

# A structural hierarchy mediated by multiple nuclear factors establishes *IgH* locus conformation

Tatiana Gerasimova,<sup>1</sup> Changying Guo,<sup>1,3</sup> Amalendu Ghosh,<sup>1</sup> Xiang Qiu,<sup>1</sup> Lindsey Montefiori,<sup>1,4</sup> Jiyoti Verma-Gaur,<sup>2</sup> Nancy M. Choi,<sup>2</sup> Ann J. Feeney,<sup>2</sup> and Ranjan Sen<sup>1</sup>

<sup>1</sup>Laboratory of Molecular Biology and Immunology, National Institute on Aging, Baltimore, Maryland 21224, USA; <sup>2</sup>The Scripps Research Institute, La Jolla, California 92037, USA

**Conformation of antigen receptor gene loci spatially juxtaposes rearranging gene segments in the appropriate cell lineage and developmental stage. We describe a three-step pathway that establishes the structure of the 2.8-Mb immunoglobulin heavy chain gene (*IgH*) locus in pro-B cells. Each step uses a different transcription factor and leads to increasing levels of structural organization. CTCF mediates one level of compaction that folds the locus into several 250- to 400-kb subdomains, and Pax5 further compacts the 2-Mb region that encodes variable ( $V_H$ ) gene segments. The 5' and 3' domains are brought together by the transcription factor YY1 to establish the configuration within which gene recombination initiates. Such stepwise mechanisms may apply more generally to establish regulatory fine structure within megabase-sized topologically associated domains.**

[*Keywords:* chromosome conformation; looping; architectural proteins]

Supplemental material is available for this article.

Received April 15, 2015; revised version accepted July 29, 2015.

Immunoglobulin heavy chain (*IgH*) genes are assembled by rearrangement of variable ( $V_H$ ), diversity ( $D_H$ ), and joining ( $J_H$ ) gene segments spread over 2.8 Mb of the genome (Fig. 1A; Subrahmanyam and Sen 2012). The order of recombination is precise, with  $D_H$  to  $J_H$  recombination occurring first followed by  $V_H$  recombination to newly created  $DJ_H$  junctions. Regulated access of lymphocyte-specific recombination-activating gene 1 (RAG1) and RAG2 initiates the process by introducing DNA double-strand breaks at specific recombination signal sequences associated with rearrangeable gene segments (Schatz and Ji 2011). Thereafter, breaks are repaired by the ubiquitously expressed nonhomologous end-joining (NHEJ) machinery; genomic instability inherent in a process that simultaneously introduces double-strand breaks at two sites in the genome requires V(D)J recombination to be tightly regulated. This is determined in large part by epigenetic means (Bossen et al. 2012; Shih and Krangel 2013).

One key mechanism involves folding the *IgH* locus into conformations that bring  $V_H$  gene segments located 0.1–2 Mb away into spatial proximity with  $D_H$  gene segments. The earliest evidence for this was the demonstration

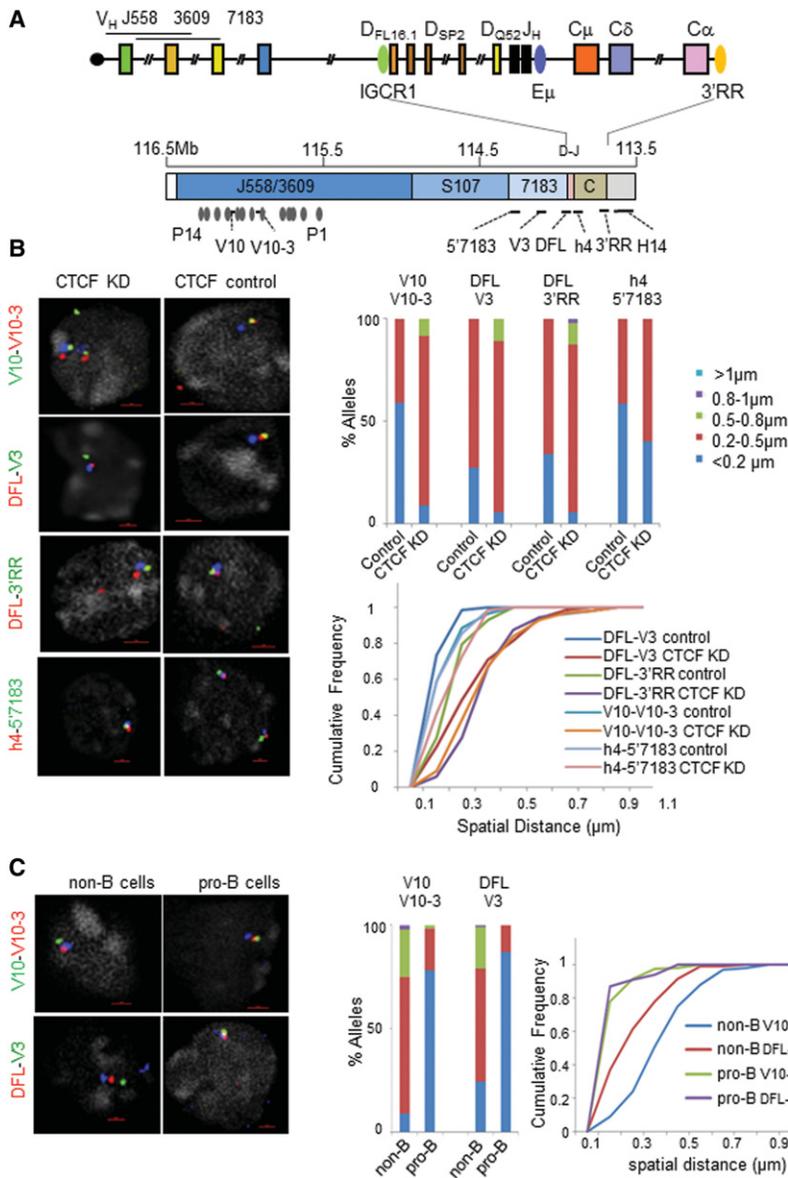
that 5' and 3' ends of the *IgH* locus are located in spatial proximity in developing B cells where *IgH* genes rearrange (Kosak et al. 2002; Sayegh et al. 2005). This phenomenon is referred to as locus compaction (or contraction), and its functional significance is inferred from the strong correlation between reduced locus compaction and restricted use of  $D_H$ -distal  $V_H$  gene segments in pro-B cells of mice bearing several genetic mutations that affect B-cell development. The most prominent of these are deletions of genes encoding the transcription factors Pax5 (Fuxa et al. 2004) and YY1 (Liu et al. 2007; Verma-Gaur et al. 2012) and “*cis*” mutations within the *IgH* locus that delete the intronic enhancer  $E_\mu$  (Guo et al. 2011a). A compacted *IgH* locus structure may also minimize hazardous DNA translocation events during V(D)J recombination. The transcription factor CTCF plays a key role in organizing the genome (Phillips-Cremins and Corces 2013). Its ability to alter chromosome conformation is mediated via interaction with cohesin, a protein that is known to form multimolecular complexes. CTCF binding is widespread across the *IgH* locus (Degner et al. 2009), and *IgH* locus compaction is diminished in CTCF knockdown pro-B cells (Degner et al. 2011). However, lack of CTCF-binding sites

Present addresses: <sup>3</sup>Cellular and Molecular Medicine Program, Boston Children's Hospital, Department of Pediatrics, Harvard Medical School, Boston, MA, 02115, USA; <sup>4</sup>Program in Genetics, University of Chicago, Chicago, IL 60637, USA

Corresponding author: [senra@mail.nih.gov](mailto:senra@mail.nih.gov)

Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.263871.115>.

© 2015 Gerasimova et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genesdev.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.



**Figure 1.** CTCF nucleates 300- to 400-kb B-lymphocyte lineage-specific loops that are E $\mu$ -independent. (A) Schematic representation of the unrearranged *IgH* locus showing multiple V<sub>H</sub>, D<sub>H</sub>, and J<sub>H</sub> gene segments. DNA recombination juxtaposes a V<sub>H</sub>, D<sub>H</sub>, and J<sub>H</sub> gene segment to generate VDJ rearranged alleles that can express antibody heavy chain proteins. Positions of relevant regulatory sequences are indicated. (E $\mu$ ) A tissue-specific transcriptional enhancer in the J<sub>H</sub>-C $\mu$  intron; (IGCR1) intergenic control region 1 that comprises a cluster of DNase I-hypersensitive sites with two binding sites for the transcription factor CTCF; (3'RR) the 3' regulatory region located at the 3' end of the *IgH* locus that also comprises a cluster of DNase I-hypersensitive sites that have CTCF-binding sites. The lower line shows a scale view of the *IgH* locus. J558/3609 refers to families of V<sub>H</sub> gene segments. Members of V<sub>H</sub>558 and V<sub>H</sub>3609 gene families are interspersed throughout this region. S107 and 7183 refer to smaller V<sub>H</sub> gene families. The eight to 12 D<sub>H</sub> gene segments and four J<sub>H</sub> gene segments are located within the region marked D-J. Exons for eight classes of antibody molecules are located within the region marked C. The distribution of 14 Pax5-activated intergenic repeat (PAIR) sequences (Ebert et al. 2011) is indicated by gray ovals (P1-P14). The locations of probes used for fluorescent in situ hybridization (FISH) studies are indicated by black lines. BAC RP23-201H14 was used to mark *IgH* alleles and is located ~44 kb 3' of HS7 of the 3'RR. (B) CTCF expression was attenuated (labeled CTCF KD) by shRNA expression (Degner et al. 2011) in primary pro-B cells from RAG1-deficient mice cultured on OP9 stromal cells; control cells expressed a scrambled shRNA in the same cells. Short probes are color-coded as shown at the left of representative nuclei from FISH experiments; BAC RP23-201H14 is shown in blue. Distances between probes were measured after deconvolution of images, and the average separation was calculated from 88–110 nuclei (Supplemental Table 1). Bar graphs show the percentages of *IgH* alleles in which the average distances between the indicated probe pairs fell in the ranges shown by different colors. The cumulative frequency distributions of spatial distance measurements for each color-coded probe combination are shown below the bar graph. D statistics and P-values for differences between control and knockdown cells were calculated using a two-sample Kolmogorov-Smirnov test and are shown in Supplemental Table 1. (C) Non-B cells were purified from the bone marrow of RAG2-deficient mice by removing pro-B cells using anti-CD19-coupled magnetic beads. FISH was carried out using the indicated small probes from the V<sub>H</sub> and D<sub>H</sub> regions and BAC RP23-201H14 (shown in blue). Distance measurements and quantitation were as in B using 91–108 nuclei.

