



CTCF and ncRNA regulate the three-dimensional structure of antigen receptor loci to facilitate V(D)J recombination

Nancy M. Choi and Ann J. Feeney*

Department of Immunology and Microbial Science, The Scripps Research Institute, La Jolla, CA, USA

Edited by:

Ananda L. Roy, Tufts University School of Medicine, USA

Reviewed by:

Amy L. Kenter, University of Illinois College of Medicine, USA
Kay L. Medina, Mayo Clinic, USA

*Correspondence:

Ann J. Feeney, Department of Immunology and Microbial Science, The Scripps Research Institute, 10550 North Torrey Pines Road, IMM-22, La Jolla, CA 92037, USA
e-mail: feeney@scripps.edu

At both the immunoglobulin heavy and kappa light chain loci, there are >100 functional variable (V) genes spread over >2 Mb that must move into close proximity in 3D space to the (D)J genes to create a diverse repertoire of antibodies. Similar events take place at the T cell receptor (TCR) loci to create a wide repertoire of TCRs. In this review, we will discuss the role of CTCF in forming rosette-like structures at the antigen receptor (AgR) loci, and the varied roles it plays in alternately facilitating and repressing V(D)J rearrangements. In addition, non-coding RNAs, also known as germline transcription, can shape the 3D configuration of the *Igh* locus, and presumably that of the other AgR loci. At the *Igh* locus, this could occur by gathering the regions being transcribed in the V_H locus into the same transcription factory where $I\mu$ is being transcribed. Since the $I\mu$ promoter, $E\mu$, is adjacent to the DJ_H rearrangement to which one V gene will ultimately rearrange, the process of germline transcription itself, prominent in the distal half of the V_H locus, may play an important and direct role in locus compaction. Finally, we will discuss the impact of the transcriptional and epigenetic landscape of the *Igh* locus on V_H gene rearrangement frequencies.

Keywords: V(D)J recombination, antigen receptor, chromatin, non-coding RNA, CTCF, histone modification, chromatin loop

INTRODUCTION

Antigen receptor (AgR) loci are facing a uniquely difficult task to produce a great diversity of receptors in order to recognize the limitless possibility of antigens present in the environment of an organism. With the advent of next generation sequencing, we can now determine the actual diversity of AgRs by sequencing all of the rearrangements from developing B and T cells. This diversity is created through the combinatorial recombination of multiple variable (V), diversity (D), and joining (J) gene segments at AgR loci by the RAG1/2 recombinase complex, along with the extensive junctional diversity at the V–D, D–J, and V–J junctions.

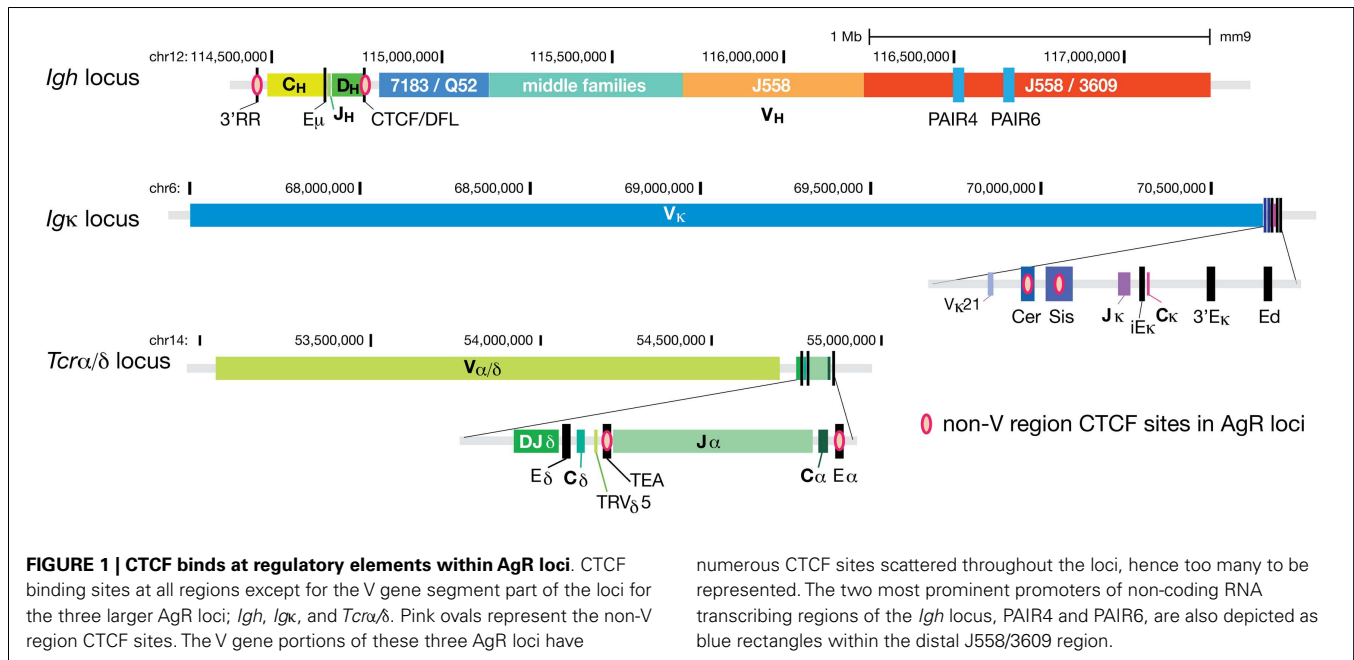
One of the most extensively studied AgR loci is the mouse *Igh* locus where the V_H , D_H , and J_H gene segments span a region of ~2.8 Mb (Figure 1). The 8–13 D_H genes, the four J_H genes, and all of the constant region genes and enhancers are located within a relatively small 300 kb region. In contrast, the 195 V_H genes, of which ~100 were deemed to be functional, are spread out over ~2.5 Mb. To create the greatest combinatorial diversity, all V genes would have to be able to access the D_H and J_H genes relatively equally regardless of their genomic distance. The question is then, how is this equality achieved?

