Published in final edited form as:

Immunity. 2013 August 22; 39(2): 229–244. doi:10.1016/j.immuni.2013.08.011.

Flexible Long-Range Loops in the V_H Gene Region of the *Igh* Locus Facilitate the Generation of a Diverse Antibody Repertoire

Jasna Medvedovic[1,8], **Anja Ebert**[1], **Hiromi Tagoh**[1], **Ido M. Tamir**[1,2], **Tanja A. Schwickert**[1], **Maria Novatchkova**[1,3], **Qiong Sun**[1], **Pim J. Huis in 't Veld**[1], **Chunguang Guo**[4], **Hye Suk Yoon**[4], **Yves Denizot**[5], **Sjoerd J.B. Holwerda**[6], **Wouter de Laat**[6], **Michel Cogné**[5], **Yang Shi**[7], **Frederick W. Alt**[4], and **Meinrad Busslinger**[1,*]

[1]Research Institute of Molecular Pathology, Vienna Biocenter, Dr. Bohr-Gasse 7, 1030 Vienna, Austria [2]Campus Science Support Facilities, Dr. Bohr-Gasse 5, 1030 Vienna, Austria [3]Institute of Molecular Biotechnology of the Austrian Academy of Sciences, Dr. Bohr-Gasse 3, 1030 Vienna, Austria [4]Howard Hughes Medical Institute, Children's Hospital, Immune Disease Institute, Harvard Medical School, 300 Longwood Avenue, Boston, MA 02115, USA [5]UMR CNRS 7276, Centre Nationale de la Recherche Scientifique, Université de Limoges, 87025 Limoges Cedex, France [6]Hubrecht Institute-KNAW and University Medical Center Utrecht, Uppsalalaan 8, Utrecht 3584 CT, the Netherlands [7]Department of Cell Biology, Harvard Medical School, Division of Newborn Medicine, Children's Hospital, Boston, MA 02115, USA

SUMMARY

The immunoglobulin heavy-chain (*Igh*) locus undergoes large-scale contraction in pro-B cells, which facilitates V_H-DJ_H recombination by juxtaposing distal V_H genes next to the DJ_H-rearranged gene segment in the 3′ proximal *Igh* domain. By using high-resolution mapping of long-range interactions, we demonstrate that local interaction domains established the three-dimensional structure of the extended *Igh* locus in lymphoid progenitors. In pro- B cells, these local domains engaged in long-range interactions across the *Igh* locus, which depend on the regulators Pax5, YY1, and CTCF. The large V_H gene cluster underwent flexible long-range interactions with the more rigidly structured proximal domain, which probably ensures similar participation of all V_H genes in V_H-DJ_H recombination to generate a diverse antibody repertoire. These long-range interactions appear to be an intrinsic feature of the V_H gene cluster, because they are still generated upon mutation of the Eμ enhancer, IGCR1 insulator, or 3′ regulatory region in the proximal *Igh* domain.

*Correspondence: busslinger@imp.ac.at.
[8]Present address: Lymphoid Tissue Development Unit, Institut Pasteur, 75015 Paris, France

ACCESSION NUMBERS

The 4C-seq and ChIP-seq data are available in the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/gds) under the accession number GSE43008.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.immuni.2013.08.011.

INTRODUCTION

B lymphocytes generate humoral immunity to foreign pathogens by creating a diverse antigen receptor repertoire through V(D)J recombination, which assembles the variable regions of immunoglobulin (Ig) genes from variable (V), diversity (D), and joining (J) gene segments during B cell development. The recombination of *Ig* genes is tightly controlled within the B lymphoid lineage: the Ig heavy-chain (*Igh*) locus undergoes rearrangements in early B cell development prior to Ig light-chain genes in pre-B cells. D_H-J_H rearrangements at the *Igh* locus are initiated in lymphoid progenitors followed by V_H-DJ_H recombination in pro-B cells. The temporal order of V(D)J recombination is largely determined by the accessibility of the different Ig gene segments to the V(D)J recombinase, which is controlled by multiple epigenetic mechanisms (Jhunjhunwala et al., 2009; Perlot and Alt, 2008).

The *Igh* locus is composed of the 3′ proximal region of 266 kb length consisting of 16 D_H, 4 J_H, and 8 C_H gene segments and of the distal V_H gene cluster extending over a 2.44 Mb region, which contains 195 V_H genes with the largest V_H gene family consisting of 89 V_HJ558 genes (Johnston et al., 2006). V_H-DJ_H recombination at the *Igh* locus is regulated at several levels including the relocation of *Igh* alleles from peripheral to central nuclear positions (Fuxa et al., 2004; Kosak et al., 2002) and antisense transcription in the V_HJ558 gene region (Bolland et al., 2004). Both *Igh* alleles also undergo homologous pairing in pro-B cells, which ensures that V_H-DJ_H recombination simultaneously takes place on only one of the two *Igh* alleles (Hewitt et al., 2009).

The *Igh* locus furthermore contracts by looping in pro-B cells, which juxtaposes distal V_H genes next to proximal D_H segments to facilitate V_H-DJ_H rearrangements (Fuxa et al., 2004; Jhunjhunwala et al., 2008; Kosak et al., 2002; Roldá n et al., 2005; Sayegh et al., 2005). Moreover, the *Igh* locus undergoes decontraction at the next developmental stage, which separates V_H genes from the proximal *Igh* domain, thereby preventing V_H-DJ_H rearrangement of the second, DJ_H-rearranged *Igh* allele in pre-B cells (Roldá n et al., 2005). The pro-B cell-specific contraction of the *Igh* locus depends on the B cell commitment factor Pax5 (Fuxa et al., 2004) and the ubiquitous transcriptional regulator YY1 (Liu et al., 2007). Long-range chromatin looping at complex loci is known to depend on the CCCTC-binding factor (CTCF) and its associated cohesin complex (Hadjur et al., 2009; Nativio et al., 2009; Splinter et al., 2006). Notably, multiple CTCF- and cohesin-binding sites are colocalized throughout the V_H gene cluster (Degner et al., 2011; Ebert et al., 2011), and shRNA knockdown experiments implicated CTCF in the regulation of *Igh* locus contraction in pro-B cells (Degner et al., 2011).

Several *cis*-acting elements control the activity of the *Igh* locus. The intronic Eμ enhancer is essential for *Igh* recombination by regulating germline transcription and chromatin accessibility in the D_H-J_H region (Afshar et al., 2006; Chakraborty et al., 2009; Perlot et al., 2005). The Eμ enhancer was also implicated in *Igh* locus contraction by mediating loop formation with two V_H gene regions (Guo et al., 2011a). The 3′ regulatory region (3′RR) downstream of the C_H region consists of four DNase I hypersensitive sites (HS1–HS4), which function as potent enhancers in late B cell development (Vincent-Fabert et al., 2010). The downstream 3′CBE region also consists of four DNase I hypersensitive sites (HS5–HS7

and site “38’ [here referred to as HS8]) and may constitute the 3′ boundary of the *Igh* locus (Garrett et al., 2005), because it contains nine CTCF-binding elements (CBEs) that colocalize with cohesin-binding sites (Degner et al., 2011; Ebert et al., 2011). The intergenic control region 1 (IGCR1) with its two CBEs is located 2.1 kb upstream of the D_HFL16.1 segment in the 100 kb region separating the D_H and V_H gene regions (Guo et al., 2011b). Specific mutation of these two CTCF-binding sites in IGCR1/CBE mutant mice revealed that they function as insulator elements to regulate ordered and lineage-specific V(D)J recombination at the *Igh* locus (Guo et al., 2011b). The large V_H gene cluster contains 14 Pax5-activated intergenic repeat (PAIR) elements, which are interspersed together with the V_H3609 genes in the distal V_HJ558 gene region (Ebert et al., 2011). The PAIR elements, which bind Pax5, E2A, CTCF, and cohesin, give rise to long noncoding antisense transcripts only in pro-B cells, suggesting that they regulate distal V_H-DJ_H recombination possibly by controlling *Igh* locus contraction (Ebert et al., 2011).

The 3D architecture of the *Igh* locus was so far studied at low spatial resolution in single cells by DNA fluorescent in situ hybridization (DNA-FISH) (Fuxa et al., 2004; Jhunjhunwala et al., 2008; Roldá n et al., 2005). Here, we have used 4C sequencing (4Cseq), which provides high-resolution analysis of chromatin loops at the cell population level and is based on the chromosome conformation capture-on-chip (4C) technique adapted for deep sequencing (van de Werken et al., 2012). 4C sequencing revealed that the *Igh* locus consists of local interaction domains in lymphoid progenitors. In committed pro-B cells, the local domains in the V_H gene region participate in long-range interactions across the entire *Igh* locus by generating a continuum of flexible loops that probably contribute to the generation of a diverse immunoglobulin repertoire by providing a similar opportunity for each V_H gene to contact the proximal DJ_H-rearranged gene segment prior to V_H-DJ_H recombination.

RESULTS

High-Resolution Mapping of Long-Range Interactions in the *Igh* Locus by 4C Sequencing

To determine the 3D architecture of the *Igh* locus, we used short-term cultured pro-B cells on a recombination-deficient *Rag2* mutant background for 4C sequencing with 16 different viewpoints across the *Igh* locus, which are referred to by the closest annotated feature and the combination of primary (HindIII [H], EcoRI [E], or BglII [B]) and secondary (DpnII [D] or NlaIII [N]) restriction enzymes used for 4C template preparation (see Supplemental Experimental Procedures for a general description of the 4C-seq concept). We performed these 4Cseq experiments in the presence or absence of *Igh* locus contraction by analyzing $Rag2^{-/-}$ and $Pax5^{-/-}Rag2^{-/-}$ pro-B cells, respectively (Fuxa et al., 2004). Figures 1A and 1B show a typical 4C-seq result obtained with the HS3B-HD viewpoint in the *Igh* 3′ regulatory region (3′RR). Most 4C-seq reads (70%) mapped to the contracted *Igh* locus in Pax5-expressing $Rag2^{-/-}$ pro-B cells, whereas fewer sequence reads (49%) with a narrower distribution were localized to the decontracted *Igh* locus in $Pax5^{-/-}Rag2^{-/-}$ pro-B cells. 4C-seq experiments with other viewpoints revealed the same trend (Figure S1A available online). We therefore conclude that long-range interactions are largely confined to the *Igh* locus in pro-B cells undergoing locus contraction, whereas *Igh* sequences more readily interact with chromosomal sites throughout the genome in the absence of locus contraction.

Within the *Igh* locus, sequences at the HS3B-HD viewpoint efficiently associated only with elements of the *Igh* 3′ region in the absence of *Igh* locus contraction in cultured *Pax5*$^{-/-}$ *Rag2*$^{-/-}$ pro-B cells and ex vivo isolated thymocytes (Figure 1C). In contrast, interactions with multiple sequences throughout the entire *Igh* locus were seen in both cultured and ex vivo isolated *Rag2*$^{-/-}$ pro-B cells, which validates the use of cultured pro-B cells for 4C-seq analysis (Figure 1C). Because the observed interaction patterns were highly reproducible in three independent 4C-seq experiments (Figure S1B), we subsequently show the average pattern of replica experiments, if available. Of note, the long-range interactions observed with the HS3B-HD viewpoint abruptly terminated at the 3′CBE region and the most 5′ V_H gene (J558.89pg.195) of the *Igh* locus (Figure 1C).