within E $\mu$  and the strong effect of E $\mu$  deletion on *IgH* locus compaction suggest a more complex mechanism. Busslinger and colleagues (Ebert et al. 2011; Medvedovic et al. 2013) recently proposed a model in which *IgH* locus compaction is driven by direct interactions between Pax5 (which binds to a cluster of Pax5-activated intergenic repeat [PAIR] elements dispersed through the 5' V<sub>H</sub> region) (Fig. 1A) and CTCF bound throughout the locus. The role of E $\mu$  and the transcription factor YY1 in this model is not clear. One possibility could be that YY1

binds to a subset of PAIR elements and regulates antisense transcripts at PAIR4, PAIR6, and PAIR8 (Verma-Gaur et al. 2012). We had previously proposed a two-step model for generating *IgH* locus conformation (Guo et al. 2011a). The first step, which is E $\mu$ -independent, generates multiple 250- to 400-kb subdomains in the V<sub>H</sub> region. Because we identified these domains using anti-CTCF chromatin immunoprecipitation (ChIP) loop assays, we proposed that they would be CTCF-dependent. The second step involves E $\mu$ -dependent interactions with distant sites in the

$V_H$  region that juxtapose subdomains in the  $V_H$  part of the locus with the 3' end of the *IgH* locus. The major loop sub-compartment (MLS) structure derived by Murre and colleagues (Jhunjhunwala et al. 2008) in E2A-deficient pro-B cells, where  $E_{\mu}$  is inactive, likely represents a locus that has undergone only the first step of locus compaction. Because  $E_{\mu}$  contains a YY1-binding site, we proposed that  $E_{\mu}$ -dependent effects could be mediated by this transcription factor. In light of these contrasting models, it is imperative to clarify the molecular mechanisms by which  $E_{\mu}$ , YY1, CTCF, and Pax5 coordinately configure the functional structure of the prearrangement *IgH* locus in pro-B cells.

Here we provide a unifying model that defines a structural hierarchy by which these transcription factors and  $E_{\mu}$  establish *IgH* locus conformation. We demonstrate that integrity of  $E_{\mu}$ -dependent loops requires YY1 and uses the condensin components Smc2 and Smc4. In contrast, subdomains within the  $V_H$  region are CTCF-dependent but YY1-independent. Furthermore, these CTCF-dependent subdomains are B-lineage-specific and Pax5-independent. In addition, CTCF helps to configure the 3' 280 kb of the *IgH* locus; however, the integrity of this domain also requires  $E_{\mu}$  and YY1. These observations highlight combinatorial mechanisms by which the *IgH* locus structure is established via three levels of compaction mediated by three different transcription factors and lead to a novel model for the *IgH* locus. CTCF- and Pax5-dependent interactions compact the  $V_H$  region, working in domains of a few hundred kilobases and over a megabase, respectively. The compacted  $V_H$  region is brought into proximity of the 3' end of the locus in a step that requires YY1 and  $E_{\mu}$ . We suggest that such stepwise generation of chromosome conformation may apply more generally in folding megabase-sized chunks of the genome.

## Results

### Configuring the 5' *IgH* ( $V_H$ ) region

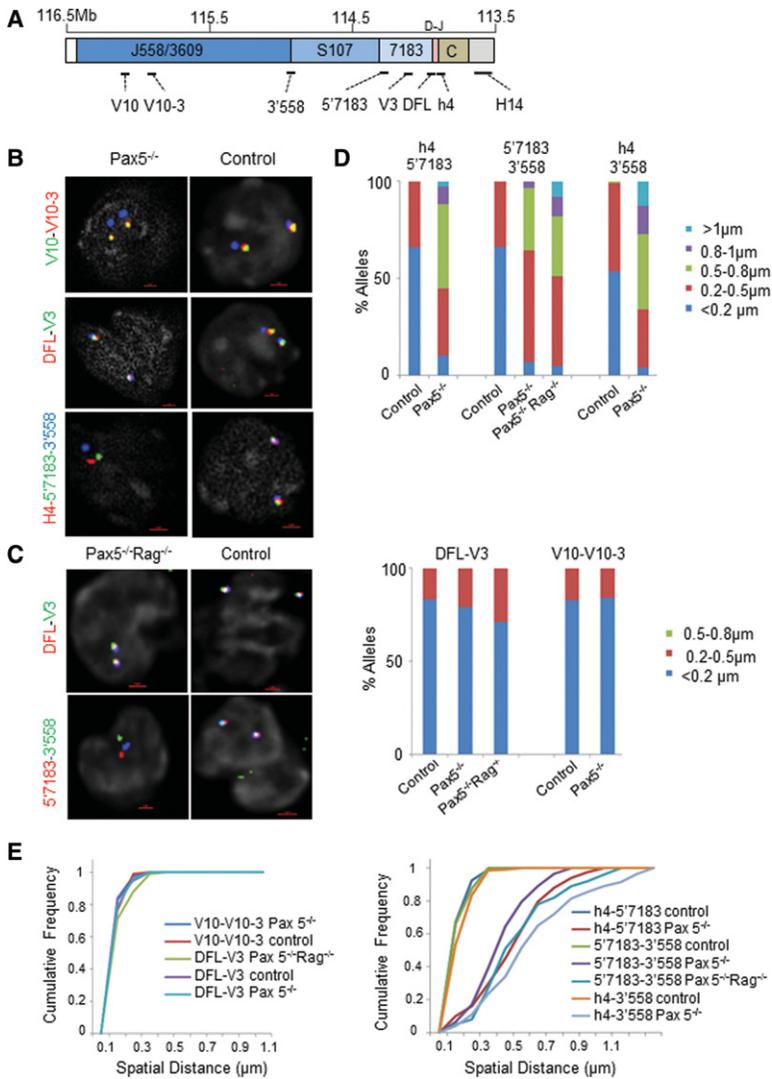
The chromatin-organizing factor CTCF binds at multiple sites throughout the  $V_H$  region, and *IgH* locus compaction is reduced in CTCF knockdown pro-B cells (Degner et al. 2011). In our model of the *IgH* locus, CTCF is involved in generating  $V_H$  region subdomains that are  $E_{\mu}$ -independent. Using anti-CTCF ChIP-loop, we previously identified two sets of interactions in the  $V_H$  part of the *IgH* locus. One 288-kb domain (V10–V10-3) (Fig. 1A) lies at the 5' end of the  $V_H$  locus, and the second 250-kb domain (DFL–V3) lies at the 3' end. These interactions are  $E_{\mu}$ -independent, and we proposed that they would be CTCF-dependent. To test this hypothesis, we used primary pro-B cell cultures in which CTCF expression was attenuated by appropriate shRNA expression. These conditions led to 75% reduction in the level of CTCF mRNA and protein (Degner et al. 2011). We found that both previously identified  $E_{\mu}$ -independent loops were disrupted in CTCF knockdown cells (Fig. 1B; Supplemental Table 1). The site labeled DFL corresponds closely with three DNase1-hypersensitive sites that comprise the intergenic control

region (IGCR1). Guo et al. (2011b) showed that mutating two CTCF-binding sites within IGCR1 reduced looping of this site to the 3' regulatory region (3'RR) located at the 3' end of the *IgH* locus and accentuated proximal  $V_H$  rearrangement to DQ52. We found that DFL–3'RR interactions were disrupted in CTCF knockdown cells, thereby confirming the importance of CTCF in configuring the structure of this end of the *IgH* locus (Fig. 1B). The three CTCF-dependent chromatin loops ranged in size between 250 and 400 kb.

As a control, we analyzed a 400-kb  $E_{\mu}$ -dependent interaction between 5'7183 and  $E_{\mu}$  (probe h4). Interaction between these two sites was minimally reduced in CTCF knockdown cells (Fig. 1B). To quantitate looping disruptions caused by CTCF deficiency, we determined the extent of decompaction between specific sites as the ratio of the compaction values (Jhunjhunwala et al. 2008) in CTCF-sufficient and CTCF-deficient cells. For both  $E_{\mu}$ -independent loops, the decompaction in CTCF-deficient cells was ~1.9-fold, whereas for the  $E_{\mu}$ -dependent h4–5'7183 loop, the value was 1.3 (Supplemental Table 1). Maximal disruption of h4 and 5'7183 interaction on  $E_{\mu}$ -deficient alleles resulted in 3.8-fold decompaction (Guo et al. 2011a). We infer that CTCF knockdown partially disrupted  $E_{\mu}$ -dependent looping. Because there are many CTCF-binding sites within  $E_{\mu}$ -dependent loops, our interpretation is that the partial disruption resulted from attenuation of CTCF-dependent interactions within the  $E_{\mu}$ –5'7183 loop.