With growing appreciation for how three-dimensional structural changes at the locus may bring V genes into proximity of the (D)J rearrangement to which one V gene will ultimately rearrange, current studies are employing cutting edge technologies to further understand this process. Chromatin conformation capture (3C) and its more recent modifications, 4C, 5C, and Hi-C (1–3), have allowed the identification of long-range chromosomal interactions, which facilitate the rearrangement of distant V genes by

making critical connections between the V genes and enhancers downstream (4). Next generation sequencing technologies coupled with chromatin immunoprecipitation (ChIP) (ChIP-seq) have allowed us to determine the binding sites of transcription factors throughout the genome as well as the genome-wide epigenetic landscape. Deep sequencing of RNA reveals the entire transcriptional profile of cells for both coding and non-coding RNA (ncRNA). Together, these techniques supply us with a bounty of information regarding the transcriptional and epigenetic profile of AgR loci at varying stages of differentiation. In this review, we will summarize and discuss how these recent studies have advanced our understanding of how long-range chromatin interactions and epigenetic changes may regulate V(D)J recombination at mouse AgR loci.

AgR LOCI UNDERGO LARGE SCALE THREE-DIMENSIONAL CONFORMATIONAL CHANGES DURING V(D)J REARRANGEMENT

All B cell and T cell receptor (BCR, TCR) subunits are formed through the process of V(D)J recombination. The BCR consists of two immunoglobulin heavy chains (*Igh*) and two identical light chains encoded by either the kappa (*Igk*) or lambda (*Igl*) loci. The TCR alpha (*Tcr α*) and beta (*Tcr β*), or delta (*Tcr δ*) and gamma (*Tcr γ*) chains constitute the TCR complex of the two major T cell subsets. The *Igh* and *Igk* are of similarly large sizes of approximately 2.8 and 3.2 Mb, while the *Tcr $\alpha\delta$* and *Tcr β* loci are smaller at 1.7 and 0.66 Mb. In comparison, the *Igl* and *Tcr γ* loci are much smaller, each only being about 200 kb. The challenge, which is particularly great for the large receptor loci, is to give all V genes



a chance to undergo rearrangement in order to create a diverse repertoire. How an AgR locus brings the V genes into proximity to the (D)J genes to create this diversity is still an unanswered question.

The original observations that showed three-dimensional structural changes at the *Igh* locus, presumably facilitating the creation of a diverse AgR repertoire, came from fluorescent *in situ* hybridization (FISH) studies (5). It was found that the *Igh* and *Igk* loci were predominantly located at the periphery of the nucleus in non-recombining cell types, but were found in more centralized locations in B cells. The nuclear periphery is generally considered a transcriptionally silent environment and is associated with repressive chromatin modifications, whereas gene dense active regions of the genome are more centrally located (6). Using two colors of probes at proximal and distal ends of the V_H locus, it was also shown for the first time that the *Igh* locus was in a more compacted conformation in recombining B cells. Subsequently, lineage- and developmental stage-specific locus contraction was observed for all of the large AgR loci: *Igk*, *Tcrα/δ*, and *Tcrβ* (7–10). This process of locus contraction is reversible, as demonstrated by the extension of the *Igh* locus in pre-B cells, when *Igh* rearrangement is complete (7). Contraction and re-extension of the distal end of the *Tcrα/δ* locus was also observed in double positive (DP) T cells (8). At this locus, contraction is necessary in double negative (DN) T cells for the accessibility of V genes used in TCR δ rearrangements, but in DP thymocytes, rearrangement of the more J-proximal V_α genes occurs before the rearrangement of distal V_α genes, so extension of the distal V_α genes would facilitate the ordered rearrangement of TCR V_α genes.