The *Igh* locus contains many interspersed repeat sequences (Johnston et al., 2006), which prevent mapping of the corresponding 4C-seq reads to the *Igh* locus. Consequently, the end sequences of HindIII, BglII, and EcoRI fragments differ in their mappability along the *Igh* locus (Figures 1C and 1D, gray bars). We therefore used (in addition to HindIII), BglII and EcoRI for the preparation of 4C templates. The 4C-seq data obtained with the overlapping viewpoints HS8-BN and HS8-ED in the 3′CBE region demonstrated that their interactions in pro-B cells were also confined to the 3′ proximal region in the absence of Pax5 and extended along the entire *Igh* locus in the presence of Pax5 (Figures 1D and 1E), similar to the results obtained with the HS3B-HD viewpoint (Figure 1C). Together, these 4Cseq data demonstrate that the 3′RR and 3′CBE regions form multiple loops across the entire *Igh* locus in committed pro-B cells.

Loop Formation in the 3′ Proximal *Igh* Domain

The proximal *Igh* region is known to form loops between the regulatory IGCR1, Eμ, 3′RR, and 3′CBE regions in pro-B cells (Degner et al., 2011; Guo et al., 2011a, 2011b). Our 4C-seq analyses with IGCR1, Eμ, HS3B (3′RR), and HS8 (3′CBE) viewpoints revealed that strong interactions in the proximal *Igh* region were restricted to a ~0.5 Mb long region in the absence of locus contraction in *Pax5*$^{-/-}$*Rag2*$^{-/-}$ pro-B cells (Figure 2) and thymocytes (Figure S2A). Notably, the interaction patterns observed for each viewpoint were similar in *Rag2*$^{-/-}$ pro-B cells compared to *Pax5*$^{-/-}$*Rag2*$^{-/-}$ pro-B cells (Figure 2) and thymocytes (Figure S2A), indicating that locus contraction does not influence loop formation in the proximal *Igh* domain.

Our 4C-seq analyses confirmed the existence of the previously identified chromatin loops between the IGCR1, Eμ, 3′RR (HS3B), and 3′CBE (HS8) regions (Figure 2, highlighted in green). Our 4C-seq analyses detected two new contact sites in the *Igh* constant gene region (Figure 2, highlighted in orange). A region close to the Cγ3 gene segment was contacted by the viewpoints HS3B-HD and Eμ-HD (Figure 2A). The viewpoints IGCR1-ED, Eμ-HD, HS3B-HD, and HS8-ED all interacted with a region located between the Cγ1 and Cγ2b gene segments (Figures 2A and 2C), which contained two DNase I hypersensitive (DHS) regions (referred to as Cγ1-2b DHS sites 1 and 2) in *Rag2*$^{-/-}$ pro-B cells (Figure 2D). Pax5 not only bound to both Cγ1-2b DHS sites together with other transcription factors but also was required for the formation of these DHS sites and their active chromatin state (H3K4me2, H3K9ac; Figure 2D).

Notably, four interaction sites were identified upstream of the IGCR1 region in the absence of *Igh* locus contraction. The viewpoints HS3B-HD, Eμ-HD, and IGCR1-ED strongly associated with sequences close to the D_H5-1 pseudogene, which is located in the 100 kb region separating the D_H and V_H gene regions (Figures 2A and 2C). The proximal V_H7183.2.3 gene region (also known as V_H81X) was efficiently contacted by the viewpoints HS3B-HD and Eμ-HD, the upstream V_HQ52.3.8 gene sequences by the four viewpoints IGCR1-ED, IGCR1-BN, HS8-ED, and HS8-BN, and the V_H7183.9.15 gene region by the viewpoints HS3B-HD and HS8-BN (Figures 2A–2C). Not all interaction sites were equally well detected by the two overlapping viewpoints in the IGCR1 (ED and BN) or HS8 (ED and BN) region. The observed differences may be explained by the different mappabilities of the EcoRI and BglII end sequences (Figures 2B and 2C), the presence of unique features in the nonoverlapping part of these viewpoints (Figure S3), or possible PCR biases. Despite these potential limitations, our systematic 4C-seq analysis detected new chromatin loops within the proximal *Igh* domain in pro-B cells. Hence, more chromatin loops than previously anticipated contribute to the 3D architecture of the proximal *Igh* region in the absence of locus contraction.

Folding Pattern of the Distal V_H Gene Region

We next investigated the 3D conformation at the *Igh* 5′ end by using viewpoints corresponding to the last two V_HJ558 and upstream *Zfp386* genes (Figure S3). The viewpoint Zfp386-HD almost exclusively interacted with sequences upstream of the last V_HJ558.89pg.195 gene in both *Pax5*$^{-/-}$*Rag2*$^{-/-}$ and *Rag2*$^{-/-}$ pro-B cells (Figure 3A). In contrast, the J558.89-HD viewpoint efficiently interacted with upstream and downstream sequences in both pro-B cell types (Figure 3A) similar to the situation observed with the two HS8 viewpoints at the *Igh* 3′ end (Figures 2B and 2C). The J558.88-ED viewpoint predominantly contacted sequences in the distal V_H gene cluster, although it is separated by only 6.3 kb from the upstream J558.89-HD viewpoint. Hence, these data indicate that the 5′ and 3′ boundaries of the *Igh* locus are located at or near the V_HJ558.89pg.195 gene and HS8 site, respectively.

Three viewpoints in the PAIR region (PAIR4,5,8-HD) revealed an interaction pattern in the distal V_H gene region that was unique in two aspects. First, the interactions detected with all three viewpoints extended from the *Igh* 5′ boundary over a 1.3 Mb region spanning almost the entire V_HJ558 gene cluster in *Pax5*$^{-/-}$ *Rag2*$^{-/-}$ and *Rag2*$^{-/-}$ pro-B cells (Figure 3B) and is thus 2.5 times larger than the 0.5 Mb domain in the 3′ proximal *Igh* region in the absence of locus contraction. Second, the 4C-seq profiles of both pro-B cell types were similar for not just one, but for all three PAIR viewpoints (Figure 3B) as well as for the J558.70-HD viewpoint (Figure S2B). In contrast, the J606.5-HD viewpoint near the 3′ end of the distal V_H gene cluster was part of a different local interaction region that minimally overlapped with the PAIR element-containing interaction domain in *Pax5*$^{-/-}$*Rag2*$^{-/-}$ pro-B cells (Figure S2B). These data therefore identified a large interaction domain, which spans most of the distal V_HJ558 gene cluster in the absence of locus contraction.

Flexible Long-Range Interactions across the V_H Gene Cluster Mediate Locus Contraction

As expected, the Eμ-HD and HS3B-HD viewpoints of the proximal *Igh* domain specifically interacted with the distal V_H gene region only upon locus contraction in *Rag2*$^{-/-}$ pro-B cells, but not in *Pax5*$^{-/-}$*Rag2*$^{-/-}$ pro-B cells (Figure 4A). Notably, the long-distance interaction patterns of these proximal viewpoints were highly similar not only between each other (Figure 4A) but also in comparison with the local interaction patterns of the three PAIR (4,5,8-HD) viewpoints (Figure 3B). Likewise, the long-distance interaction profiles of *Rag2*$^{-/-}$ pro-B cells were also comparable between the proximal HS8-BN and IGCR1-BN or HS8-ED and IGCR1-ED viewpoints (Figures 4B and 4C). However, the 4C-seq profiles of *Rag2*$^{-/-}$ pro-B cells differed in the distal V_H gene region between the three pairs of proximal region viewpoints similar to the observed changes in the mappability of the end sequences between the HindIII, BglII, and EcoRI sites that were used for 4C template preparation (Figures 4A–4C). One possible explanation for the above results could be that the observed peaks of the 4C-seq profiles identify interacting sequences at the base of chromatin loops. Alternatively, the 4C-seq patterns may result from a continuum of flexible long-range loops in the distal V_H gene region, particularly in view of the fact that each 4C-seq experiment was performed with approximately 1 million pro-B cells, thus revealing an average of all interactions in the entire cell population (Figure S4A).

To test these hypotheses, we generated two random 4C-seq libraries with purified DNA of nonoverlapping C57BL/6 BAC clones that covered the entire mouse *Igh* locus (Figures S4B–S4D). To this end, HindIII-digested BAC DNA was ligated at high concentration to favor random ligation, and the randomly ligated HindIII fragments were subsequently processed for deep sequencing in the same way as the 4C samples of pro-B cells (see Supplemental Experimental Procedures). In addition to the random assortment of HindIII sites, the sequencing of random BAC libraries also controlled for two potential artifacts, because it was subjected to the same mappability issues and potential PCR biases as the 4C-seq analysis of pro-B cells. As shown in Figure 4D, the sequence pattern of the BAC libraries was similar to the *Rag2*$^{-/-}$ pro-B cell interaction profile in the distal V_H gene region. We conclude therefore that the V_H gene region is characterized by a continuum of flexible long-range interactions.

Locus Contraction Promotes Interaction of Distal V_H Genes with Proximal CBE Regions

As shown by previous DNA-FISH analyses, the distal V_H gene region interacts with the proximal *Igh* domain only upon locus contraction in *Rag2*$^{-/-}$ pro-B cells. Here we revealed the details of these interactions by 4C-seq analysis by using distal J558 (89,70-HD) and PAIR (4,5,8-HD) viewpoints (Figures 5A, 5B, and S2B). The long-range interactions of these distal viewpoints with the proximal *Igh* domain were significantly different (Figure 5B) from the local interaction patterns observed with proximal viewpoints (Figure 2). In particular, the J558.89-HD and three PAIR (4,5,8-HD) viewpoints showed only minimal interactions with the Eμ enhancer and elements in the C_H region (Figure 5B), which were, however, strongly contacted by proximal viewpoints. Instead, the distal viewpoints efficiently interacted with the IGCR1/D_H and 3′CBE (HS7 and HS8) regions (Figure 5B), which are the only elements containing CTCF-binding sites in the proximal *Igh* domain (Degner et al., 2011; Ebert et al., 2011). These data suggest that locus contraction brings the

5′ and 3′ *Igh* boundaries into close proximity and may mediate long-range interactions between distal and proximal *Igh* sequences through CTCF-binding sites in the proximal *Igh* domain. Moreover, analysis of the *Igh* BAC libraries with the J558.89-HD viewpoint indicated that the random sequencing pattern of the BACs and the 4C-seq interaction profile of *Rag2*$^{-/-}$ pro-B cells were largely similar in the proximal and middle V_H gene regions but significantly differed in the *Igh* 3′ proximal domain (Figures 5C and S4E). These results therefore indicate that the proximal *Igh* domain assumes a more rigid 3D architecture compared to the flexible loops observed in the V_H gene region.