Degner et al. (2009) previously showed that CTCF bound to *IgH* in pro-B cells but not in mouse embryo fibroblasts, indicating that CTCF binding to the *IgH* locus was lineage-specific despite the factor being ubiquitously expressed. These observations predicted that CTCF-dependent  $V_H$  locus compaction would be lymphoid-specific. To test this, we assayed CTCF-dependent interactions in primary non-B-lineage cells isolated from the bone marrow of RAG2-deficient mice. We found that both V10–V10-3 and DFL–V3 interactions were absent in non-B-lineage cells (Fig. 1C). Quantitation of fluorescent in situ hybridization (FISH) data showed that spatial distances between pairs of probes were comparable between CTCF knockdown and non-B-lineage cells (Supplemental Table 1).

The first transcription factor shown to be involved in *IgH* locus compaction was Pax5 (Fuxa et al. 2004). In recent studies, Medvedovic et al. (2013) proposed that Pax5-dependent locus compaction occurred via direct interactions between Pax5 and CTCF. Therefore, it was of interest to examine the role of Pax5 in establishing CTCF-dependent subdomains. For this, we expanded Pax5-deficient pro-B cells on stromal cell cultures (Pongubala et al. 2008) and assayed V10–V10-3 and DFL–V3 interactions by FISH; as a control, we used Pax5-sufficient pro-B cells from RAG2-deficient mice that were expanded similarly. Both of these CTCF-dependent,  $E_{\mu}$ -independent interactions were intact in Pax5<sup>-/-</sup> pro-B cells (Fig. 2B,D,E; Supplemental Table 2). Because these cells were recombinase-sufficient and therefore likely to have partial *IgH* rearrangements, we also assayed a subset of interactions in Pax5- and RAG2-double-deficient



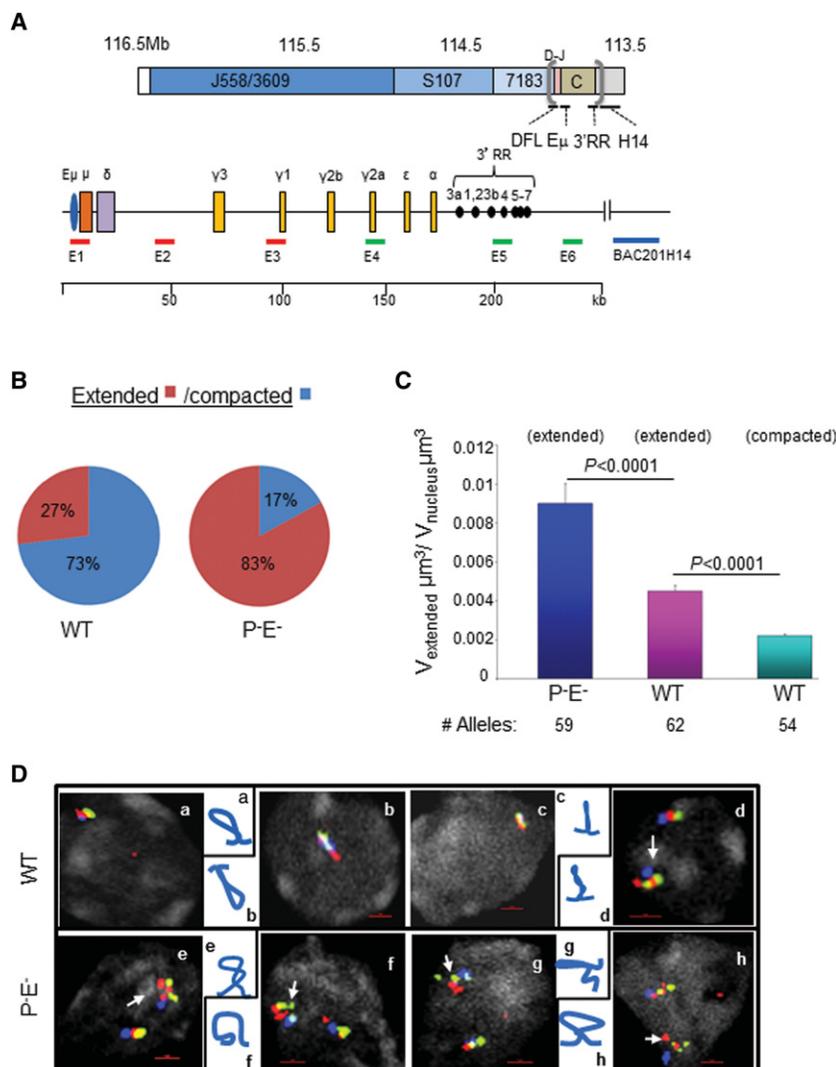
**Figure 2.** CTCF-mediated compaction in Pax5-deficient pro-B cells. (A) *IgH* locus schematic indicating the locations of short probes and BAC RP23-201H14. (B) Pax5-deficient pro-B cells (*Pax5*<sup>-/-</sup>) were expanded on S17 stromal cell cultures for FISH; control pro-B cells were obtained from RAG2-deficient bone marrow and expanded on OP9 stromal cells. FISH was carried out with the probe combinations indicated at the left of representative nuclei; BAC RP23-201H14 is shown in blue in the top two panels. (C) RAG2-deficient, Pax5-deficient pro-B cells (*Pax5*<sup>-/-</sup>*RAG*<sup>-/-</sup>) were expanded on S17 stromal cell cultures for FISH with the indicated probes; control pro-B cells were obtained from RAG2-deficient bone marrow and expanded on OP9 stroma. (D) Distances between probes were measured after image deconvolution, and the average spatial distance was calculated from 118–121 alleles. The bar graph shows the percentage of *IgH* alleles in which the average distance between the indicated probe pairs fell in the ranges shown by the different colors. Pro-B-cell genotypes are indicated below the bars. (E) Cumulative frequency distributions of spatial distance measurements for each color-coded probe combination in each genotype. D statistics and *P*-values were calculated as described in the legend for Figure 1 and are shown in Supplemental Table 2.

(*Pax5*<sup>-/-</sup>*RAG*<sup>-/-</sup>) pro-B cells in which *IgH* alleles would be in germline configuration. DFL-V3 interactions were unaffected in *Pax5*<sup>-/-</sup>*RAG*<sup>-/-</sup> pro-B cells (Fig. 2C–E). In contrast, association between probes located within distal (3'558) and proximal (5'7183) V<sub>H</sub> genes was disrupted in both *Pax5*<sup>-/-</sup> and *Pax5*<sup>-/-</sup>*RAG*<sup>-/-</sup> pro-B cells (Fig. 2B,C; Supplemental Table 2). Although Pax5 does not bind E<sub>μ</sub>, interactions between E<sub>μ</sub> and V<sub>H</sub> sites were also reduced in *Pax5*<sup>-/-</sup> pro-B cells. The basis for this remains unclear, and possible models are considered in the Discussion. We conclude that lymphoid-specific access of CTCF to the *IgH* locus leads to the establishment of CTCF-dependent subdomains that are independent of E<sub>μ</sub>, YY1 (see below), and Pax5. An additional level of compaction of the V<sub>H</sub> region is conferred by Pax5 that juxtaposes distal and proximal V<sub>H</sub> gene segments.

*Configuring the 3' IgH domain*

The 280-kb 3' *IgH* domain extends from the IGCR1, located 5' of DFLI6.1 (Fig. 1A), to the 3'RR. Chromosome con-

formation capture (3C) studies show that IGCR1, E<sub>μ</sub>, and the 3'RR are spatially proximal in pro-B cells, leading us to propose a three-loop configuration for this part of the *IgH* locus (Degner et al. 2011; Guo et al. 2011a). However, 3C techniques that read out proximity between sequences as a population average leave open questions of the frequency and variability of chromatin loops in individual cells. Although this is somewhat ameliorated by FISH, which permits assessment of the association frequency between two genomic regions in single cells, even this technique does not reveal the structures of putative loops or the path of the chromatin fiber within loops. To directly visualize chromosome loops in single nuclei, we developed a multiprobe FISH technique and used it to study the largest proposed loop of the 3' *IgH* domain (250 kb) that extends from the E<sub>μ</sub> to the 3'RR. To visualize the E<sub>μ</sub>-3'RR loop, we generated six short FISH probes (E1–E6) that were separated by 30–50 kb (Fig. 3A). Probes E1–E3, which are close to E<sub>μ</sub>, were labeled with red dye, and probes E4–E6 in the 3' half were labeled with green dye. All six probes as well as a BAC located 44 kb from E6 (RP23-201H14) were



**Figure 3.** Barcoding the 3' *IgH* domain with multiple short FISH probes. (A) Schematic of the *IgH* locus with the 3' *IgH* domain expanded to show the location of the FISH probes used for FISH. BAC RP23-201H14 used to mark the *IgH* alleles is located ~44 kb 3' of HS7 of the 3'RR. (B) Probes E1–E3 labeled with Alexa fluor 594 (red) and probes E4–E6 labeled with Alexa fluor 488 (green) were hybridized simultaneously to bone marrow pro-B cells containing wild-type (WT) *IgH* alleles or alleles that lacked 700-base-pair (bp) E $\mu$  and a promoter close to DQ52 (E<sup>-</sup>P<sup>-</sup>) (Afshar et al. 2006). Previous studies showed that the DQ52 promoter does not contribute to the locus conformation or epigenetic states of *IgH* alleles (Guo et al. 2011a). Alleles on which three or more of the six probes could be distinguished were categorized as “extended,” and the rest were categorized as “compacted.” Pie charts show the proportion of extended and compacted *IgH* alleles in pro-B cells of the indicated *IgH* genotypes; all pro-B cells were RAG2-deficient to maintain *IgH* in an unrearranged configuration. Data were obtained from three dimensional (3D) deconvolution of 100–120 alleles. (C) The volume occupied by *IgH* alleles was estimated using Nikon NIS software as described in the Materials and Methods. The proportion of the total nuclear volume occupied by *IgH* alleles of the indicated genotypes is shown. The number of alleles used for quantitation is shown below the graph; *P*-values were calculated based on two-tailed *t*-test in Microsoft Excel. (D) Representative views of four wild-type and P-E<sup>-</sup> alleles to exemplify the diversity of loop conformations visualized by multiprobe bar coding. One section is shown for each nucleus. The inset represents proposed loop conformations based on the analysis of 30–40 Z sections visualized for each nucleus (Supplemental Fig. 1).

simultaneously hybridized to bone marrow pro-B cells of defined *IgH* genotypes. We refer to this new procedure as bar-coding chromatin loops.