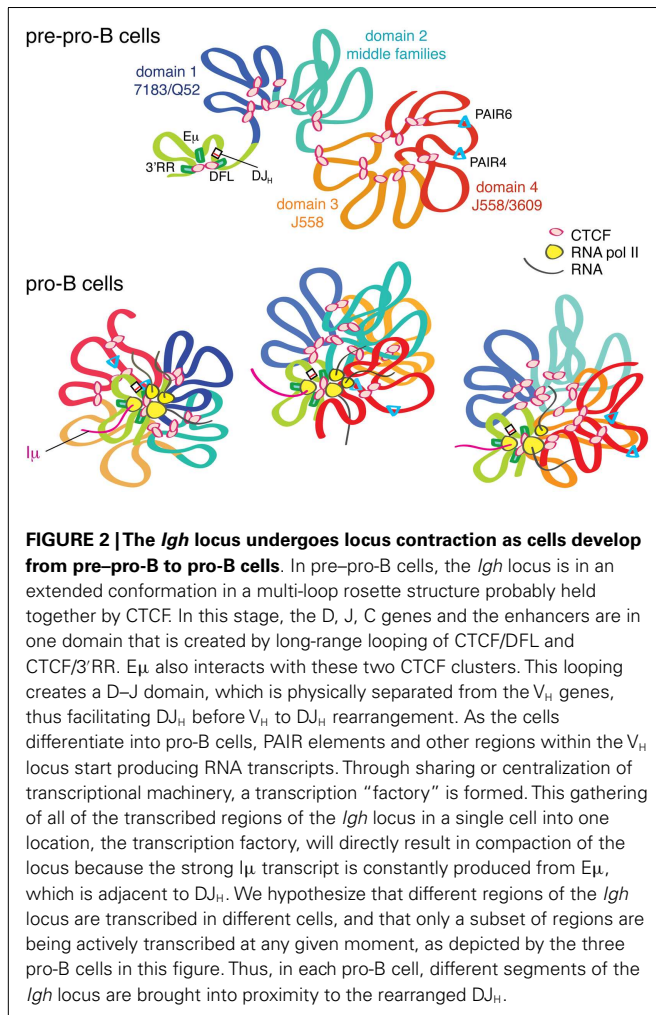
Greater insight to how such large-scale locus contraction may occur came from a 3D-FISH study by Jhunjunwala et al. that used multiple 10 kb probes spanning the entire *Igh* locus followed by 3D computational reconstruction of the location of all the probe binding sites (11). The results showed that the locus could be divided

into three ~1 Mb compartments in pre-pro-B cells in which multiple chromatin loops formed rosette-like structures (Figure 2). These compartments then collapsed into a single globule as cells developed into pro-B cells. This brought the distal V_H region into closer proximity within 3D space to the DJ $_H$ genes and regulatory elements, and in fact the distal V_H genes were found to be a similar distance away from the DJ $_H$ region as the proximal V_H genes (11).

It has been demonstrated that locus contraction of the *Igh* locus is regulated, directly or indirectly, by several key transcription factors. Mice deficient in YY1, Pax5, or the histone methyltransferase Ezh2 were impaired in locus contraction and in the rearrangement of distal V_H genes (12–15). Ikaros has also been implicated in *Igh* locus contraction (16), but Rag1/2 is not required for this process (5). Together, these studies suggest that contraction is a pre-requisite state for efficient recombination of distal V_H genes. Nonetheless, while AgR locus contraction is well established as a shared process among the large AgR loci that brings distal regions into closer 3D proximity to J genes prior to recombination, it has not been firmly determined what factors may be executing this task in the different lineages.

CTCF AND COHESIN BIND EXTENSIVELY WITHIN AgR LOCI

CTCF is an 11 zinc-finger protein that is the only known insulator binding protein in vertebrates (17, 18). Insulators are genetic regions that prevent heterochromatin on one side of the insulator from spreading into the other side. They can also prevent against positional effect variegation, or varied expression of transgenes, depending upon the site of integration in relation to where the insulator is located. Some insulators also have enhancer-blocking activity, where an enhancer cannot activate a promoter when separated by an insulator. It is now known that insulators function through CTCF that creates long-range chromatin interactions by binding to other CTCF bound sites (19). In this way, a domain is created by these chromatin loops, and activity or inactivity of



the genes within the domain is insulated from the activity of neighboring domains. In fact, CTCF has been found to play a major role in the establishment of the higher order organization of chromosomes genome-wide, and it is found at the boundaries of topological domains in numerous Hi-C studies (20–22).

CTCF is aided in this domain-creating function by cohesin. Cohesin's only known function until a few years ago was to hold sister chromatids together during mitosis by forming a ring around the sister chromatids with its four protein subunits (23). Now it is well recognized that cohesin is bound to many active CTCF sites, and thought to reinforce the loops created by the long-range CTCF–CTCF binding (24–26).