Normal Long-Range Interactions in the Absence of Individual 3′ Regulatory Elements

To investigate the influence of regulatory elements in the 3′ proximal domain on long-range *Igh* interactions, we examined the dual CBE mutation of the IGCR1 element (Guo et al., 2011b), the core Eμ enhancer deletion (Perlot et al., 2005), a deletion of the entire 3′RR (Vincent-Fabert et al., 2010), and the J_HT mutation (eliminating the D_HQ52, J_H and Eμ elements; Gu et al., 1993) in 129Sv pro-B cells (for further explanation see Supplemental Experimental Procedures). As shown in Figures 6A and S5A, 4C-seq analyses with proximal (HS8-ED) and distal (J558.89-HD) viewpoints revealed similar long-range interaction patterns across the entire V_H gene region in homozygous J_HT pro-B cells as well as in *Rag2*$^{-/-}$ pro-B cells containing one of the 3′RR, Eμ, or IGCR1/CBE mutations compared to control *Rag2*$^{-/-}$ pro-B cells. We conclude therefore that individual mutation of either the IGCR1, Eμ, or 3′RR region had no effect on longrange interactions across the *Igh* locus in *Rag2*$^{-/-}$ pro-B cells.

To confirm these findings with an independent method, we analyzed the same double mutant pro-B cells by 3D DNA-FISH with BAC probes (Figure 6A) from the 5′ and 3′ end of the V_H gene cluster. As shown in Figure 6B, the distribution and average value of the measured distances between the two DNA signals were similar in homozygous J_HT pro-B cells and *Rag2*$^{-/-}$ pro-B cells containing one of the 3′RR, Eμ, or IGCR1/CBE mutations compared to control *Rag2*$^{-/-}$ pro-B cells. In contrast, the distance between the probe signals was much greater for the decontracted *Igh* locus in *Pax5*$^{-/-}$*Rag2*$^{-/-}$ pro-B cells, as published (Fuxa et al., 2004). In summary, the results of both 4C-seq and 3D DNA-FISH analyses demonstrate that the long-range interactions mediating *Igh* locus contraction in pro-B cells do not depend on functional IGCR1, Eμ, or 3′RR elements.

CTCF-Binding Sites in the IGCR1 Region Contribute to Long-Range Looping

Although the IGCR1/CBE mutation had no global effect on long-distance interactions in the *Igh* locus, we next investigated whether the two CTCF-binding sites (CBEs) in the IGCR1 region contribute to the formation of long-range loops with distal *Igh* sequences. To this end, we performed 3C-qPCR experiments with mutant and control *Rag2*$^{-/-}$ pro-B cells by using primers located in HindIII fragments containing $V_HJ558.89$, PAIR8, PAIR4, and IGCR1 sequences (Figure 6C). Whereas the relative crosslinking frequency was high for the interaction between all three distal *Igh* elements and the IGCR1 region in *Rag2*$^{-/-}$ Eμ mutant and *Rag2*$^{-/-}$ 3′RR mutant pro-B cells, it was low in the locus contraction-deficient *Pax5*$^{-/-}$*Rag2*$^{-/-}$ pro-B cells and was reduced by about 50% in *Rag2*$^{-/-}$ IGCR1/CBE mutant pro-B cells compared to *Rag2*$^{-/-}$ pro-B cells (Figure 6C). We conclude therefore

that functional CBEs in the IGCR1 element are required for efficient loop formation with sequences in the distal V_H gene cluster.

Because CTCF and Pax5 are required for long-range interactions and bind to multiple sites in the *Igh* locus (Ebert et al., 2011), we hypothesized that both transcription factors may mediate looping by binding to each other. To test this idea, we took advantage of *Pax5*$^{Bio/Bio}$ pro-B cells expressing a Pax5 protein with a C-terminal biotin acceptor sequence that is efficiently biotinylated in vivo by coexpression of the *E. coli* biotin ligase BirA (McManus et al., 2011). Streptavidin pull-down experiments with nuclear extracts of *Pax5*$^{Bio/Bio}$ pro-B cells or BirA-expressing *Pax5*$^{+/+}$ pro-B cells revealed that CTCF specifically coprecipitated with the biotinylated Pax5-Bio protein in contrast to the control BirA extract (Figure 6D). In the reverse order, Pax5 was coimmunoprecipitated with anti-CTCF but not with unspecific IgG antibodies from nuclear extracts of *Rag2*$^{-/-}$ pro-B cells even after nuclease-mediated DNA digestion (Figure 6E). We next performed in vitro binding and GST pull-down assays with affinity-purified human CTCF and GST-Pax5 fusion proteins, which were expressed in baculovirus-infected Sf9 insect cells (Figure S5B). These in vitro binding experiments revealed that CTCF could directly interact with full-length Pax5 (G-P5-A) and a Pax5 mutant lacking the paired domain (G-P5-B), but not with polypeptides containing only central (G-P5-C) or C-terminal (G-P5-D) Pax5 sequences (Figure 6F). Similar results were obtained with GST pull-down assays with bacterially expressed GST-Pax5 fusion proteins (Figure S5C). These in vitro binding data are consistent with the possibility that Pax5 and CTCF may directly interact with each other in pro-B cells. As shown by ChIP-seq analysis of *Pax5*$^{-/-}$*Rag2*$^{-/-}$ and *Rag2*$^{-/-}$ pro-B cells (Figure S5D), the CTCF-binding patterns at the *Igh* locus were highly similar in the absence and presence of Pax5. Hence, CTCF binds to its sites in the *Igh* locus independently of its interaction with Pax5.

YY1 Controls PAIR Function and Long-Range Interactions across the *Igh* Locus

Given the involvement of the ubiquitous transcription factor YY1 in *Igh* locus contraction (Liu et al., 2007), we determined by 4Cseq analysis to what degree the conditional loss of YY1 affects long-range interactions in the *Igh* locus. Because B cell development was stringently arrested at the CD19$^+$ pro-B cell stage in *Rag1*$^{Cre/+}$*Yy1*$^{fl/fl}$ mice (Figure S6A), we used CD19 MACS sorting to isolate *Yy1*-deleted pro-B cells with high purity from the bone marrow of these mice (Figures S6B and S6C). 4C-seq analysis with the Eμ-HD and J606.5-HD viewpoints demonstrated that long-range interactions along the *Igh* locus could not be detected in YY1-deficient pro-B cells (Figures 7A and S6D), consistent with recent 3C-qPCR results indicating that YY1 is essential for looping of Eμ with PAIR4 and PAIR6 (Verma-Gaur et al., 2012). In contrast to *Pax5*$^{-/-}$ pro-B cells (Figure 2), local interactions in the proximal *Igh* domain were also altered in YY1-deficient pro-B cells, as shown by the loss of the Eμ-3′CBE interaction (Figure 7A). We therefore conclude that YY1 controls local as well as long-range interactions in the *Igh* locus.

YY1 is known to bind to the Eμ enhancer (Park and Atchison, 1991), which is, however, dispensable for long-range *Igh* looping (Figures 6A and 6B). To search for YY1-binding sites involved in long-range interactions, we generated a *Yy1*^{ihCd2} allele expressing a YY1-

Bio protein that can be biotinylated by the *E. coli* biotin ligase BirA (Figures S6E–S6H). B cell development was normal in *Yy1*$^{ihCd2/+}$ mice (data not shown), so we used short-term cultured pro-B cells from *Yy1*$^{ihCd2/+}$*Rosa26*$^{BirA/BirA}$*Rag2*$^{-/-}$ mice for streptavidin-mediated chromatin precipitation coupled with deep sequencing (Bio-ChIP-seq). As shown in Figures 2D and 7B, two prominent YY1 peaks were identified by Bio-ChIP sequencing at the Eμ enhancer and Cγ1-2b DHS site 1 in the proximal *Igh* domain. YY1 peaks were unevenly distributed in the V_H gene cluster with most peaks being present in the distal V_H gene region (Figure 7B). Several of these YY1-binding sites were located together with Pax5 peaks at PAIR elements and their associated V_H3609 genes (Figure 7C). We next examined by RNA sequencing whether YY1 controls the expression of known regulators of locus contraction. *Pax5*, *Ikzf1* (Ikaros), *Tcf3* (E2A), and *Ctcf* as well as *Smc1* and *Rad21* (encoding two cohesin subunits) were similarly expressed in sorted *Rag1*$^{Cre/+}$*Yy1*$^{fl/fl}$ and control *Rag1*$^{Cre/+}$*Yy1*$^{fl/+}$ pro-B cells (Figure 7D), consistent with a direct role of YY1 in controlling *Igh* locus contraction. In contrast, the noncoding antisense transcripts originating from PAIR4 and PAIR6 were strongly reduced in the absence of YY1 (Figure 7E). Collectively, these data demonstrate that YY1 controls antisense transcription and long-range interactions by binding to multiple sites in the distal V_H gene cluster.

DISCUSSION

V_H-DJ_H recombination depends on long-range interactions between the V_H gene cluster and proximal *Igh* domain containing the DJ_H-rearranged gene segment (Fuxa et al., 2004; Kosak et al., 2002). Here, we have investigated these interactions by high-resolution 4C sequencing, which revealed important aspects of *Igh* looping in lymphoid progenitors and committed pro-B cells.

In pro-B cells, all long-range *Igh* interactions were strictly confined to the *Igh* locus, which constitutes one large chromatin- folding unit. We identified the *Igh* 5′ and 3′ boundaries near the first 5′ V_HJ558 gene and at the 3′CBE with its nine CTCF-cohesin-binding sites, because the distal J558.89 and proximal HS8 viewpoints efficiently interacted with both *Igh* and adjacent non-*Igh* sequences, as expected for boundary elements. We were, however, unable to further localize the 5′ boundary region because of the absence of conspicuous features such as a specific chromatin signature or clustering of CTCF-binding sites in the V_HJ558.89 gene region (Ebert et al., 2011).

Distance measurements between 12 distinct *Igh* FISH probes combined with computer simulation predicted that the *Igh* locus is organized into compartments with clusters of loops separated by linkers (MLS model) in uncommitted pre-pro-B cells (Jhunjhunwala et al., 2008). Our 4C-seq analyses have provided experimental evidence for this notion, because the *Igh* locus in its “extended’ configuration consists of local interaction domains ranging from 0.5 Mb (3′ proximal region) to 1.3 Mb (5′ PAIR region) in uncommitted lymphoid progenitors and thymocytes. As shown by global Hi-C analyses, the entire mouse genome is organized into local chromatin interaction domains referred to as topological domains, which range from 0.5 to 3 Mb with a median size of 0.88 Mb (Dixon et al., 2012; Lin et al., 2012; Nora et al., 2012). The local interaction domains of the *Igh* locus also fit this size range, exhibit a similar 4C-seq pattern in both lymphoid progenitors and thymocytes, and

thus probably correspond to topological domains, which are largely invariant between cell types (Dixon et al., 2012; Lin et al., 2012). However, the local interaction domains in the *Igh* locus partially overlap with each other and thus differ from topological domains, which are separated by stringent boundaries (Dixon et al., 2012; Nora et al., 2012). Notably, the local interactions in the V_H gene cluster do not require Pax5 or PAIR activity for their formation in lymphoid progenitors and may thus reflect a default folding state of the *Igh* locus prior to full activation of the V_H gene region in committed pro-B cells.