We compared the structures of wild-type and P-E<sup>-</sup> *IgH* alleles in a RAG2-deficient background. P-E<sup>-</sup> alleles lack the DQ52 promoter and E $\mu$ . Based on previous studies that showed no effect of DQ52 promoter deletion on the epigenetic state of *IgH* (Guo et al. 2011a), we infer that differences between the two genotypes reflect loss of E $\mu$ . We quantified the differences between wild-type and E $\mu$ -deleted *IgH* alleles in two ways. First, we determined the proportion of alleles in which three or more probes could be visually resolved on at least one *IgH* allele; we refer to these alleles as being in an “extended” configuration. We found that only 27% of wild-type *IgH* alleles were extended in primary pro-B cells; in contrast, 83% of P-E<sup>-</sup> *IgH* alleles were in an extended configuration. Second, we estimated the nuclear volume occupied by *IgH* alleles as described in the Materials and Methods. We found that the volume of extended wild-type *IgH* alleles was significantly different from compacted wild-type alleles (Fig.

3C). Importantly, the volume occupied by extended E $\mu$ -deficient alleles was almost twofold greater than wild-type extended alleles (Fig. 3C). Taken together, these observations demonstrate that the E $\mu$  determines the spatial volume of the 3' 250 kb of the *IgH* locus.

To understand the basis for the increased volume of P-E<sup>-</sup> compared with wild-type alleles, we examined the “fine structure” of extended *IgH* alleles of both genotypes. Analyses of Z sections of wild-type alleles (Supplemental Fig. 1A) revealed looped structures with different conformations in different nuclei. Many wild-type alleles showed a yellow dot (reflecting colocalization of at least one red and one green probe and several distinct red and green signals) (Fig. 3D, panels a–d). Based on the composite pattern from individual Z sections, we infer these to be classical loops, with the yellow representing the base of the loop and little or no overlap of the other probes (Fig. 3D, insets a and b). In other nuclei, we found more than one yellow dot, indicating marked differences in chromosome conformation in individual cells. The additional fine structure revealed by more than one yellow dot could

represent loops in which the stem is further twisted (as schematized in the insets c and d of Figure 3D) or loops in which the arms touch at one or more points without interwrapping. Remarkably, we rarely saw yellow dots with  $E^-P^-$  alleles; instead, we usually resolved several red and green signals. Based on multiple Z sections (Supplemental Fig. 1B), we interpret this to indicate that the base of the loop is disrupted in  $E^-P^-$  alleles (Fig. 3D, insets e–h), thereby precluding additional loop conformations.

We used a similar set of six short FISH probes that hybridized approximately every 50 kb (Supplemental Fig. 2A) in the 3' *IgH* domain to study the structure of the locus in normal and CTCF-depleted pro-B cells. Thirty-four percent of alleles were in extended configuration in control pro-B cell nuclei; in contrast, 80% of alleles were in extended configuration in CTCF-depleted pro-B cells (Supplemental Fig. 2B). We also estimated the nuclear volume occupied by the domain in normal and CTCF-depleted pro-B cells. We found that the 3' *IgH* domain occupied a larger nuclear volume in CTCF-depleted pro-B cells (Supplemental Fig. 2C). The conformational similarity between  $E\mu$ -deficient alleles and wild-type alleles in CTCF-depleted cells suggests a shared mechanism that establishes the 3' *IgH* domain.

#### Bridging the 5' ( $V_H$ ) and 3' *IgH* domains

We previously identified two sites in the  $V_H$  region of the unrearranged *IgH* locus that interacted with  $E\mu$  (in the 3' *IgH* domain). One site, denoted as 5'7183, is located ~400 kb from  $E\mu$ , and the other, denoted as 3'558, is located ~1.2 Mb from  $E\mu$  (Fig. 4A). We proposed that these interactions bring  $V_H$  gene segments into spatial proximity of the  $D_H/J_H$  part of this locus for  $V_H$  recombination. However, factors driving this association remain unclear. To test the hypothesis that  $E\mu$ -dependent loops were configured by YY1, we depleted YY1 in a RAG2-deficient pro-B-cell line, D345, via shRNA expression (Supplemental Fig. 3A,B) and assayed *IgH* locus conformation by FISH and 3C assays. We found that interaction of 5'7183 with the  $D_H-C\mu$  part of the locus that contains  $E\mu$  as well as spatial colocalization of 5'7183 and 3'558 probes (a distance of ~800 kb) were significantly reduced in YY1 knockdown cells (Fig. 4B; Supplemental Table 3). In contrast, loops previously classified as  $E\mu$ -independent, such as V10–V10-3 and DFL–V3, were not affected in YY1 knockdown cells (Supplemental Fig. 3C). Disruption of these interactions was also confirmed by quantitative (3C) assays in YY1 knockdown cells (Supplemental Fig. 3D,E).

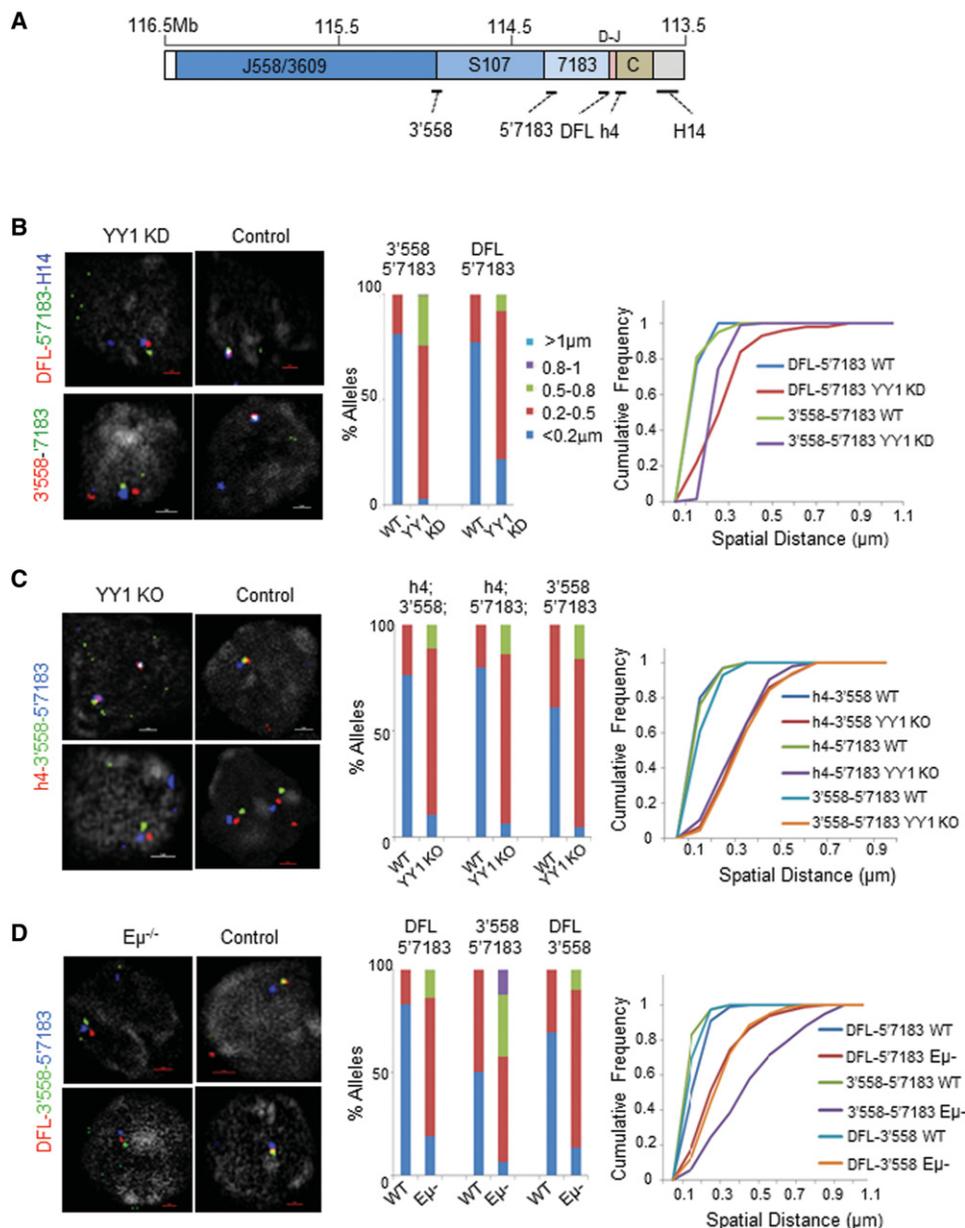
We further interrogated these interactions in YY1-deficient primary pro-B cells. To focus on the effects of YY1 in establishing the structure of the unrearranged *IgH* locus, we carried out these studies in VDJ recombinase-deficient cells that lacked YY1. Freshly isolated bone marrow pro-B cells from YY1<sup>fl/fl</sup> × mb1-cre × RAG1<sup>-/-</sup> mice were used for FISH without further expansion in vitro. We found that spatial proximity of  $E\mu$  (probe h4) to both 3'558 and 5'7183 probes was substantially altered in YY1-deficient pro-B cells compared with control pro-B cells (YY1<sup>fl/+</sup> mb1-cre × RAG1<sup>-/-</sup>) (Fig. 4C; Supplemental Table 3).