Because of the capability of CTCF to form long-range loops, we hypothesized that if CTCF were present at many sites in the AgR loci, it may play a role in determining the 3D structure of the loci and could possibly even influence locus contraction. Thus, we performed ChIP-chip, and subsequently ChIP-seq, to demonstrate that indeed CTCF was bound at numerous sites within the Ig loci, and was therefore an excellent candidate for creating multiple long-range loops (27, 28). If CTCF also had an important role in locus contraction, then we would predict that it would only be bound to the *Igh* locus in pro-B cells, the stage at which the *Igh*

locus undergoes contraction. However, we found by ChIP/qPCR that CTCF had a similar pattern of binding in pre-B cells and even in thymocytes, showing that CTCF binding was not lineage- or stage-specific (28). However, widespread binding of CTCF within the *Igh* locus was not observed in fibroblasts, demonstrating that the binding was at least lymphoid-specific. We then analyzed the binding pattern of cohesin by performing a ChIP/qPCR for Rad21, one of the cohesin subunits. This revealed that the level of Rad21 binding was higher in pro-B cells than in pre-B cells or thymocytes for many sites, suggesting cohesin may have a greater role than CTCF in specifying the developmental stage in which *Igh* recombination occurs (28).

CTCF displayed more lineage- and developmental stage-specific binding at the *Igk* locus (28). Some sites were only bound in pre-B cells, while others showed lower levels of binding in pro-B cells or thymocytes. Rad21 binding also displayed similar lineage and stage-specificity at the *Igk* locus. Investigation of ChIP-seq of CTCF binding at the large TCR loci showed various extents of lineage- and stage-specificity. At all AgR loci, however, we observed that the binding of cohesin was highest in the appropriate lineage and developmental stage. From these observations, it can be seen that CTCF and Rad21 may have different degrees of function in regulating lineage and stage-specific 3D structures at each AgR locus.

CTCF AND COHESIN INFLUENCE THE THREE-DIMENSIONAL STRUCTURE OF ANTIGEN RECEPTOR LOCI

To determine if CTCF made long-range loops that contributed to the compacted structure of the *Igh* locus in pro-B cells, we knocked down CTCF expression in RAG^{-/-} pre-B cells that were cultured in IL7 for 4 days (27). 3D-FISH was performed 4 days after knock-down of CTCF, and the spatial distance between two probes at the far ends of the *Igh* locus did increase, although not to the extent observed in YY1-deficient pro-B cells. This could be due to the fact that while CTCF binding was significantly reduced it was not completely eliminated at the *Igh* locus in the knocked-down pro-B cells as determined by ChIP. However, it is likely that CTCF is only one of many factors that are involved in the compacted structure of the *Igh* locus.

Further insight into the contribution of CTCF to the 3D structure of the *Igh* locus came from the 4C studies of Guo et al. (4). They described two different kinds of loops that formed at the *Igh* locus: E μ -dependent and E μ -independent loops. Using a CTCF ChIP-loop assay, they showed that the proximal regions had several CTCF-dependent and E μ -independent interactions, spanning a region of ~140 kb, as well as interactions with CTCF/DFL. Using a probe in the distal J558 region in the CTCF ChIP-loop assay, they demonstrated four sites of interaction within a 500 kb region, about half of the number of sites seen in 4C with the same distal probe. Importantly, none of the distal CTCF-dependent loops interacted with any other part of the *Igh* locus, and similarly the loops in the proximal region only interacted locally. Jhunjhunwala et al. previously demonstrated that the *Igh* locus consisted of three distinct rosette-like multi-looped structures in pre-pro-B cells that compacted upon themselves during locus contraction (11). Thus, it may be that most of the CTCF-dependent loops described by Guo et al. are local interactions that form the basic rosette-like

loops within the *Igh* locus. In addition to CTCF-mediated loops, locus contraction results from further large-scale interactions of these rosettes that are dependent upon $E\mu$. It may be that the longer range interactions require other key transcription factors such as YY1 and Pax5. YY1 binds to $E\mu$, and Pax5 binds to PAIR elements, the sites of greatest antisense transcription (29, 30). Whether these are the regions of most importance for YY1 and Pax5 binding with regard to locus contraction, or whether their primary influence is indirect, is not known. Our previous results that showed an increase in spatial distance between the two ends of the *Igh* locus after CTCF knockdown may reflect a loosening of the individual rosette structures while still being held together by other locus contraction regulating factors.

INSULATOR CTCF SITES BETWEEN THE V REGIONS AND D/J GENES AT AgR LOCI REGULATE REPERTOIRE DIVERSITY

The *Igh* locus has a pair of CTCF sites 3–5 kb upstream of the last functional D_H gene, DFL16.1 (28) (Figure 1). We and others have shown that this pair of CTCF sites (CTCF/DFL) has enhancer-blocking insulator activity in a traditional *in vitro* insulator assay (28, 31). By 3C, we have shown that CTCF/DFL loops to the cluster of nine CTCF sites downstream of the 3' regulatory region (3'RR) and to $E\mu$ (27), and this was subsequently confirmed by two other groups (4, 32). Coincidentally, Jhunjhunwala et al. utilized a probe near CTCF/DFL in their trilateration study (11), so we know that this D_H and J_H gene containing-loop is located far from the V_H genes in pre-pro-B cells, but it moves in close proximity to V_H genes in pro-B cells (Figure 2). We hypothesized that this loop creates a domain that contains all the D_H , J_H and constant region genes as well as the $E\mu$ enhancer, but excludes V_H genes (27). This would provide a physical environment in which D_H to J_H rearrangement could occur without any V_H genes in the vicinity.