In contrast, the proximal *Igh* domain is already active under the control of the Eμ enhancer in lymphoid progenitors (Afshar et al., 2006; Perlot et al., 2005). The local 3D architecture of the proximal domain consists of loops between the IGCR1, Eμ, 3′RR, and 3′CBE regions (Degner et al., 2011; Guo et al., 2011a, 2011b). Our systematic 4C-seq analysis identified additional interactions of these regulatory regions with elements in the constant gene region and the first proximal V_H genes in thymocytes, lymphoid progenitors, and committed pro-B cells. One of these interaction sites is located between the Cγ1 and Cγ2b gene segments in a region containing two DHS sites that bind Pax5 and other transcription factors in pro-B cells. The two Cγ1-2b DHS sites may function as Pax5-dependent enhancers regulating V(D)J or class switch recombination. The incorporation of proximal V_H genes into the local 3D architecture of the proximal *Igh* domain nicely explains why only the first proximal V_H genes undergo V_H-DJ_H recombination in Pax5-deficient progenitors (Hesslein et al., 2003; Roldá n et al., 2005). Surprisingly, the most proximal V_H genes also interact with the proximal *Igh* domain in thymocytes, although they never undergo V_H-DJ_H recombination during T cell development, unless the CBEs of the IGCR1 insulator are inactivated (Guo et al., 2011b). We conclude therefore that a functional IGCR1 element can prevent V_H-DJ_H recombination even if V_H genes participate in interactions with the proximal *Igh* domain in thymocytes. Moreover, the function of the IGCR1 insulator seems to be neutralized already in the absence of locus contraction in Pax5-deficient progenitors.

Igh locus contraction was extensively studied at the single-cell level by low-resolution DNA-FISH analysis. Recent 3C experiments, which provide high-resolution analysis of individual loops in a population of a million pro-B cells, have so far identified only four focal long-distance interactions between the Eμ enhancer and the V_H7183.18, V_HJ558.1, PAIR4, and PAIR6 regions (Guo et al., 2011a; Verma-Gaur et al., 2012). In contrast, our systematic 4C-seq analyses based on 16 different viewpoints have revealed a continuum of flexible long-range interactions across the entire V_H gene cluster in committed pro-B cells. This conclusion is based on the following evidence. First, long-range interactions were observed only in committed pro-B cells but not in Pax5- or YY1-deficient pro-B cells, indicating that these interactions were not caused by a 4C sequencing artifact. Second, 4C templates prepared with the same primary restriction enzyme (HindIII, EcoRI, or BglII) gave rise to similar long-range interaction patterns in the V_H gene region when analyzed from different proximal viewpoints. Third, the interaction profiles were distinct between 4C-seq experiments performed with different restriction enzymes that differ in the mappability of their end sequences along the V_H gene cluster. Finally, the long-range interaction patterns in the V_H gene region of committed pro-B cells were similar to the sequencing profiles obtained with random *Igh* BAC libraries, thus supporting the idea that the sum of all flexible

loops in a population of a million pro-B cells results in a continuum of long-range interactions. In contrast, the 3′ proximal *Igh* domain forms a more rigid 3D architecture, as shown by the fact that long-range loops from the distal V_H gene region primarily interacted with the IGCR1/D_H and 3′CBE regions in a manner that differed from the sequencing pattern of random *Igh* BAC libraries. We therefore hypothesize that the flexible long-range loops provide, at the pro-B cell population level, a similar probability for each V_H gene to be juxtaposed next to the proximal DJ_H-rearranged gene segment to undergo V_H-DJ_H recombination. By promoting equal participation of all V_H genes in V_HDJ_H recombination, flexible long-range loops contribute to the generation of a diverse antibody repertoire, which is essential for an effective immune response to foreign pathogens.

The question thus arises what sequence features may determine the structure of flexible loops in the V_H gene cluster. Unfortunately, it is the flexible nature of the long-range loops itself that prevents the identification of interacting elements, which may be present at regular intervals similar to the dense distribution of CTCF- and Pax5-binding sites along the V_H gene cluster (Ebert et al., 2011; Revilla-I-Domingo et al., 2012). Importantly, the local interaction domains in Pax5-deficient progenitors and the corresponding long-range loops in committed pro-B cells revealed similar 4C-seq patterns in the distal V_H gene region. Hence, the flexible loops along the V_H gene cluster may be a characteristic feature of the local interaction domains, which coalesce into a large interaction territory upon locus contraction by a so-far-unknown mechanism.

As shown by 4C sequencing, long-range interactions across the V_H gene cluster were absent in pro-B cells lacking Pax5 or YY1, consistent with previous DNA-FISH analyses (Fuxa et al., 2004; Liu et al., 2007). In the absence of Pax5, local *Igh* domains were normally formed in lymphoid progenitors, as shown by the fact that Pax5 expression is activated only later in committed pro-B cells. Local *Igh* domains were also generated in the absence of the ubiquitous regulator YY1, although specific local interactions, such as the Eμ-3′CBE loop, required YY1 function. Bio-ChIP sequencing identified YY1-binding sites in the 3′ proximal domain and middle V_H gene region. However, the majority of YY1-binding sites were found in the distal V_HJ558 region, where YY1 peaks colocalized together with Pax5-binding sites at six PAIR elements (2, 4, 6, 7, 11, and 12) and their associated V_H3609 genes. Consistent with this finding, noncoding antisense transcripts originating from PAIR4 and PAIR6 were lost in YY1-deficient pro-B cells (Verma-Gaur et al., 2012; this study) similar to *Pax5* mutant pro-B cells (Ebert et al., 2011). Hence, Pax5 and YY1 both control the activity of PAIR elements and mediate long-range interactions across the V_H gene cluster, thus supporting a role for PAIR elements in regulating *Igh* locus contraction.

The PAIR elements 4, 6, and 11 exhibit the highest transcriptional activity within the V_H gene cluster and have therefore been postulated to induce long-range looping by relocating the PAIR-containing V_H gene sequences to transcription factories in the nucleus where they interact with the proximal *Igh* domain, which is transcriptionally activated by the Eμ enhancer (Verma-Gaur et al., 2012). This hypothesis predicts that the PAIR elements with the highest transcriptional activity should constitute interaction hotspots in the distal V_H gene region. In marked contrast, our 4C-seq experiments indicate that all V_H gene sequences contact the proximal *Igh* domain with similar probability. Alternatively, the regularly spaced

CTCF-cohesin-binding sites in the V_H gene region may be involved in the formation of flexible loops within local interaction domains in lymphoid progenitors. Once expressed in committed pro-B cells, Pax5 may then provide another layer of interactions between the many Pax5-binding sites and CTCF-cohesin-binding elements in the *Igh* locus, thus promoting long-range interactions among local interaction domains. In support of this hypothesis, we have shown that Pax5 and CTCF interact with each other in pro-B cells and in vitro binding assays. Moreover, long-range interactions between IGCR1 and distal V_H gene sequences depend on functional CTCF-binding sites in the IGCR1 region. Finally, long-range loops from the V_H gene region primarily contact the IGCR1 and 3′CBE regions, which are the only elements containing CTCF-binding sites in the proximal *Igh* domain.

The chromatin accessibility and transcriptional activation of the proximal *Igh* domain is stringently controlled by the Eμ enhancer (Afshar et al., 2006; Chakraborty et al., 2009; Perlot et al., 2005). The V_H-DJ_H recombination block, which is observed in Eμ mutant pro-B cells, is probably a secondary effect of the strongly reduced number of DJ_H-rearranged *Igh* alleles (Afshar et al., 2006; Perlot et al., 2005). Notably, the recombination of proximal and distal V_H genes is equally affected by deletion of the Eμ enhancer (Perlot et al., 2005), implying that the V_H-DJ_H recombination phenotype is not caused by the loss of *Igh* locus contraction. Moreover, germline transcription of proximal and distal V_H genes proceeds normally in Eμ mutant pro-B cells, indicating that the activity of the Eμ enhancer is restricted to the proximal *Igh* domain (Afshar et al., 2006; Perlot et al., 2005). A recent report has, however, implicated the Eμ enhancer in the control of *Igh* locus contraction and thus V_H-DJ_H recombination (Guo et al., 2011a). This conclusion was primarily based on 3D DNA-FISH analysis of the *Igh* contraction state in *Rag2*$^{-/-}$ Eμ mutant and J_HT-deficient pro-B cells compared to control *Rag2*$^{-/-}$ pro-B cells (Guo et al., 2011a). By performing 4C-seq and 3D DNA-FISH experiments with the same pro-B cell types, we could, however, not confirm a role for the Eμ enhancer in the regulation of *Igh* locus contraction by these two independent methods. Similar to the Eμ enhancer, our 4C-seq and 3D DNAFISH analyses did not provide evidence for a role of the IGCR1 and 3′RR elements in *Igh* long-range looping, indicating that individual mutations of these three regulatory elements in the proximal *Igh* domain do not affect locus contraction. These data are compatible with the testable hypothesis that the V_H gene region and proximal domain of the *Igh* locus have separable functions. A major and intrinsic function of the V_H gene cluster seems to be the formation of long-range loops to present V_H genes to the proximal domain, which functions as the business end of the *Igh* locus where V(D)J recombination takes place.

EXPERIMENTAL PROCEDURES

Detailed methods can be found in the Supplemental Information available online. All animal experiments were carried out according to valid project licenses, which were approved and regularly controlled by the Austrian Veterinary Authorities.

4C Sequencing

4C templates were prepared from pro-B cells and thymocytes, as described (Simonis et al., 2007). For each viewpoint, 16 individual PCR amplifications with 35 cycles were performed

with ~150 ng of 4C template with specific primers that were extended with Illumina sequencing adaptors (Table S1). The pooled PCR products were selected for a fragment size less than 1 kb on agarose gels prior to single-end sequencing with the HiSeq 2000 system resulting in a read length of 100 nucleotides.

Random 4C-Seq BAC Library

For template generation, equimolar DNA mixtures of C57BL/6 BAC clones covering the mouse *Igh* locus (Figure S4B) were digested with HindIII and ligated a high DNA concentration followed by Sau3AI digestion, ligation at low DNA concentration, and PCR amplification with viewpoint-specific primers.

3C-qPCR Analysis

3C templates of pro-B cells were prepared and subjected to TaqMan PCR analysis with primers listed in Table S2, as described (Hagége et al., 2007).

Bio-ChIP-Seq Analysis of YY1 Binding

Chromatin was prepared from short-term cultured pro-B cells of *Yy1*$^{ihCd2/+}$ *Rosa26*$^{BirA/BirA}$*Rag2*$^{-/-}$ mice followed by streptavidin pull-down, as described (Ebert et al., 2011). The precipitated DNA were submitted to paired-end sequencing with the Illumina GAIIx system resulting in a read length of 76 nucleotides. YY1 peaks were called with MACS (version 1.3.6.1) and filtered for p values of $<10^{-10}$.

