2016). Total reads in each cell line were aligned to 128 V_H genes in 129 strain mice (Retter et al., 2007). Sequences of the first 10 proximal V_H accounted for 99% of reads. Proportion of V_H utilization (Y axis) was calculated as a fraction of reads of a specific gene fragment for each cell line. Genomic DNA from 2 independent RAG2 infection experiments were combined for sequencing. Further details are provided in Table S4 and S5.

See also Figure S8–10 and Tables S1, 3–5.

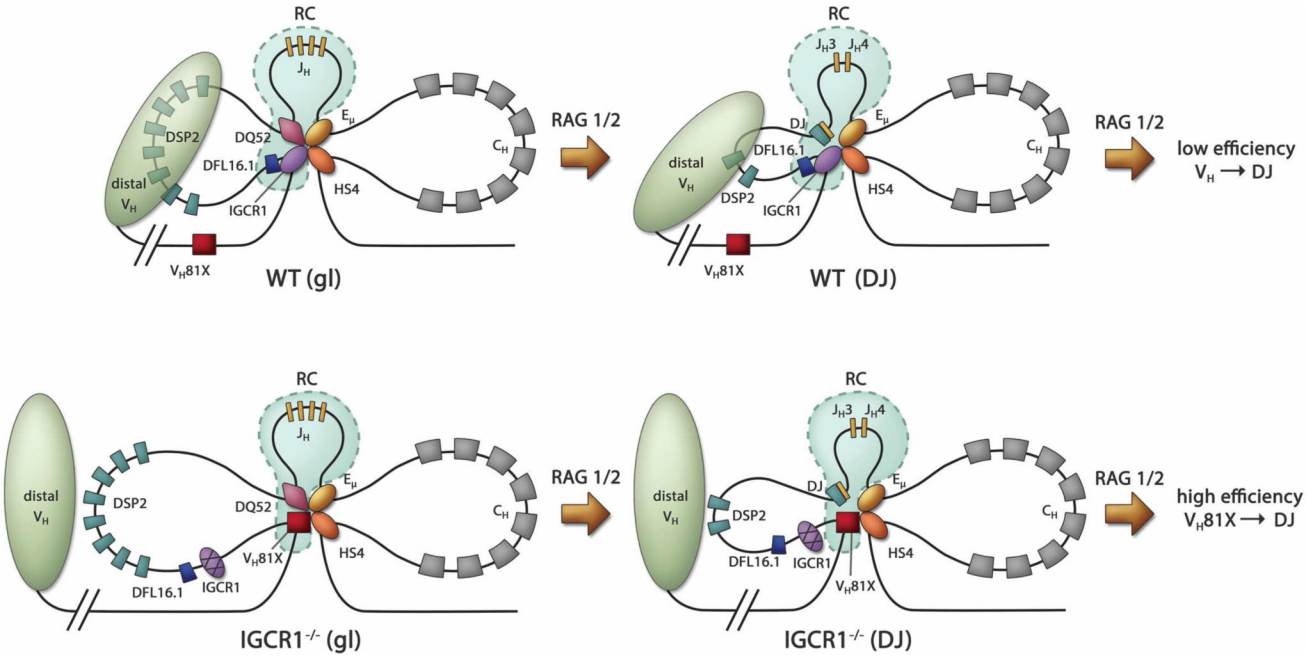


Figure 7. Model for Eμ- and IGCR1-dependent configuration of the 3' *IgH* domain

Top and bottom panels show proposed structures of WT and IGCR1-deficient *IgH* alleles, respectively, in unrearranged (left) and DJ_H recombined (right) configurations.

Left: Altered structure and recombination center formation on IGCR1-mutated *IgH* alleles prior to initiation of recombination (gl = germline alleles). Interactions between IGCR1 (blue oval), the DQ52 promoter (purple diamond), Eμ (yellow oval) and HS4 of the 3' regulatory region (orange oval) nucleate a 3-loop configuration (top). RAG1/2-rich recombination center (RC) is shown within dotted lines (light blue). The intermediate-sized loop contains all the D_H gene segments. The nearest functional V_H gene segment, V_H81X, is located 90 kb from IGCR1. The compacted domain of distal V_H gene segments is shown as a light green oval. Mutation of IGCR1 (bottom) leads to looping of Eμ to the next available CTCF binding site close to V_H81X. DFL16.1 lies near the middle of the new 160 kb loop, V_H81X is located spatially close to DQ52 and the RC, and the distal V_H domain moves away. High levels of RAG1/2 at V_H81X reflect an altered RC on IGCR1-mutated alleles.

Right: Proposed structure of WT (top) and IGCR1-deficient (bottom) DJ_H recombined *IgH* alleles as discussed in the text.