Since the D_H genes have much antisense transcription, it was hypothesized that perhaps the function of CTCF/DFL was to stop antisense transcription from extending into the proximal V_H genes, preventing accessibility of those V_H genes (31). Indeed, deletion of the entire 96 kb intervening region between DFL16.1 and 7183.2.3 resulted in increased levels of D_H antisense transcription and extension of this transcription into the proximal V_H locus (33). However, knockdown of CTCF in pro-B cells with an intact *Igh* locus only resulted in extension of the antisense transcription for ~4 kb, and the antisense transcription dropped precipitously at the 3' Adam6 gene (27). Thus, preventing D_H region antisense transcription from extending into the V_H region does not seem to be the function of CTCF/DFL.

Importantly, Guo et al. deleted or mutated the CTCF/DFL sites, and the consequences were profound (32). Ordered rearrangement was perturbed, such that V_H to D_H rearrangement occurred as well as D_H to J_H rearrangement. More strikingly, rearrangements were confined to the two most proximal V_H genes. This shows that one critical function of these CTCF/DFL sites is to allow the creation of a diverse repertoire of *Igh* rearrangement, fully utilizing all of the V_H genes, although the mechanism by which this is achieved is not clear (34). In addition to these striking changes, deletion of CTCF/DFL resulted in a lack of lineage restriction, with V_H rearrangement being observed in thymocytes. Thus, two of the basic tenets of the accessibility hypothesis, ordered rearrangements

and lineage- and stage-specific restriction of V(D)J rearrangement, are regulated by this pair of CTCF binding sites at CTCF/DFL.

The *Igk* locus has two pairs of CTCF sites between the V_k and J_k genes (28) (Figure 1). One pair is within a region called "Sis" (Silencer in the Intervening Sequence), which also contains several Ikaros binding sites (35). When Garrard and colleagues deleted the 650 bp Sis element in the germline (36), these mice showed a modest preference for rearranging proximal V_k over distal V_k genes, and sense non-coding transcription over V_k genes was also slightly increased. Much more striking was the germline deletion of the strong CTCF sites upstream of Sis in the region called "Cer" (Contracting Element for Recombination) (37). In the $Cer^{-/-}$ mice, sense transcription over a few proximal V_k genes was increased modestly, but there was a very strong bias toward rearrangement of the most proximal V_k genes and a great reduction of rearrangement of the remainder of genes. This effect was reminiscent of the strong over utilization of the most proximal V_H genes in the CTCF/DFL deletion mice (32). Significantly, some *Igk* rearrangement was observed in thymocytes in $Cer^{-/-}$ mice (although mainly limited to J_k1), suggesting that the insulator sequences downstream of the V genes in both *Igh* and *Igk* loci are major contributors to the lineage restriction of Ig rearrangement. It should be mentioned that the *Igk* locus contraction was also reduced in $Cer^{-/-}$ mice, meaning extension of the locus could be a reason for the strong bias toward the most proximal V genes. Nonetheless, CTCF/DFL knockout mice did not display any change in *Igh* locus compaction (32), suggesting different modes of repertoire restriction at the two AgR loci.

In addition to the above studies in which the CTCF sites downstream of the V loci have been deleted, CTCF-deficient mice have been studied for their effects on repertoire formation. Hendriks and colleagues examined the *Igk* locus in mice carrying a B lineage-specific deletion of CTCF (38). By expressing a rearranged *Igh* gene they partially rescued development into pre-B cells. Absence of CTCF in pre-B cells resulted in a strong shift of usage to the most proximal V_k genes, where most rearrangements occurred at the 10 most proximal genes within the first ~200 kb in the knockout mice. V_k ncRNA were increased in this region, while the remainder of V_k ncRNA remained the same. Using Sis as an anchor/viewpoint for 4C-seq, it was demonstrated that the interactions of Sis with the 300 kb proximal region increased significantly. In contrast, iEk and 3'Ek viewpoints demonstrated that the enhancer interactions increased with sites up to 1 Mb into the V_k locus. However, other than a minor decrease of interaction of 3'Ek with the end of the V_k locus, the interactions of these three regulatory regions with the distal half of the V_k locus was unchanged. From these results, it seems that the majority of these long-range interactions between the enhancers or Sis with the distal 2/3 of the V_k locus are CTCF-independent interactions. Considering that the complete absence of CTCF in the cells gave a similar phenotype as the $Cer^{-/-}$ mice, the predominant effect of CTCF depletion throughout the *Igk* locus may be primarily due to the absence of CTCF binding at Cer.