Nuclear Extract Preparation and Coprecipitation Analysis

Pax5-Bio was precipitated from nuclear extracts of Abl-MLV-transformed *Pax5*$^{Bio/Bio}$ and *Rosa26*$^{BirA/BirA}$ pro-B cells by streptavidin pull-down, and CTCF was precipitated from nuclear extracts of cultured *Rag2*$^{-/-}$ pro-B cells with a CTCF mAb (clone 48, BD), as described (McManus et al., 2011).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank A. Sommer and P. Stolt-Berger at the Campus Science Support Facilities for deep sequencing and protein expression, E. Axelsson for RNA-seq analysis, W. van Ijcken and F. Grosveld for advice on 4C sequencing, J. Skok for help with DNA-FISH analysis, and J.-M. Peters for advice on CTCF purification. This research was supported by Boehringer-Ingelheim, the Vienna Science and Technology Fund (WWTF), an ERC Advanced Grant (291740-LymphoControl) from the European Community's Seventh Framework Program, the Austrian GEN-AU initiative (financed by the Bundesminsterium für Bildung und Wissenschaft), and an EMBO fellowship (to T.S.).

REFERENCES

Afshar R, Pierce S, Bolland DJ, Corcoran A, Oltz EM. Regulation of IgH gene assembly: role of the intronic enhancer and 5'D_{Q52} region in targeting D_HJ_H recombination. J. Immunol. 2006; 176:2439–2447. [PubMed: 16456003]

Bolland DJ, Wood AL, Johnston CM, Bunting SF, Morgan G, Chakalova L, Fraser PJ, Corcoran AE. Antisense intergenic transcription in V(D)J recombination. Nat. Immunol. 2004; 5:630–637. [PubMed: 15107847]

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]

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. USA. 2011; 108:9566–9571. [PubMed: 21606361]

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012; 485:376–380. [PubMed: 22495300]

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, Roldán 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]

Garrett FE, Emelyanov AV, Sepulveda MA, Flanagan P, Volpi S, Li F, Loukinov D, Eckhardt LA, Lobanenkov VV, Birshtein BK. Chromatin architecture near a potential 3′ end of the *igh* locus involves modular regulation of histone modifications during B-Cell development and in vivo occupancy at CTCF sites. Mol. Cell. Biol. 2005; 25:1511–1525. [PubMed: 15684400]

Gu H, Zou YR, Rajewsky K. Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-*loxP*-mediated gene targeting. Cell. 1993; 73:1155–1164. [PubMed: 8513499]

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]

Hadjur S, Williams LM, Ryan NK, Cobb BS, Sexton T, Fraser P, Fisher AG, Merkenschlager M. Cohesins form chromosomal *cis*-interactions at the developmentally regulated *IFNG* locus. Nature. 2009; 460:410–413. [PubMed: 19458616]

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

Hesslein DGT, Pflugh DL, Chowdhury D, Bothwell ALM, 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]

Hewitt SL, Yin B, Ji Y, Chaumeil J, Marszalek K, Tenthorey J, Salvagiotto G, Steinel N, Ramsey LB, Ghysdael J, et al. RAG-1 and ATM coordinate monoallelic recombination and nuclear positioning of immunoglobulin loci. Nat. Immunol. 2009; 10:655–664. [PubMed: 19448632]

Jhunjhunwala S, van Zelm MC, Peak MM, Cutchin S, Riblet R, van Dongen JJM, Grosveld FG, Knoch TA, Murre C. The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. Cell. 2008; 133:265–279. [PubMed: 18423198]

Jhunjhunwala S, van Zelm MC, Peak MM, Murre C. Chromatin architecture and the generation of antigen receptor diversity. Cell. 2009; 138:435–448. [PubMed: 19665968]

Johnston CM, Wood AL, Bolland DJ, Corcoran AE. Complete sequence assembly and characterization of the C57BL/6 mouse Ig heavy chain V region. J. Immunol. 2006; 176:4221–4234. [PubMed: 16547259]

Kosak ST, Skok JA, Medina KL, Riblet R, Le Beau MM, Fisher AG, Singh H. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. Science. 2002; 296:158–162. [PubMed: 11935030]

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, Shi Y. Yin Yang 1 is a critical regulator of B-cell development. Genes Dev. 2007; 21:1179–1189. [PubMed: 17504937]

McManus S, Ebert A, Salvagiotto G, Medvedovic J, Sun Q, Tamir I, Jaritz M, Tagoh H, Busslinger M. The transcription factor Pax5 regulates its target genes by recruiting chromatin-modifying proteins in committed B cells. EMBO J. 2011; 30:2388–2404. [PubMed: 21552207]

Nativio R, Wendt KS, Ito Y, Huddleston JE, Uribe-Lewis S, Woodfine K, Krueger C, Reik W, Peters J-M, Murrell A. Cohesin is required for higher-order chromatin conformation at the imprinted *IGF2-H19* locus. PLoS Genet. 2009; 5:e1000739. [PubMed: 19956766]

Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012; 485:381–385. [PubMed: 22495304]

Park K, Atchison ML. Isolation of a candidate repressor/activator, NF-E1 (YY-1, δ), that binds to the immunoglobulin κ 3′ enhancer and the immunoglobulin heavy-chain μ E1 site. Proc. Natl. Acad. Sci. USA. 1991; 88:9804–9808. [PubMed: 1946405]

Perlot T, Alt FW. Cis-regulatory elements and epigenetic changes control genomic rearrangements of the *IgH* locus. Adv. Immunol. 2008; 99:1–32. [PubMed: 19117530]

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

Revilla-I-Domingo R, Bilic I, Vilagos B, Tagoh H, Ebert A, Tamir IM, Smeenk L, Trupke J, Sommer A, Jaritz M, Busslinger M. The B-cell identity factor Pax5 regulates distinct transcriptional programmes in early and late B lymphopoiesis. EMBO J. 2012; 31:3130–3146. [PubMed: 22669466]

Roldán E, Fuxa M, Chong W, Martinez D, Novatchkova M, Busslinger M, Skok JA. Locus ‘decontraction’ and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. Nat. Immunol. 2005; 6:31–41. [PubMed: 15580273]

Sayegh CE, Jhunjhunwala S, Riblet R, Murre C. Visualization of looping involving the immunoglobulin heavy-chain locus in developing B cells. Genes Dev. 2005; 19:322–327. [PubMed: 15687256]

Simonis M, Kooren J, de Laat W. An evaluation of 3C-based methods to capture DNA interactions. Nat. Methods. 2007; 4:895–901. [PubMed: 17971780]

Splinter E, Heath H, Kooren J, Palstra R-J, Klous P, Grosveld F, Galjart N, de Laat W. CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. Genes Dev. 2006; 20:2349–2354. [PubMed: 16951251]

van de Werken HJ, de Vree PJ, Splinter E, Holwerda SJ, Klous P, de Wit E, de Laat W. 4C technology: protocols and data analysis. Methods Enzymol. 2012; 513:89–112. [PubMed: 22929766]

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. USA. 2012; 109:17004–17009. [PubMed: 23027941]

Vincent-Fabert C, Fiancette R, Pinaud E, Truffinet V, Cogné N, Cogné M, Denizot Y. Genomic deletion of the whole IgH 3′ regulatory region (hs3a, hs1,2, hs3b, and hs4) dramatically affects class switch recombination and Ig secretion to all isotypes. Blood. 2010; 116:1895–1898. [PubMed: 20538806]

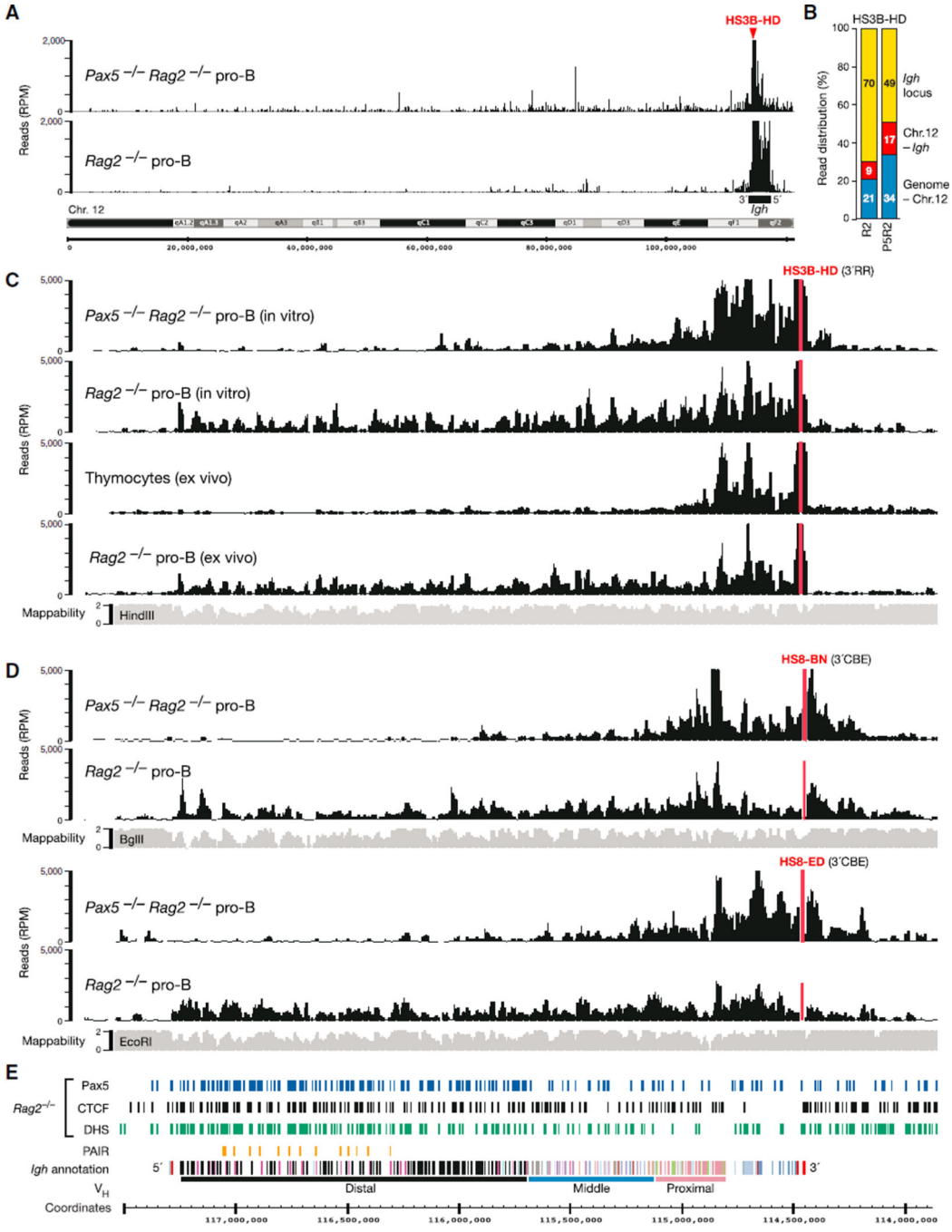

Figure 1. High-Resolution Analysis of Long-Range Interactions within the *Igh* Locus

(A) Interaction of the HS3B-HD viewpoint (red) with chromosome 12 sequences in cultured *Pax5*$^{-/-}$*Rag2*$^{-/-}$ and *Rag2*$^{-/-}$ pro-B cells. The 20 kb running mean values of the 4C-seq reads were plotted as reads per million mapped sequence reads (RPMs). A black bar denotes the *Igh* locus.

(B) Relative distribution of the 4C-seq reads (HS3B-HD viewpoint) in the *Igh* locus, rest of chromosome 12 (Chr.12 – *Igh*), or entire genome minus chromosome 12 (Genome – Chr. 12).