YY1 deficiency also altered the proximity between 3'558 and 5'7183 probes (Fig. 4C; Supplemental Table 3). The structure of  $E\mu$ -deleted *IgH* alleles in pro-B cells resembled that in YY1-deficient pro-B cells insofar as revealing reduced interactions between  $V_H$  region probes (5'7183 and 3'558) and a probe within the 3' *IgH* domain (DFL) (Fig. 4D). However, 3'558–5'7183 interactions were more severely reduced on  $E\mu$ -deleted alleles compared with YY1 knockout pro-B cells. One possibility for the difference may be that these sites are brought together on wild-type alleles by participating in a three-way interaction that includes  $E\mu$ . In the absence of  $E\mu$ , the three-way interaction would be disrupted, thereby spacing apart 5'7183 and 3'558. However, each of these sites could independently maintain some proximity to the 3' *IgH* domain due to residual YY1-dependent locus compaction. Alternatively, the differences could be due to the use of untransformed YY1 knockout pro-B cells and an Abelson virus transformed cell line with  $E\mu$ -deficient *IgH* alleles. We conclude that the  $E\mu$ -dependent compaction of *IgH* alleles requires YY1.

YY1-dependent looping has been proposed to be mediated via interactions with components of the condensin complex (Pan et al. 2013). To determine whether YY1-binding sites within *IgH* interacted with condensin components, we carried out ChIP with anti-SMC2 and anti-SMC4 antibodies. We detected both SMC2 and SMC4 at all prominent YY1-binding sites within the 3' *IgH* domain, including  $E\mu$  and 5'DFL16.1, and within HS5–7 of the 3'RR (Supplemental Fig. 4). YY1 binding was much lower and SMC2/SMC4 was barely discernible at 3'558 and 5'7183 compared with any of the sites in the 3' *IgH* domain. The residual binding could reflect cross-linking of  $E\mu$ - or DFL16-bound YY1/SMC2/4 to these distal sites due to looping-induced spatial proximity. To determine whether SMC proteins played a functional role in *IgH* locus conformation, we reduced SMC2 expression by retroviral shRNA expression in D345 pro-B cells (Supplemental Fig. 5A,B) and assayed  $E\mu$ -dependent interactions by quantitative 3C assays. We found that 3'558 loops to 5'7183,  $E\mu$ , and the 3'RR were significantly reduced, whereas CTCF-dependent V10–V10-3 interactions were unaffected, in SMC2 knockdown cells (Supplemental Fig. 5D). Disruption of these loops in both YY1 knockdown and SMC2 knockdown pro-B cells suggested that YY1-bound chromosomal regions were brought together by condensin-mediated interactions. We propose that spatial juxtaposition of 5' ( $V_H$ ) *IgH* and 3' *IgH* domains occurs via  $E\mu$ - and YY1-dependent interactions that involve SMC2.

#### Interchromosomal interactions mediated by CTCF

Recent studies provide genome-wide evidence that the frequency of double-strand breaks and spatial proximity of sequences underpin the likelihood of interchromosomal translocations (Zhang et al. 2012). Both parameters have been associated with *IgH/c-Myc* translocations that are prevalent in plasmacytomas (tumors of antibody-secreting plasma cells). DNA breaks are introduced by the enzyme activation-induced deaminase (AID) (Chiarle et al.



**Figure 4.** YY1 deficiency disrupts  $E\mu$ -dependent chromatin loops in the *IgH* locus. (A) *IgH* locus schematic showing the positions of the short FISH probes used; BAC RP23-201H14 was used to mark *IgH* alleles. (B) YY1 expression was reduced by shRNA expression in D345 pro-B cells (YY1 knockdown [KD]) (Supplemental Fig. 3) followed by FISH using the indicated probes; cells infected with an empty vector were used as controls. D345 is a recombinase-deficient Abelson virus transformed pro-B-cell line. Distances between FISH probes were measured after image deconvolution from 80–120 nuclei as described in the Materials and Methods (Supplemental Table 3). (Middle) Bar graphs show the percentages of *IgH* alleles in which the average distances between the indicated probe pairs fell in the ranges shown by different colors. Cumulative frequency distributions of spatial distance measurements for each color-coded probe combination are shown at the right. D statistics and *P*-values for differences between control and YY1 knockdown cells were calculated as described in the legend for Figure 1 and are shown in Supplemental Table 3. (C) FISH analysis using the indicated probes was carried out in YY1-deficient or control pro-B cells isolated from the bone marrow of YY1<sup>fl/+</sup> × mb-1-cre × RAG1<sup>-/-</sup> or YY1<sup>fl/+</sup> × mb-1-cre × RAG1<sup>-/-</sup> mice, respectively, and used without further ex vivo culture. Probes were color-coded as indicated at the left of the representative nucleus shown, and distances between probe pairs were measured in 75–125 nuclei (Supplemental Table 3). (Middle and right) Data quantitation and statistics are as in B. (D) FISH analysis using the indicated probes in RAG2-deficient pro-B-cell lines with normal (control) or  $E\mu$ -deficient *IgH* alleles. Data quantitation and statistics are as in B, based on measurements of 87–92 nuclei.  $E\mu^{-/-}$  cells are recombinase-deficient Abelson virus transformed pro-B-cell lines whose *IgH* alleles lack 220 bp of the core  $E\mu$  enhancer (Perlot et al. 2005).

2011; Klein et al. 2011), and the two loci undergo inducible juxtaposition in the nuclei of activated mature B cells (Roix et al. 2003; Osborne et al. 2004). Hi-C (chromosome capture followed by high-throughput sequencing) studies in human B lymphoblastoid cells also demonstrated increased interaction between *IgH* and *c-Myc* loci (located on human chromosomes 14 and 8, respectively) (Lieberman-Aiden et al. 2009; Engreitz et al. 2012). However, factors involved in interchromosomal *IgH/c-Myc* interactions have not been identified.

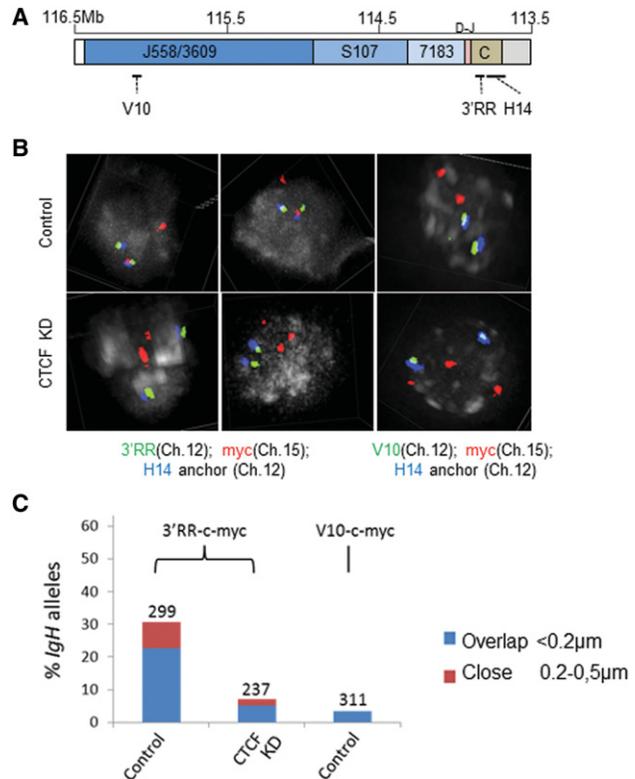
Translocations between *IgH* and *c-Myc* loci also occur in pro-B cells, where they are initiated by DNA breaks induced during V(D)J recombination (Difilippantonio et al. 2002; Zhu et al. 2002; Tsai et al. 2008; Gostissa et al. 2009). Recent studies found that AID activity can also be induced in pro-B cells, providing another source of translocation-initiating DNA double-strand breaks (Kumar et al. 2013). However, Hi-C studies in pro-B cells have not revealed selective association of these two loci in pro-B cells (Lin et al. 2012; Zhang et al. 2012). To determine whether *IgH* and *c-Myc* colocalize in primary pro-B cells, we used FISH probes specific to the 5' and 3' ends of the *IgH* locus (Fig. 5A) as well as one that overlaps *c-Myc*-coding exons. We found that the 3'RR probe was close to or overlapped with the *c-Myc* probe in ~30% of control pro-B cells; in contrast, a probe in the 5' V<sub>H</sub> region labeled V10 was close to *c-Myc* in <5% of pro-B cells (Fig. 5B,C; Supplemental Table 4). These observations demonstrate that the two loci interact and that the 3' end of the *IgH* locus serves as the major site of interaction. Because the 3' end of the *IgH* locus has more prominent CTCF binding compared with the 5' end (Degner et al. 2011) and because CTCF binding to *c-Myc* is well-established (Gombert et al. 2003; Ishihara et al. 2006), we hypothesized that CTCF could be involved in this interaction. To determine the role of CTCF, we carried out FISH in CTCF-depleted pro-B cells. We found that 3'RR-*c-Myc* association was reduced to the level of V10-*c-Myc* association in the absence of CTCF (Fig. 5B,C; Supplemental Table 4). Thus, CTCF is essential for *IgH/c-Myc* association in primary pro-B cells. We propose that B-lineage-specific binding of CTCF to the *IgH* locus increases the propensity of interchromosomal interactions between *IgH* and *c-Myc* and thereby the potential for hazardous translocations.