As mentioned above, Rad21 (a subunit of cohesin) binds to CTCF sites in the AgR loci when rearrangement occurs (28, 39, 40). Seitan et al. analyzed the role of cohesin in V(D)J rearrangement at the *Tcr α / δ* locus (Figure 1) through the use of Rad21-deficient DP

thymocytes (39). Because cells cannot progress through cell division in the absence of cohesin, its role can only be ascertained in cells that do not divide, making DP thymocytes an appropriate cell type to study. They demonstrated that Rad21-deficiency resulted in reduced long-range looping between the CTCF/cohesin sites at TEA, the promoter of the germline transcripts of the 10 most 5' J α genes, and E α that also contains a CTCF/cohesin binding site. They also found an altered pattern of germline transcription in the J α region and reduced rearrangement to all but the most 5' J α genes in these Rad21-deficient mice.

A more detailed analysis of the role of CTCF/cohesin in TCR α rearrangement was performed using CTCF-deficient thymocytes (40). Shih et al. demonstrated by 3C that TEA and E α strongly interacted in wild type DP thymocytes, weakly in DN thymocytes, and not at all in B cells. TEA and E α also interacted with several proximal V α genes and with some J α genes, predominantly at the 5' portion of the J α region. In the Tcr α/δ locus, most functional V α genes have CTCF sites bound adjacent to the promoters, and thus it appears that normally CTCF nucleates a hub of proximal V α genes, a subset of J α genes, and the enhancer to create a functional recombination center. This entire hub of interactions was greatly reduced in E α -deficient DP thymocytes, and thus dependent upon E α . Deletion of TEA resulted in a shift of the peak of interaction of E α to the middle J α genes, likely explaining the previous observations that TEA deletion shifted the predominant rearrangements and germline transcription to the middle J α genes (41). In contrast to these results in wild type mice, 3C analysis of CTCF-deficient DP thymocytes revealed a reduction in the E α interactions with TEA, 5' J α , and certain 3' V α genes, and the level of rearrangement at the Tcr α locus was greatly reduced. Strikingly, the CTCF-deficient DP thymocytes showed increased E α contacts with the Tcr δ gene segments that are just upstream of TEA. Thus, it appears that the role of CTCF is to promote E α interactions with the 3' V α and 5' J α genes, while discouraging interactions with the intervening Tcr δ genes. 3D-FISH experiments demonstrated that the 3' end of the locus was still contracted in CTCF-deleted DP thymocytes, but 3C results showed that the long-range interactions were reduced for some 3' V α genes in DP thymocytes in the absence of CTCF. The level of transcription paralleled the new contacts as TEA-dependent transcription was decreased and transcription of Tcr δ genes was increased. Notably, this pattern of altered transcription and 3C contacts paralleled that seen in TEA^{-/-} mice, suggesting that it is the CTCF binding to TEA in WT DP thymocytes that directs E α to interact with 5' V α and 3' J α and promotes their transcription and subsequent rearrangement. CTCF binding to TEA also presumably directs E α to skip over the more proximal Tcr δ genes and instead interact with the 5' V α genes further away in the locus. In this way, the function of the CTCF-binding region at TEA resembles that of CTCF/DFL and Cer/Sis in that it prevents interactions with the immediately proximal genes, and instead directs interactions to V genes that are further away, allowing the creation of a diverse repertoire of AgR.

3D CHANGES CAUSED BY NON-CODING RNA

For many years we have known that the J and C genes of each AgR locus undergo high levels of non-coding transcription when the

locus is undergoing rearrangement (42, 43). In addition, V genes can produce low levels of sense ncRNA (or “germline transcription”) when they are accessible for rearrangement (44). In a few cases it has been demonstrated that these sense ncRNAs begin at the V gene’s promoter and stop shortly after the RSS and presumably this is the extent of most sense ncRNA. More recently, ncRNA in the antisense direction was described, and these ncRNAs are largely intergenic and longer (45). We performed directional RNA-seq of the *Igh* locus, thus defining all of the sense and antisense ncRNA within the locus in pro-B cells (29). Strikingly, there were three major regions of antisense ncRNA, and two minor antisense regions. The three major transcripts began at three of the PAIR elements, PAIR 4, 6, and 11. The 14 PAIR elements, or Pax5 Inter-genic Repeat elements, consist of binding sites for Pax5, E2A, and CTCF. These regions have high levels of H3K4me3 and H3ac, as would be expected since they are so highly transcribed (29). The two minor regions of antisense ncRNA were in the proximal J558 region, the site of the originally described antisense RNA (45), and near the J606 genes.