(C) Pax5-dependent long-range interactions of the HS3B region in ex vivo sorted and in vitro cultured cells of the indicated genotypes.
(D) 4C-seq patterns of the HS8-BN and HS8-ED viewpoints in cultured *Pax5*$^{-/-}$*Rag2*$^{-/-}$ and *Rag2*$^{-/-}$ pro-B cells. A 20 kb running mean mappability track (gray) indicates to what degree (on a scale from 0 to 2) the end sequences of HindIII (C), BglII (D), and EcoRI (D) fragments can be unambiguously mapped to genomic *Igh* sequences. One representative 4C-seq experiment is shown for each viewpoint and cell type (C and D).
(E) Annotation of the C57BL/6 Igh locus. The distinct V_H gene families (different colors) in the distal, middle, and proximal V_H gene regions (Johnston et al., 2006) as well as the D_H (gray) and C_H (blue) elements and Eμ and 3′RR enhancers (red) in the 3′ proximal Igh domain are shown together with the PAIR elements (orange) and the mm9 genomic coordinates of mouse chromosome 12. The indicated DNase I hypersensitive (DHS) regions and CTCF- and Pax5-binding sites were determined in Rag2$^{-/-}$ pro-B cells by paired-end sequencing and peak calling (MACS; p value < 10^{-10}) (Ebert et al., 2011; Revilla-I-Domingo et al., 2012).
See also Figure S1.

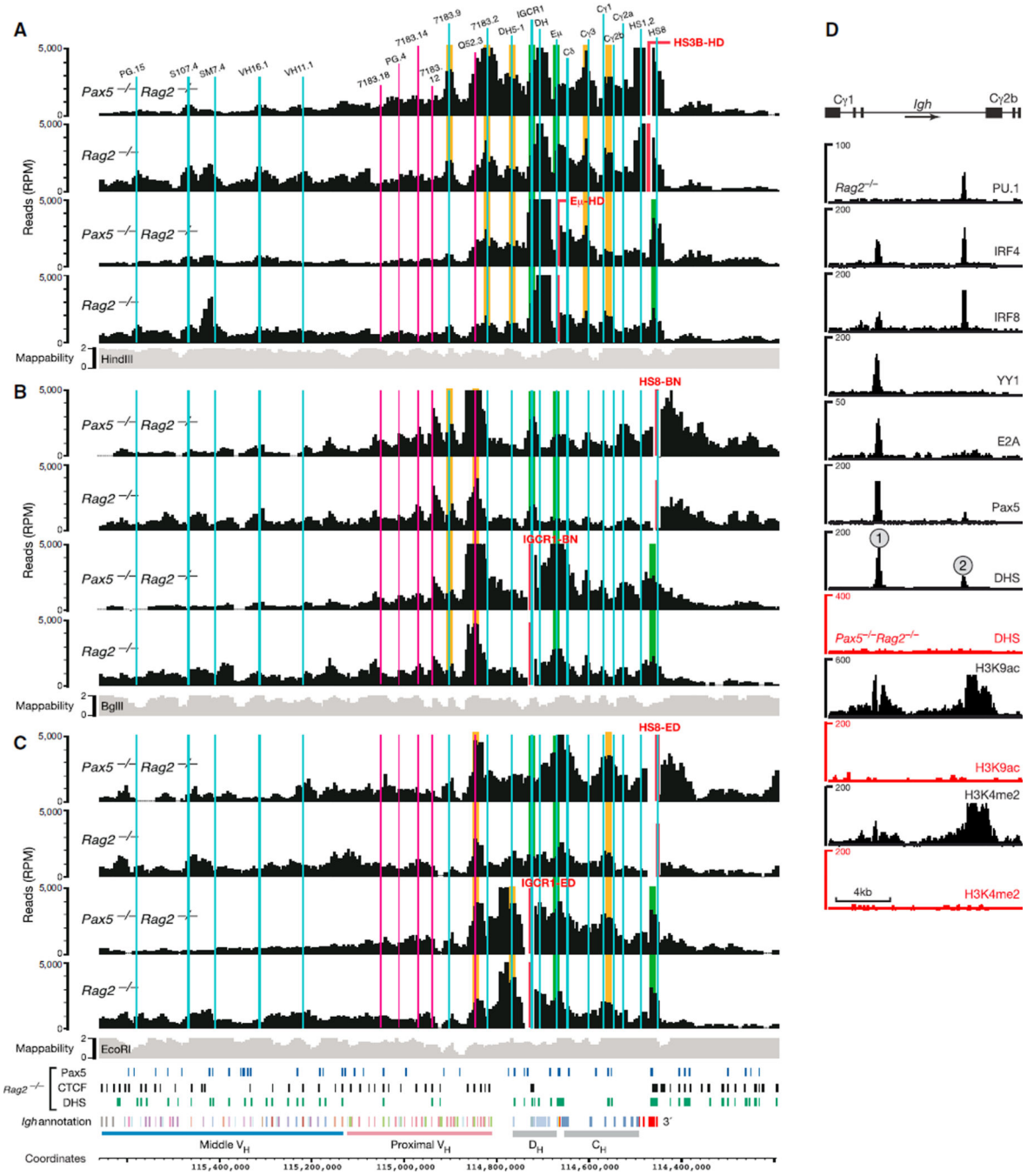

Figure 2. Loop Formation in the 3′ Proximal *Igh* Region

(A–C) 4C-seq interaction patterns revealed by the indicated viewpoints (red) in short-term cultured *Pax5*$^{-/-}$*Rag2*$^{-/-}$ and *Rag2*$^{-/-}$ pro-B cells. Vertical lines indicate the positions of relevant V_H, D_H, and C_H gene segments as well as IGCR1, Eμ, and DNase I hypersensitive sites constituting the 3′RR (HS1–HS4) and 3′CBE (HS5–HS8) regions. Previously known and newly identified interactions are highlighted in green and orange, respectively. See legend of Figure 1 for further explanations. The data in (A) are based on average values of

three independent experiments, and the results shown in (B) and (C) are derived from one experiment.

(D) Characterization of the Cγ1-Cγ2b region. The sequences between the Cγ1 and Cγ2b gene segments were analyzed for binding of the indicated transcription factors and the presence of DHS sites and active histone modifications (H3K4me2 and H3K9ac) by deep sequencing of *Pax5*$^{-/-}$*Rag2*$^{-/-}$ (red) and *Rag2*$^{-/-}$ (black) pro-B cells (Revilla-I-Domingo et al., 2012).

See also Figure S2.

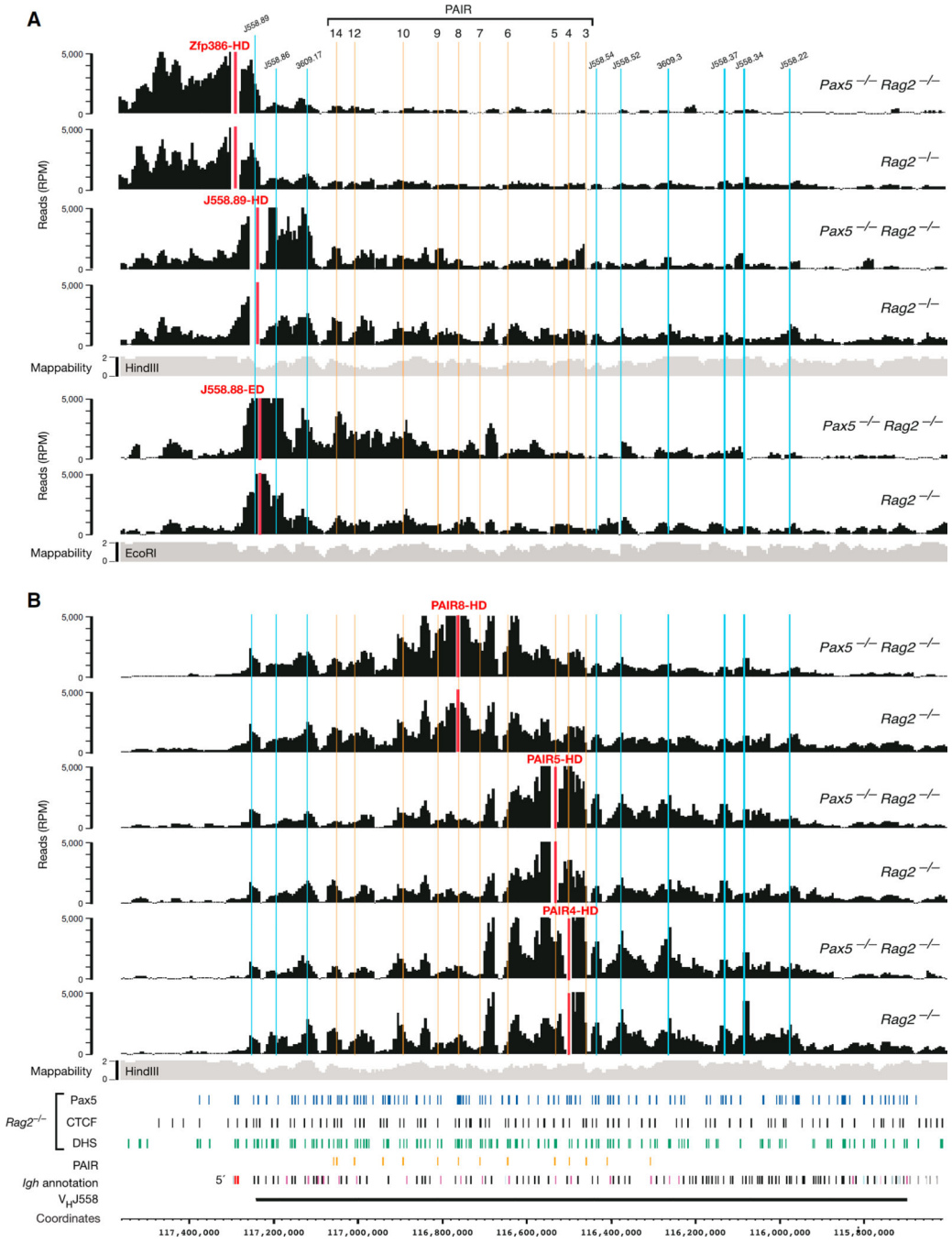

Figure 3. Local Interactions within the Distal V_H Gene Cluster in the Absence of Locus Contraction

(A) 4C-seq interaction profiles at the *Igh* 5′ end in cultured *Pax5*$^{-/-}$*Rag2*$^{-/-}$ and *Rag2*$^{-/-}$ pro-B cells. Vertical lines indicate the positions of PAIR elements (yellow) and selected V_H genes (blue).

(B) 4C-seq interaction patterns detected with the indicated PAIR viewpoints (red) in the distal V_H gene region. See legend of Figure 1 for further explanations. One 4C-seq experiment was performed with viewpoints Zfp386-HD and J558.88-ED, and the data of all other viewpoints are based on average values of three experiments.

See also Figure S3.