## Discussion

Our observations reveal several levels of chromosomal compaction that fold the 2.8-Mb *IgH* locus into a conformation that optimizes tissue-specific gene recombination and expression.

### Level 1

First, the locus folds into domains a few hundred kilobases in size mediated by CTCF. We provide three examples of such domains. One lies entirely within the V<sub>H</sub> region (V10-V10-3), a second bridges the proximal V<sub>H</sub> genes to the D<sub>H</sub>-C<sub>μ</sub> part of the locus (V3-DFL), and the third en-



**Figure 5.** *IgH-c-Myc* interactions mediated by CTCF. (A) Schematic of the *IgH* locus with the locations of probes at its 5' and 3' ends. The 5-kb *c-Myc* probe encompassed most of the *c-Myc* gene on mouse chromosome 15. (B) Primary pro-B cells from RAG1-deficient mice expressing either CTCF shRNA (CTCF knockdown [CTCF-KD]) or scrambled shRNA (control) as described in Figure 2 were hybridized to either V10 or 3'RR (green) FISH probes and the *c-Myc* probe (red); BAC RP23-201H14 (blue) was used as an *IgH* marker. Representative nuclei are shown. (C) Distances between *IgH* and *c-Myc* probes were measured after image deconvolution, and the average distances were calculated based on measurements of 237–299 nuclei as indicated. The bar graph shows the percentage of *IgH* alleles in which signals between *IgH* and *c-Myc* overlap (separation of <0.2 μm) or were close (separation between 0.2 and 0.3 μm).

compasses the 3'-most part of the *IgH* locus (DFL-3'RR). Each domain provides insights into the structural organization of the locus.

The V10-V10-3 loop was first identified by anti-CTCF 4C (circular 3C) assays using a prominent CTCF-binding site as the bait. The spatial proximity of FISH probes corresponding to these sites led us to infer that they came together at the base of a loop that defined a chromatin domain within the V<sub>H</sub> region. We do not intend to imply that this pairing is strict, in the sense that V10 can only interact with V10-3. Rather, our interpretation is that V10 and V10-3 are prominent, but not exclusive, sites of interaction. In other words, these sites may also engage with alternative CTCF-bound sites within a population of cells. Such variations could represent the dynamic state of chromatin loops within a cell as well as cell-to-cell variations within a population. Because the V<sub>H</sub> region contains >100

CTCF-bound sites (Choi et al. 2013), we expect that there are other similarly sized CTCF-dependent domains within the  $V_H$  region in addition to V10–V10-3. We propose that such CTCF-dependent domains provide the first level of compaction to the *IgH* locus and, like V10–V10-3 and V3–DFL, would be  $E\mu$ -, Pax5-, and YY1-independent. We hypothesize that such a partially folded pro-B-cell-specific structure serves as the substrate for subsequent steps of locus compaction (see below).

It is interesting to note that these domains form only in pro-B cells but not in nonlymphoid hematopoietic cells, although they are generated by a ubiquitously expressed protein. This is entirely consistent with lymphoid-specific binding of CTCF to the  $V_H$  region and indicates that a level of B-lineage-specific locus activation precedes the first level of chromatin compaction. The nature of this key developmental step remains unknown. We speculate that it may be akin to the state of the murine  $\beta$  globin locus that lacks the locus control region yet maintains several features of erythroid-specific locus activation (Schubeler et al. 2001; Sawado et al. 2003).

Characteristics of the DFL–3'RR loop, which encompasses the 3' end of the *IgH* locus, reveal additional features of structural organization. Our high-resolution barcode FISH analysis of this domain showed considerable conformational heterogeneity of the chromatin path on a cell-to-cell basis. Among extended wild-type alleles, which represent less than a third of the total, we noted simple loops (two contact points that serve as the base of the loop) as well as loops with additional twists or loops with additional points of contact away from the base. We speculate that higher numbers of additional twists or contacts preclude visualization of more than three probes on compacted alleles. Because compacted alleles represent >70% of the total, we infer that the majority of wild-type alleles are not folded into simple loops. To the best of our knowledge, this is the first demonstration of the conformational state of chromatin loops in mammalian interphase nuclei. Like V10–V10-3 and V3–DFL loops, the DFL–3'RR loop requires CTCF. However, despite being sized similarly to V10–V10-3 or V3–DFL loops, the integrity of the DFL–3'RR domain also requires YY1 (Medvedovic et al. 2013) and the *cis*-regulatory element  $E\mu$  (Guo et al. 2011a). The involvement of YY1 and  $E\mu$  likely stems from prominent YY1-binding sites close to DFL, at  $E\mu$ , and within the 3'RR (Guo et al. 2011a). However, this conjecture needs to be directly verified by mutating YY1-binding sites at these locations. It remains unclear why the DFL–3'RR loop has such different structural requirements from the  $V_H$  region V10–V10-3 and V3–DFL loops. One possibility is that the DFL–3'RR chromatin domain serves very different functions from the CTCF-dependent  $V_H$  domains. For example, this domain has the highest levels of activating histone modifications in the *IgH* locus, sense and antisense transcripts initiate here, and the RAG1/2-rich recombination center forms within this domain. In contrast the V10–V10-3 and V3–DFL domains are mostly heterochromatic, lack significant transcriptional activity, and may therefore be largely structural.

### Level 2

The second level of locus compaction generates megabase-sized domains in the  $V_H$  region using the transcription factor Pax5. We infer this based on the observation that FISH probes located within 5' and 3'  $V_H$  gene families (labeled 3'558 and 5'7183) were spatially disassociated in Pax5-deficient pro-B cells, whereas CTCF-dependent loops remained intact. Based on our current analysis, it remains unclear whether CTCF-dependent compaction is a prerequisite for Pax5-dependent compaction. Preservation of CTCF-dependent loops in Pax5-deficient pro-B cells provides a plausible explanation for a puzzling feature of  $V_H$  gene recombination in Pax5-deficient pro-B cells. It has been known for some time that proximal  $V_H$  gene recombination (7183 family) is normal in Pax5<sup>-/-</sup> pro-B cells, whereas distal  $V_H$  gene (J558 family) recombination is profoundly inhibited (Hesslein et al. 2003). We discovered that a CTCF-dependent loop (V3–DFL), which is unaffected by Pax5 deficiency, places a subset of proximal  $V_H$  genes close to the  $D_H$ - $C\mu$  region. This interaction may permit proximal  $V_H$  genes to recombine in Pax5<sup>-/-</sup> pro-B cells.

The structure of a Pax5-compacted  $V_H$  region and the mechanism by which it is generated remain unclear. Medvedovic et al. (2013) proposed that this compaction occurs via direct interactions between Pax5 bound to PAIR sequences and CTCF bound throughout the locus. At a mechanistic level, this model did not distinguish between a structure that is generated by both proteins acting coordinately and one that is generated by each protein acting independently via complementary partial structures. The former mechanism predicts that all  $V_H$  region structure will be lost in the absence of either protein, while the second model predicts that the residual structure will be retained in the absence of one or the other protein. Our observations revealed that the B-lineage-specific structure mediated by CTCF was undisturbed in the absence of Pax5, thereby providing evidence in favor of the second mechanism. It remains to be determined whether Pax5 also induces some degree of organization in the absence of CTCF.

### Level 3

The first two levels of compaction produce an *IgH* locus with a structured 2.5-Mb  $V_H$  region and a 300-kb 3' domain. Several lines of reasoning suggest that yet another step is required to fully establish the prerrearrangement structure of the locus. First, YY1 deficiency results in locus decompaction and preferential recombination of the proximal  $V_H$  genes (Liu et al. 2007). This phenotype closely resembles that of Pax5-deficient pro-B cells, yet the major YY1-binding sites are located within the 3' *IgH* domain far away from Pax5-binding PAIR elements (Verma-Gaur et al. 2012). While YY1 function may depend on its weak binding to PAIR4, PAIR6, and PAIR8 (Medvedovic et al. 2013), an alternative possibility is that YY1 is required to increase association between the structured  $V_H$  region and the 3' *IgH* domain. Second, IGCR1-mutated *IgH* alleles undergo normal  $V_H$  region

compaction (Medvedovic et al. 2013), yet distal  $V_H$  recombination is substantially reduced, while proximal  $V_H$  recombination is substantially enhanced, compared with wild-type *IgH* alleles. This phenotype is similar to that of wild-type *IgH* alleles in Pax5- or YY1-deficient pro-B cells, although neither of these transcription factors is known to bind IGCR1. These observations suggest that locus compaction is insufficient for equal utilization of all  $V_H$  gene segments and indicate the need to incorporate additional structural features into models of *IgH* locus structure. Third, we showed previously that  $E\mu$  is required for full locus compaction, whereas the first two levels of compaction are  $E\mu$ -independent.

We propose that an  $E\mu$ -dependent third level of *IgH* compaction can account for these observations. This interaction brings the compacted  $V_H$  region into proximity with the 3' *IgH* domain via YY1 bound at  $E\mu$  interacting with sequences referred to as 3'558 and 5'7183 in the  $V_H$  region. In support of this idea, we demonstrate that the conformational state of  $E\mu$ -deficient alleles closely resembles that of wild-type alleles in YY1-deficient pro-B cells. Additionally, knockdown of YY1-associated factor SMC2 also disrupted  $E\mu$ -dependent interactions. Proof of a role for  $E\mu$ -bound YY1 awaits mutational analysis of the YY1-binding site within  $E\mu$ . In addition to associating with  $E\mu$ , we found that 3'558 and 5'783 were close to each other in wild-type alleles but not in  $E\mu$ -deficient alleles. Our interpretation is that these sequences are brought together by independently interacting with  $E\mu$ . Accordingly, 3'558–5'7183 interactions were also disrupted in YY1 knockdown or SMC2 knockdown pro-B cells.