It is now widely accepted that transcription takes place in sub-nuclear compartments called transcription factories, which are clusters of RNA polymerases (46, 47). Many genes are transcribed within each transcription factory, and often co-regulated genes occupy one together regardless of their genomic distance, and even genes on separate chromosomes may co-localize to the same factory (47, 48). It can be hypothesized that if all *Igh* ncRNA were to be transcribed from the same transcription factory, any regions within the V_H part of the *Igh* locus that are being transcribed will of necessity be brought into juxtaposition with E μ , which contains the promoter of the predominant I μ germline transcript (29, 49). I μ is constantly transcribed and located 1–2.2 kb downstream of the J_H genes (50). This would mean that any V_H genes being transcribed would be close to the DJ_H region to which one of the V_H genes would ultimately rearrange in each pro-B cell (Figure 2). In support of this hypothesis, we demonstrated by 3C that PAIR4 and PAIR6, the regions of highest antisense transcription within the V_H region, directly interacted with E μ (29). We knew that YY1^{-/-} pro-B cells do not undergo locus contraction or rearrange distal V_H genes. With this in mind, we also showed that YY1^{-/-} pro-B cells did not undergo antisense transcription at PAIR elements, and their PAIR elements did not interact with E μ (29). Thus, it is possible that the lack of antisense ncRNA in the distal V_H region of YY1^{-/-} pro-B cells contributes to their lack of both locus contraction and rearrangement of distal J558 genes. We also saw a modest increase in antisense transcription at PAIR elements in CTCF-knockdown in RAG^{-/-} pro-B cells, and 3C analysis showed modestly increased interactions of PAIR and E μ . This is consistent with the idea that these interactions are taking place in a common transcription factory (27). By 3D-FISH, larger spatial distances between the proximal and distal ends of the *Igh* locus were seen in pro-B cells with CTCF knockdown, suggesting that CTCF is likely assisting in forming multiple loops within the *Igh* locus that “loosen” as its expression is reduced. However, the increase in PAIR–E μ interactions that we observed with loss of CTCF expression suggests that CTCF is not a major player in the pro-B specific locus contraction process.

DEEP SEQUENCING OF THE *Igh* REPERTOIRE IN PRO-B CELLS AND BIOINFORMATIC ANALYSES

While it is necessary to understand the effect of individual elements that regulate accessibility and chromatin structure at AgR loci, it is likely that many different factors are acting in concert for efficient production of a diverse repertoire. Recently, our lab and the Oltz lab adopted a bioinformatic approach with a goal to assign weight to the various factors that influence the frequency of rearrangement of individual V genes. To address this aim, we correlated the sequenced repertoires of mouse *Igh* and *Tcrβ* to ChIP-seq data for histone modifications and transcription factor binding and RNA-seq data for ncRNA transcripts (51, 52).

For the analysis of the mouse *Igh* repertoire in C57BL/6 mice, we sequenced 5'RACE-amplified cDNA from cell sorter purified pro-B cells to determine the pre-selection repertoire (51). Because this approach utilizes universal sequences to the 5' annealed adapter and Cμ on the expressed heavy chain transcript, it allows for an unbiased amplification of the expressed repertoire. In pro-B cells, as expected, the V_H genes were recombined at widely different frequencies throughout the locus. We assessed the histone post-translational modifications and transcript levels over each actively recombined gene and observed a significant distinction between V_H genes at the distal and proximal parts of the locus (Figure 3). Distal J558 family genes had greater enrichment for the active histone modifications (H3K4 methylation and H3 acetylation) as well as higher levels of both sense and antisense transcripts, than the proximal 7183 and Q52 families. This difference in epigenetic profiles suggests that these factors may be preferentially more influential at the distal half of the large *Igh* locus. We therefore divided the *Igh* locus into four domains based on V_H gene family locations, and found that domain 1, consisting of the 7183 and Q52 families, had very low levels of H3K4 methylation and the lowest levels of ncRNA. Domain 4, the most distal, containing all of the 3609 family as well as half of the J558 genes, had the highest levels of all the active histone modifications as well as the highest levels of both sense and antisense ncRNA. Domain 3, containing the remainder of the J558 genes, also had active chromatin marks and higher levels of ncRNA than the proximal genes.

When the relation to CTCF and Rad 21 binding was examined, all but one actively utilized gene of the proximal 7183 and Q52 families in domain 1 had a CTCF binding site within 100 bp, and all but one inactive gene had a CTCF site at ~1–20 kb distance. While at a genomic scale, a distance of 100 bp vs. >1 kb may not be of great difference, it may be enough distinction to place an RSS in close enough vicinity to the recombination center at the J_H region to provide a significant advantage to a V_H gene. CTCF binding at the base of the loop at CTCF/DFL, which is proximal to the rearranged DJ_H, and the base of the loop of functional V_H-adjacent CTCF sites in domain 1 would bring these regions in close proximity. Genes in the middle and distal regions did not show this tendency, suggesting that having a close CTCF binding site is most important for the genes at the proximal end of the *Igh* locus.