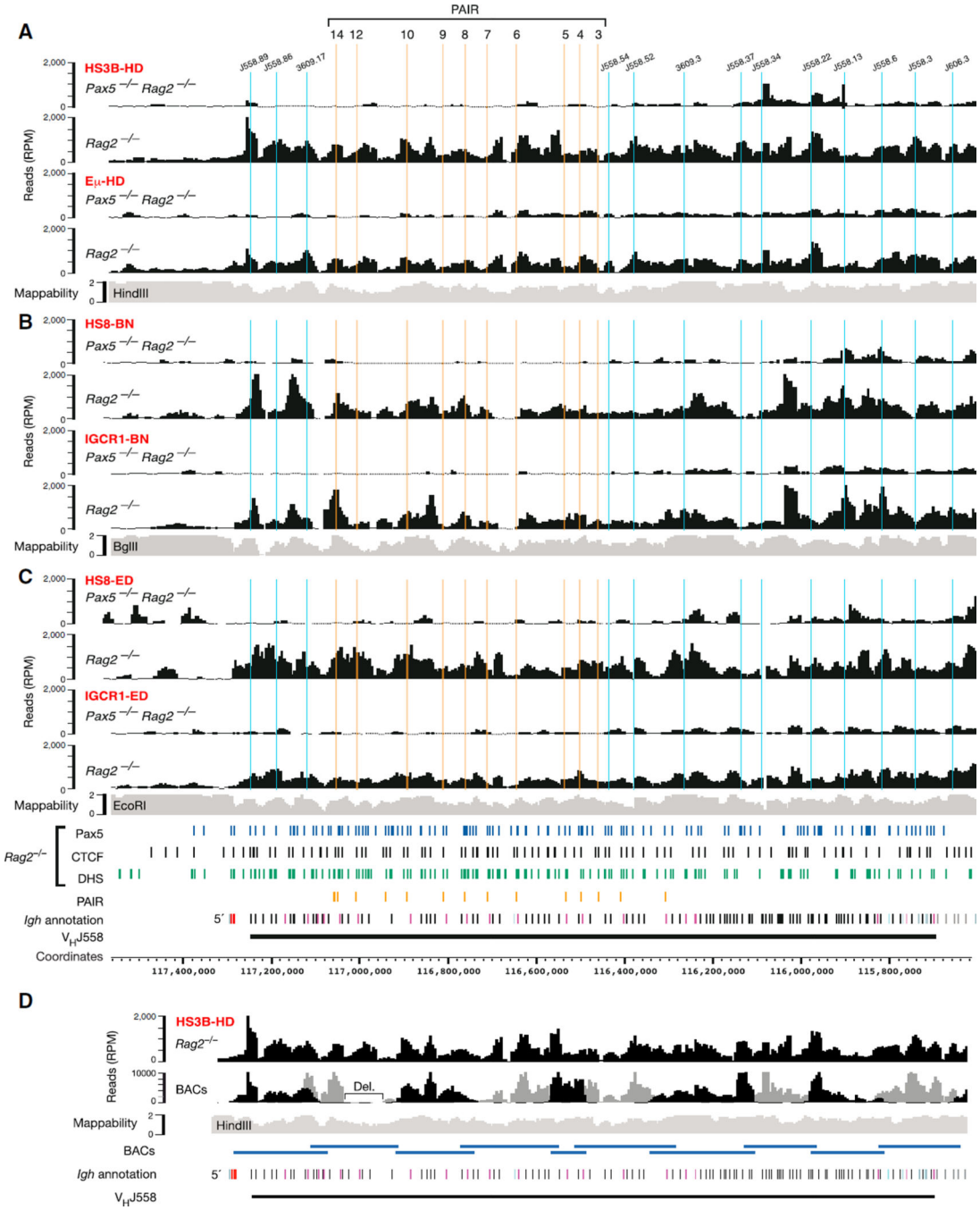

Figure 4. Flexible Long-Range Interactions along the Distal V_H Gene Region in the Presence of Locus Contraction

(A–C) Long-range interactions of the indicated proximal viewpoints (red) with the distal V_H gene region in cultured *Pax5*$^{-/-}$*Rag2*$^{-/-}$ and *Rag2*$^{-/-}$ pro-B cells. See legend of Figure 1 for further explanations. The data of the HS3B-HD and Eμ-HD viewpoints (A) are average values of three experiments, and all other viewpoints (B, C) were analyzed once.

(D) Comparison of the 4C-seq patterns obtained with random BAC libraries and *Rag2*$^{-/-}$ pro-B cells (analyzed from HS3B-HD). The sequencing pattern of each BAC (with its size

indicated below) was normalized by setting the highest peak to an RPM value of 10,000 (Figure S4B) prior to merging the patterns of the two BAC libraries (indicated in black and gray).

Abbreviation: Del., internal BAC deletion. See also Figure S4.

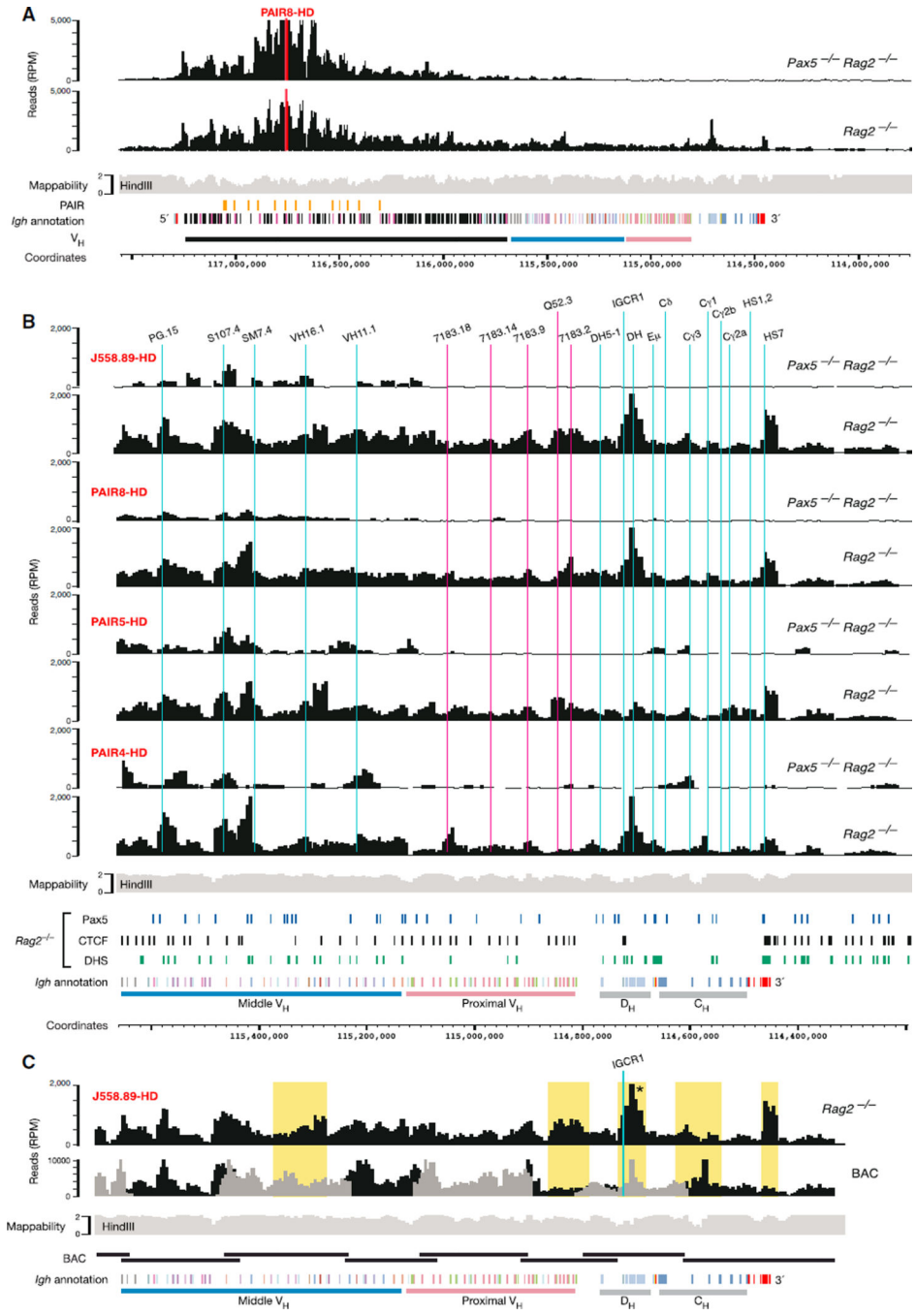

Figure 5. Long-Range Interactions of the Distal V_H Gene Region with the Proximal *Igh* Domain

(A) 4C-seq interaction pattern of PAIR8 with sequences across the *Igh* locus in cultured *Pax5*$^{-/-}$*Rag2*$^{-/-}$ and *Rag2*$^{-/-}$ pro-B cells.

(B) Long-range interactions of the indicated distal viewpoints (red) with sequences of the *Igh* 3′ region in the indicated cultured pro-B cell types. See legends of Figures 1 and 2 for further explanations. The data of all viewpoints in (A) and (B) are average values of three experiments except for the PAIR4-HD analysis of *Pax5*$^{-/-}$*Rag2*$^{-/-}$ pro-B cells (average of two experiments).

(C) Comparison of the 4C-seq patterns between random BAC libraries and *Rag2*$^{-/-}$ pro-B cells (analyzed from J558.89-HD). See legend of Figure 4D for normalization of the BAC sequencing pattern. Yellow shading denotes regions with a 4C-seq pattern that differs between the random BAC libraries and *Rag2*$^{-/-}$ pro-B cells. The specificity of the 4C-seq interactions in the IGCR1/D_H region (asterisk) of *Rag2*$^{-/-}$ pro-B cells compared to the random BAC library was confirmed by 3C-qPCR (Figure S4E).