Interestingly, 3'558–5'7183 interactions were also disrupted in Pax5-deficient pro-B cells. While the similarity between  $E\mu$  and YY1 deficiency can be readily explained by the model described above, the structural similarity in Pax5-deficient cells was unexpected. Our working model is that the  $V_H$  region must undergo both CTCF- and Pax5-dependent compaction in order for  $E\mu$  to interact with distant sites. In Pax5-deficient cells, only the first level of compaction takes place, and we infer that this is insufficient for  $E\mu$  to make stable/detectable long-range interactions. Therefore, 3'558 and 5'7183 do not come together in Pax5-deficient pro-B cells.

The similarity between the conformational states of the locus in Pax5-, YY1-, and  $E\mu$ -deficient pro-B cells requires some additional comments. We invoked a hierarchical model in the preceding paragraph to understand these observations. We favor this interpretation because the major sites of Pax5 binding (to PAIR elements) are in the distal  $V_H$  region, whereas major YY1-binding sites are in the 3' *IgH* domain at  $E\mu$  and in the 3'RR. Consequently, Pax5 deficiency is more likely to affect the structure of the distal  $V_H$  region without affecting other parts of the locus. While lack of a  $V_H$  region structure could underlie disrupted 3'558/5'7183 interactions in Pax5-deficient pro-B cells, we think that this is unlikely because (1) 5'7183 is more than a megabase away from the first PAIR element, and (2) the similarity between Pax5 deficiency and  $E\mu$ /YY1 deficiency suggests altered interactions between the  $V_H$  region and the 3' *IgH* domain mediated by  $E\mu$ . However, we note

that our data per se do not unequivocally rule out a nonhierarchical model in which Pax5 and  $E\mu$ /YY1 bring 3'558 and 5'7183 into spatial proximity by independent mechanisms. Use of additional probes of *IgH* structure will be essential to distinguish between these mechanisms.

The proposed role of  $E\mu$  in mediating the third level of locus compaction via YY1 binding is different from that of Medvedovic et al. (2013), who found that locus compaction (measured by FISH) and long-distance interactions (measured by 3C) were not altered on  $E\mu$  alleles. We surmise that the discrepancy rests on the nature and location of probes used to assay *IgH* locus conformation in the two studies. Our working hypothesis is that structural features queried by the use of short FISH probes in the present study may differ from those queried by BACs that hybridize to distal and proximal  $V_H$  genes. For example, spatial relationships determined by the use of 3'558, 5'7183, and  $E\mu$  (h4) probes may reflect site-specific interactions within a more complex, multilayered compaction state such as that revealed by the studies presented here. For example, earlier studies using BAC probes revealed relatively small effects of CTCF deletion on the overall compaction of the *IgH* locus (Degner et al. 2011). However, the use of more specific probes in the present study showed that CTCF-dependent loops in the  $V_H$  region were fully disrupted (to non-B-cell levels) in CTCF-depleted pro-B cells. This line of thinking cautions against drawing general conclusions about the structure of large loci from limited probe sets and emphasizes the continued need for high-resolution analyses to uncover principles of chromatin folding.

In summary, we propose that a three-step pathway establishes *IgH* locus conformation in pro-B cells. Each step requires a different transcription factor and leads to increasing levels of structural organization. The gradual assembly of the final structure may be enforced by the properties of each transcription factor and mechanisms of interactions involved. For example, CTCF and cohesin-mediated structural interactions may only be able to act over hundreds of kilobases, whereas YY1 and condensin-mediated interactions may be more long-range. Such principles may apply more generally to generate conformational fine structure within megabase-sized topologically associated domains.

We also demonstrate that *IgH* and *c-Myc* loci are brought together by CTCF in a substantial proportion of pro-B cells. Such interactions may predispose these loci to chromosomal translocations initiated by DNA double-strand breaks introduced at the *IgH* locus during VDJ recombination. Lymphomas carrying *IgH/c-Myc* t(12;15) translocations occur frequently in mice that are double deficient for components of the NHEJ machinery and p53 (Difilippantonio et al. 2002; Zhu et al. 2002). We envisage that DNA ends within *IgH* have time to recombine with the spatially proximal *c-Myc* gene when RAG-induced breaks are long-lived. The absence of pro-B lymphomas in single-mutant mice suggests that p53-dependent cell death guards against interchromosomal translocations even when DNA breaks are repaired slowly in the absence of NHEJ. Our studies indicate that the sites of interaction at

IgH are located at the 3' end of the locus, which contains several prominent CTCF-binding sites. Interestingly, mice that lack HS3a and HS4 of the 3'RR develop t(12;15)-containing pro-B-cell lymphomas in NHEJ<sup>-</sup>p53<sup>-</sup> mice (Gostissa et al. 2009). We infer that these regulatory sequences are not required for IgH/c-Myc interaction, leaving CTCF binding within HS5–7 as likely mediators of this phenomenon. IgH and c-Myc loci also interact in activated mature B cells (Roix et al. 2003; Osborne et al. 2004), which may predispose to translocations initiated by AID-induced DNA double-strand breaks. Since HS5–7 remain active in mature B cells, they could also serve as sites of interaction with c-Myc in the generation of mature B-cell lymphomas.

## Materials and methods

### Mice and cell culture

P<sup>+</sup>E<sup>-</sup> and RAG2-deficient mice were maintained at the National Institute on Aging animal facility. Rag1<sup>-/-</sup>YY1<sup>f/+</sup> × mb1-Cre and Rag1<sup>-/-</sup>YY1<sup>f/+</sup> × mb1-Cre mice were maintained at the breeding colony at The Scripps Research Institute (TSRI) as described previously (Verma-Gaur et al. 2012). Pro-B cells were purified from the bone marrow by positive selection using anti-CD19-coupled magnetic beads (Stem Cell Technologies) according to the manufacturer's protocol. The D345 pro-B-cell line contains an inactive RAG1 allele in a C57BL6 background (Ji et al. 2010), and the E $\mu$ <sup>-</sup> pro-B-cell line contains a 220-base deletion of E $\mu$  and lacks RAG2 (Chakraborty et al. 2009). Pax5-deficient pro-B cells were provided by Dr. Jagan Pongubala, and Pax 5<sup>-</sup>RAG2<sup>-</sup> cells were provided by Dr. Meinrad Busslinger. Both cells were expanded on S17 stromal cell cultures (Pongubala et al. 2008).

### FISH

BAC RP23-201H14 was kindly provided by Dr. Cornelis Murre (University of California at San Diego). Position-specific 10-kb probes were generated by PCR using BAC templates with the primers listed in Supplemental Table 5 or described in Guo et al. (2011a). FISH with 10-kb probes were performed as described in Guo et al. (2011a) using a Nikon T200 microscope equipped with a 100 $\times$  lens and motorized 100- $\mu$ m Piezo Z-stage (Applied Scientific Instrumentation). Depending on the size of the nucleus, 30–40 serial optical sections spaced by 0.2  $\mu$ m were acquired. The data sets were deconvolved using NIS-Elements software (Nikon). Volumes were measured using NIS-Elements software (Nikon). Statistical analyses of spatial distance measurements were carried out using a two-sample Kolmogorov-Smirnov test in R (Massey 1951).

### ChIP

ChIP experiments were performed essentially as described (Chakraborty et al. 2009). Antibodies used for ChIP were anti-YY1 (Santa Cruz Biotechnology, H414), anti-SMC2 (Abcam, ab10399), anti-SMC4 (Abcam, ab17958), and anti-YY1 (Santa Cruz Biotechnology, H414). Previously described primers for ChIP assays were from Guo et al. (2011a).

### YY1, SMC2, and CTCF knockdown

YY1 (SHCLNG-NM\_003403) and SMC2 (SHCLING-NM\_006444) shRNA lentiviral plasmids were purchased from Sigma

Aldrich. The empty vector pLKO.1 was used as the control. Silencing was carried out according to the manufacturer's instructions. CTCF knockdown was performed as previously described (Verma-Gaur et al. 2012).

### 3C assays

3C assays were performed as described (Wuerffel et al. 2007) using Hind III to digest cross-linked chromatin. 3C ligation products were measured by the TaqMan quantitative PCR technology (Hagege et al. 2007). PCR control fragments for the determination of primer efficiency of each primer combination were generated using BAC clones covering the genomic segments under study. PCR primers for 3C assay were from Guo et al. (2011a). Primers for V10-related 3C assays are in Supplemental Table 6.

## Acknowledgments

E<sup>-</sup>P<sup>+</sup> mice and Abelson virus transformed cell lines lacking E $\mu$  were generously provided by Eugene Oltz (Washington University School of Medicine, St. Louis, MO) and Fred Alt (Boston Children's Hospital, Boston, MA), respectively. Pax-5-deficient pro-B cells and S17 stromal cells were provided by Jagan Pongubala (University of Hyderabad, India), and Pax5<sup>-</sup>RAG2<sup>-</sup> pro-B cells were the gift of Meinrad Busslinger (Institute of Molecular Pathology, Vienna). We thank Dinah Singer (National Cancer Institute) for critical reading of the manuscript. This work was supported in part by the Intramural Research Program of the National Institutes of Health, by the National Institute on Aging, and in part by National Institutes of Health grant R01 AI 082918 to A.J.F. T.G., C.G., and R.S. designed the study. T.G. carried out all of the FISH experiments and C.G. and X.Q. carried out all ChIP and 3C experiments. L.M. assisted in FISH studies, and A.G. provided pro-B cells after expansion in culture on OP9 or S17 stromal cells. J.V.-G., N.M.C., and A.J.F. generated the CTCF-deficient and YY1-deficient primary pro-B cells used in these studies. R. S. wrote the manuscript, and all of the authors, especially A.J. F., provided essential comments.