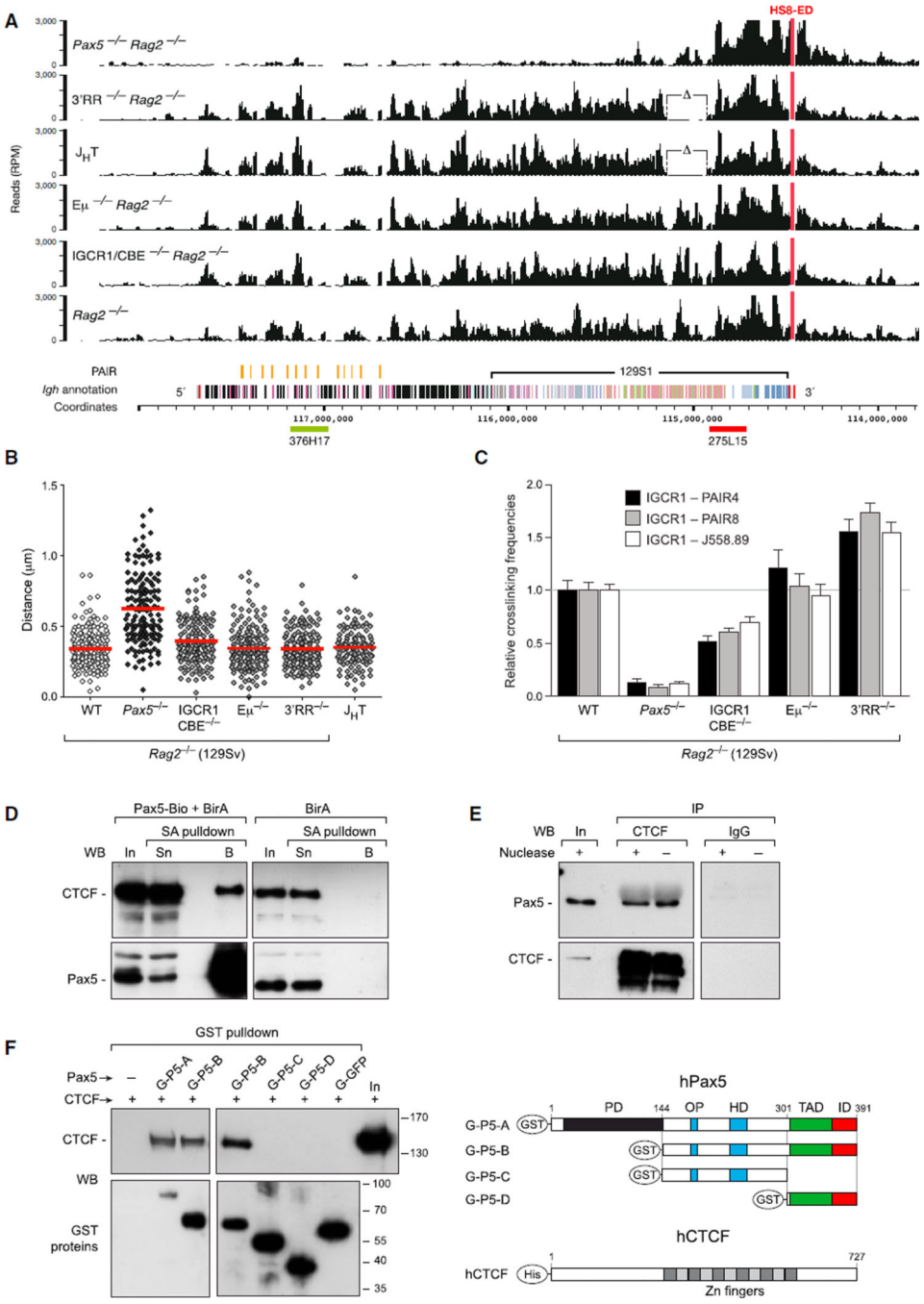

Figure 6. Role of CTCF and 3′ Regulatory Elements in Long-Range *Igh* Interactions

(A) 4C-seq analysis of cultured 129Sv pro-B cells of the indicated genotypes with the HS8-ED viewpoint (red). The homozygous J_HT deletion eliminates the D_HQ52, J_H, and Eμ elements (Gu et al., 1993). The 129Sv sequence reads were mapped to genomic C57BL/6 sequences except for the available 129S1 *Igh* sequences (see Supplemental Experimental Procedures). A 120 kb deletion (Δ) was present in the E14 (129Ola) ESCs used to generate the J_HT and 3′RR mutant alleles. 4C-seq experiments were performed once with *Rag2*$^{-/-}$

3′RR (–/–) mutant, *Rag2*$^{-/-}$ Eμ (–/–) mutant, and J_HT-deficient pro-B cells and twice with 129Sv *Rag2*$^{-/-}$, *Pax5*$^{-/-}$*Rag2*$^{-/-}$, and *Rag2*$^{-/-}$ IGCR1/CBE (–/–) mutant pro-B cells.
(B) Two-color 3D DNA-FISH analysis with *Igh* BAC probes (indicated in A). Dot plots show the distances measured between the two DNA signals of individual *Igh* alleles together with the average distance determined for each genotype.
(C) 3C-qPCR analysis of the IGCR1 interaction with distal PAIR4, PAIR8, and V_HJ558.89 sequences in cultured 129Sv pro-B cells of the indicated genotypes. A total of 12 3C templates of each genotype were used to determine the relative crosslinking frequency (see Supplemental Experimental Procedures), which was set to 1 for *Rag2*$^{-/-}$ pro-B cells and is shown with the standard error of the mean.
(D) Coprecipitation of CTCF with Pax5-Bio by streptavidin (SA) pull-down of nuclear extracts prepared from Abl-MLV-transformed pro-B cells of the *Pax5*$^{Bio/Bio}$ (Pax5-Bio) or control *Rosa26*$^{BirA/BirA}$ (BirA) genotype. The input (In; 1/100), supernatant (Sn; 1/100), and streptavidin-bound (B) precipitate were analyzed by immunoblotting (WB) with CTCF and Pax5 antibodies.
(E) Coimmunoprecipitation of Pax5 from a nuclear extract of *Rag2*$^{-/-}$ pro-B cells with CTCF antibodies followed by immunoblotting with a biotinylated rat anti-Pax5 mAb (detected with streptavidin-coupled horse radish peroxidase). Input (In; 1/100) and rabbit IgG were used as controls. Where indicated, DNA was digested with the endonuclease Benzonase during nuclear extract preparation.
(F) Direct Pax5-CTCF protein interaction. The hexahistidine-tagged CTCF and GST-Pax5 proteins (schematically depicted to the right) were affinity purified from baculovirus-infected Sf9 cells (Figure S5B) and used for in vitro binding and GST pull-down assays followed by immunoblotting of the precipitates with CTCF and GST antibodies (see Supplemental Experimental Procedures). A protein size marker (in kilodaltons) and input (In; 1/20) are shown. Empty glutathione-sepharose beads (−) and GST-GFP were used as negative controls. Abbreviations are as follows: G, GST; P5, Pax5; PD, paired domain; OP, octapeptide; HD, partial homeodomain; TAD, transactivation domain; ID, inhibitory domain.
See also Figure S5.

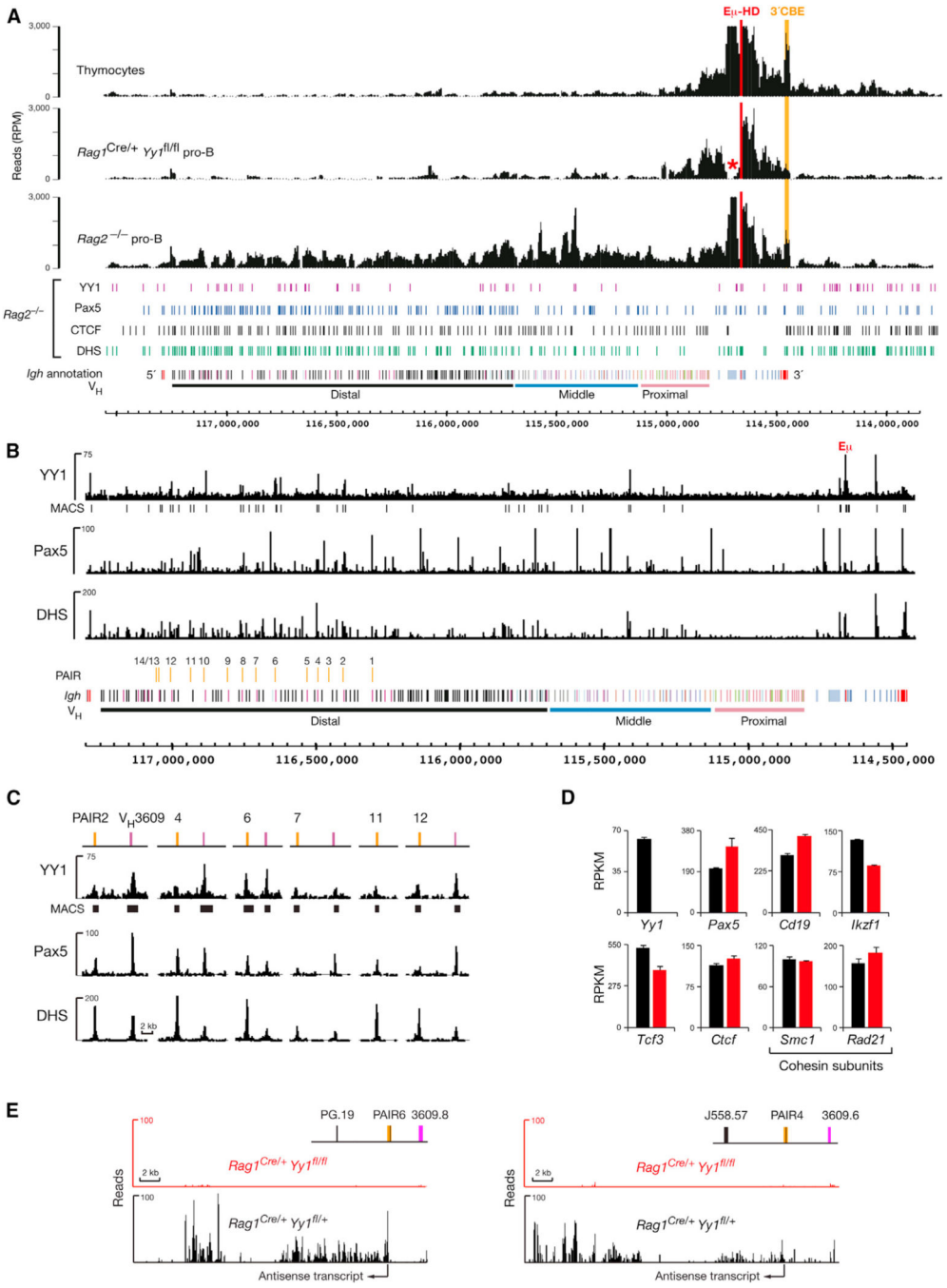

Figure 7. YY1 Controls PAIR Function and Long-Range Interactions across the *Igh* Locus

(A) 4C-seq analysis of ex vivo sorted *Rag1*$^{Cre/+}$*Yy1*$^{fl/fl}$ pro-B cells (Figures S6B and S6C), control *Rag2*$^{-/-}$ pro-B cells, and thymocytes with the viewpoint Eμ-HD (red). The Eμ-3′CBE interaction is highlighted in orange. A red asterisk indicates the loss of D_H sequences by D_H-J_H rearrangements in the recombination-proficient *Rag1*$^{Cre/+}$*Yy1*$^{fl/fl}$ pro-B cells.

(B) Identification of YY1 peaks in the *Igh* locus by Bio-ChIP-sequencing of cultured pro-B cells from *Yy1*$^{ihCd2/+}$*Rosa26*$^{BirA/BirA}$*Rag2*$^{-/-}$ mice. Vertical bars below the Bio-ChIP-seq

track identify significant YY1 peaks that were called by MACS with a p value of $<10^{-10}$. For comparison, the Pax5-binding and DHS site patterns of *Rag2*$^{-/-}$ pro-B cells are shown (Revilla-I-Domingo et al., 2012).

(C) YY1 binding at PAIR elements and associated V_H3609 genes.

(D and E) Analysis of YY1-dependent gene expression. The expression values (RPKM) of selected genes (D) and RNA-seq profiles of the antisense transcripts originating from PAIR4 and PAIR6 (E) are shown for *Rag1*$^{Cre/+}$*Yy1*$^{fl/fl}$ (red) and control *Rag1*$^{Cre/+}$*Yy1*$^{fl/+}$ (black) pro-B cells together with the standard error of the mean (based on two RNA-seq experiments for each genotype).

See also Figure S6.