## References

- Afshar R, Pierce S, Bolland DJ, Corcoran A, Oltz EM. 2006. Regulation of IgH gene assembly: role of the intronic enhancer and 5'DQ52 region in targeting DHJH recombination. *J Immunol* **176**: 2439–2447.
- Bossen C, Mansson R, Murre C. 2012. Chromatin topology and the regulation of antigen receptor assembly. *Annu Rev Immunol* **30**: 337–356.
- Chakraborty T, Perlot T, Subrahmanyam R, Jani A, Goff PH, Zhang Y, Ivanova I, Alt FW, Sen R. 2009. A 220-nucleotide deletion of the intronic enhancer reveals an epigenetic hierarchy in immunoglobulin heavy chain locus activation. *J Exp Med* **206**: 1019–1027.
- Chiarle R, Zhang Y, Frock RL, Lewis SM, Molin B, Ho YJ, Myers DR, Choi VW, Compagno M, Malkin DJ, et al. 2011. Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* **147**: 107–119.
- Choi NM, Loguercio S, Verma-Gaur J, Degner SC, Torkamani A, Su AI, Oltz EM, Artyomov M, Feeney AJ. 2013. Deep sequencing of the murine IgH repertoire reveals complex regulation of nonrandom V gene rearrangement frequencies. *J Immunol* **191**: 2393–2402.

- Degner SC, Wong TP, Jankevicius G, Feeney AJ. 2009. Cutting edge: developmental stage-specific recruitment of cohesin to CTCF sites throughout immunoglobulin loci during B lymphocyte development. *J Immunol* **182**: 44–48.
- Degner SC, Verma-Gaur J, Wong TP, Bossen C, Iverson GM, Tokamami A, Vettermann C, Lin YC, Ju Z, Schulz D, et al. 2011. 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* **108**: 9566–9571.
- Difilippantonio MJ, Petersen S, Chen HT, Johnson R, Jasin M, Kanaar R, Ried T, Nussenzweig A. 2002. Evidence for replicative repair of DNA double-strand breaks leading to oncogenic translocation and gene amplification. *J Exp Med* **196**: 469–480.
- Ebert A, McManus S, Tagoh H, Medvedovic J, Salvagiotto G, Novatchkova M, Tamir I, Sommer A, Jaritz M, Busslinger M. 2011. 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* **34**: 175–187.
- Engreitz JM, Agarwala V, Mirny LA. 2012. Three-dimensional genome architecture influences partner selection for chromosomal translocations in human disease. *PLoS One* **7**: e44196.
- Fuxa M, Skok J, Souabni A, Salvagiotto G, Roldan E, Busslinger M. 2004. Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes Dev* **18**: 411–422.
- Gombert WM, Farris SD, Rubio ED, Morey-Rosler KM, Schubach WH, Krumm A. 2003. The c-myc insulator element and matrix attachment regions define the c-myc chromosomal domain. *Mol Cell Biol* **23**: 9338–9348.
- Gostissa M, Yan CT, Bianco JM, Cogne M, Pinaud E, Alt FW. 2009. Long-range oncogenic activation of Igh-c-myc translocations by the Igh 3' regulatory region. *Nature* **462**: 803–807.
- Guo C, Gerasimova T, Hao H, Ivanova I, Chakraborty T, Selimyan R, Oltz EM, Sen R. 2011a. Two forms of loops generate the chromatin conformation of the immunoglobulin heavy-chain gene locus. *Cell* **147**: 332–343.
- Guo C, Yoon HS, Franklin A, Jain S, Ebert A, Cheng HL, Hansen E, Despo O, Bossen C, Vettermann C, et al. 2011b. CTCF-binding elements mediate control of V(D)J recombination. *Nature* **477**: 424–430.
- Hagege H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, de Laat W, Forme T. 2007. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* **2**: 1722–1733.
- Hesslein DG, Pflugh DL, Chowdhury D, Bothwell AL, Sen R, Schatz DG. 2003. Pax5 is required for recombination of transcribed, acetylated, 5' Igh V gene segments. *Genes Dev* **17**: 37–42.
- Ishihara K, Oshimura M, Nakao M. 2006. CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol Cell* **23**: 733–742.
- Jhunjhunwala S, van Zelm MC, Peak MM, Cutchin S, Riblet R, van Dongen JJ, Grosveld FG, Knoch TA, Murre C. 2008. The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. *Cell* **133**: 265–279.
- Ji Y, Resch W, Corbett E, Yamane A, Casellas R, Schatz DG. 2010. The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell* **141**: 419–431.
- Klein IA, Resch W, Jankovic M, Oliveira T, Yamane A, Nakahashi H, Di Virgilio M, Bothmer A, Nussenzweig A, Robbiani DF, et al. 2011. Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* **147**: 95–106.
- Kosak ST, Skok JA, Medina KL, Riblet R, Le Beau MM, Fisher AG, Singh H. 2002. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* **296**: 158–162.
- Kumar S, Wuerffel R, Achour I, Lajoie B, Sen R, Dekker J, Feeney AJ, Kenter AL. 2013. Flexible ordering of antibody class switch and V(D)J joining during B-cell ontogeny. *Genes Dev* **27**: 2439–2444.
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**: 289–293.
- Lin YC, Benner C, Mansson R, Heinz S, Miyazaki K, Miyazaki M, Chandra V, Bossen C, Glass CK, Murre C. 2012. Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nat Immunol* **13**: 1196–1204.
- Liu H, Schmidt-Supprian M, Shi Y, Hobeika E, Barteneva N, Jumaa H, Pelanda R, Reth M, Skok J, Rajewsky K. 2007. Yin Yang 1 is a critical regulator of B-cell development. *Genes Dev* **21**: 1179–1189.
- Massey FJ. 1951. The Kolmogorov-Smirnov test for goodness of fit. *J Am Stat Assoc* **46**: 68–78.
- 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. 2013. Flexible long-range loops in the VH gene region of the Igh locus facilitate the generation of a diverse antibody repertoire. *Immunity* **39**: 229–244.
- Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, Debrand E, Goyenechea B, Mitchell JA, Lopes S, Reik W, et al. 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet* **36**: 1065–1071.
- Pan X, Papisani M, Hao Y, Calamito M, Wei F, Quinn Iii WJ, Basu A, Wang J, Hodawadekar S, Zaprazna K, et al. 2013. YY1 controls Igk repertoire and B-cell development, and localizes with condensin on the Igk locus. *EMBO J* **32**: 1168–1182.
- Perlot T, Alt FW, Bassing CH, Suh H, Pinaud E. 2005. Elucidation of IgH intronic enhancer functions via germ-line deletion. *Proc Natl Acad Sci* **102**: 14362–14367.
- Phillips-Cremins JE, Corces VG. 2013. Chromatin insulators: linking genome organization to cellular function. *Mol Cell* **50**: 461–474.
- Pongubala JM, Northrup DL, Lancki DW, Medina KL, Treiber T, Bertolino E, Thomas M, Grosschedl R, Allman D, Singh H. 2008. Transcription factor EBF restricts alternative lineage options and promotes B cell fate commitment independently of Pax5. *Nat Immunol* **9**: 203–215.
- Roix JJ, McQueen PG, Munson PJ, Parada LA, Misteli T. 2003. Spatial proximity of translocation-prone gene loci in human lymphomas. *Nat Genet* **34**: 287–291.
- Sawado T, Halow J, Bender MA, Groudine M. 2003. The  $\beta$ -globin locus control region (LCR) functions primarily by enhancing the transition from transcription initiation to elongation. *Genes Dev* **17**: 1009–1018.
- Sayegh CE, Jhunjhunwala S, Riblet R, Murre C. 2005. Visualization of looping involving the immunoglobulin heavy-chain locus in developing B cells. *Genes Dev* **19**: 322–327.
- Schatz DG, Ji Y. 2011. Recombination centres and the orchestration of V(D)J recombination. *Nat Rev Immunol* **11**: 251–263.
- Schubeler D, Groudine M, Bender MA. 2001. The murine  $\beta$ -globin locus control region regulates the rate of transcription

- but not the hyperacetylation of histones at the active genes. *Proc Natl Acad Sci* **98**: 11432–11437.
- Shih HY, Krangel MS. 2013. Chromatin architecture, CCCTC-binding factor, and V(D)J recombination: managing long-distance relationships at antigen receptor loci. *J Immunol* **190**: 4915–4921.
- Subrahmanyam R, Sen R. 2012. Epigenetic features that regulate IgH locus recombination and expression. *Curr Top Microbiol Immunol* **356**: 39–63.
- Tsai AG, Lu H, Raghavan SC, Muschen M, Hsieh CL, Lieber MR. 2008. Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. *Cell* **135**: 1130–1142.
- Verma-Gaur J, Torkamani A, Schaffer L, Head SR, Schork NJ, Feehey AJ. 2012. 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* **109**: 17004–17009.
- Wuerffel R, Wang L, Grigera F, Manis J, Selsing E, Perlot T, Alt FW, Cogne M, Pinaud E, Kenter AL. 2007. S-S synapsis during class switch recombination is promoted by distantly located transcriptional elements and activation-induced deaminase. *Immunity* **27**: 711–722.
- Zhang Y, McCord RP, Ho YJ, Lajoie BR, Hildebrand DG, Simon AC, Becker MS, Alt FW, Dekker J. 2012. Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* **148**: 908–921.
- Zhu C, Mills KD, Ferguson DO, Lee C, Manis J, Fleming J, Gao Y, Morton CC, Alt FW. 2002. Unrepaired DNA breaks in p53-deficient cells lead to oncogenic gene amplification subsequent to translocations. *Cell* **109**: 811–821.