We previously demonstrated that RSS quality could influence V_H gene rearrangement frequency, and demonstrated that three different prototypic 7183 RSSs and a S107 RSS were more effective than a J558 RSS (53). All of the J558 RSSs are much further from the consensus RSS sequence than the 7183 RSSs. However, we also

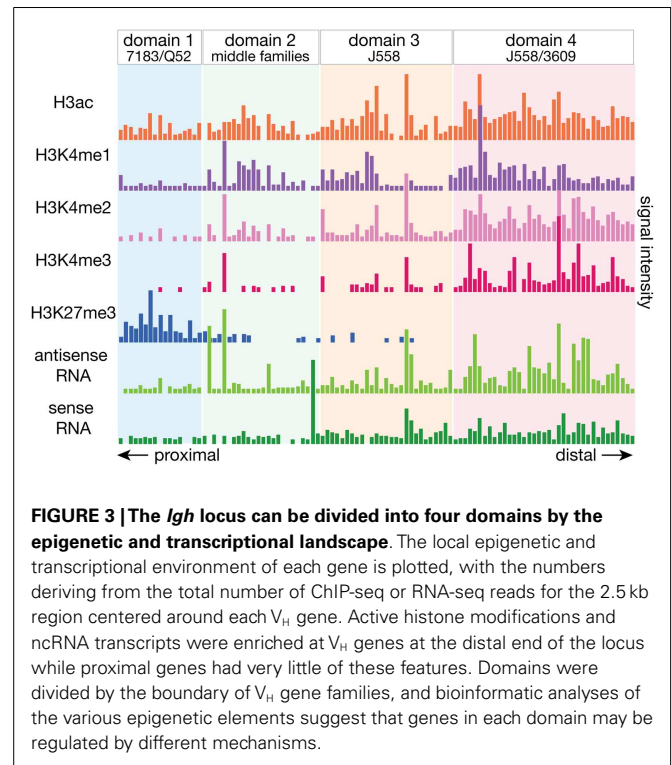


FIGURE 3 | The *Igh* locus can be divided into four domains by the epigenetic and transcriptional landscape. The local epigenetic and transcriptional environment of each gene is plotted, with the numbers deriving from the total number of ChIP-seq or RNA-seq reads for the 2.5 kb region centered around each V_H gene. Active histone modifications and ncRNA transcripts were enriched at V_H genes at the distal end of the locus while proximal genes had very little of these features. Domains were divided by the boundary of V_H gene families, and bioinformatic analyses of the various epigenetic elements suggest that genes in each domain may be regulated by different mechanisms.

showed that other parameters can override this effect, and that V genes with an identical RSS can rearrange at very different frequencies *in vivo* (53–55). Results from a computational model-building algorithm using our ChIP-seq, RNA-seq, and *Igh* repertoire deep sequencing data determined that having a functional RSS and an open chromatin environment as assessed by histone modifications were significant factors in predicting the activity of a V_H gene (51). When just the actively rearranging functional V_H genes were considered, the different domains of the V_H locus had different factors that correlated with recombination frequency. Within the proximal domain 1, proximity to the DJ_H genes was most significant, which is in agreement with the data we obtained a decade ago on another *Igh* haplotype, *Igh^a*, in pro-B cells from μMT mice (53). In contrast, at the distal domains, higher levels of active histone modifications appeared to be most important. This greater enrichment for active histone modifications at the distal V_H genes may reflect recruitment of these genes to the recombination center via transcription or some unknown factor that compensates for the disadvantages such as the distance from the DJ_H genes and their poorer RSSs.

At the *Tcrβ* locus, Gopalakrishnan et al. took a different approach of assessing individual Vβ gene usage by using a Taqman assay to measure rearrangement of genomic DNA rather than the 5'RACE approach that we used for the *Igh* repertoire (52). This approach is feasible at the *Tcrβ* locus due to the much smaller number of V genes compared to the *Igh* locus. When recombination frequency was compared to 3C interaction data, there was no rearrangement advantage observed for Vβ genes that displayed higher levels of interaction with the Dβ1 gene, leading authors to conclude that once the contraction has occurred at the relatively

smaller *Tcrβ* locus, spatial access is not a determining factor for $V\beta$ gene usage. However, it should be noted that all but two of the $V\beta$ genes are present within 235 kb at this locus, whereas the *Igh* and *Igk* V genes are spread over >2.5 kb. Therefore, proximity of V genes to (D)J genes in 3D space is much more likely to contribute to V gene rearrangement frequency in the large *Igh* and *Igk* loci. The bioinformatic analysis of all of the chromatin modifications, transcriptional activity, and 3D proximity for the *Tcrβ* locus led to the conclusion that having a functional RSS, higher nucleosome depletion (FAIRE assay), and higher RNA pol II binding were good indicators for active vs. inert $V\beta$ genes. They also concluded, for the actively rearranging genes, higher levels of active histone modifications correlated with higher levels of recombination, similarly to our conclusions for the domain 3 and 4 V_H genes.

The results from the *Tcrβ* and *Igh* locus considered together suggest that while generally accessible chromatin conformation and functional RSS sequences are both important, the different AgR loci are not governed by the same rules. In the case of the *Igh* locus, even the proximal and distal ends of the locus may be regulated by different mechanisms, which is likely due to its great expansion over a large genomic area and hence a greater need for locus contraction to bring the distal and middle V_H genes closer.

MODEL FOR THE ROLE OF CTCF AND ncRNA IN THE ESTABLISHMENT OF THE 3D STRUCTURE OF THE AgR LOCI

CTCF and its partner cohesin play important structural roles in creating large domains throughout the entire genome. Within AgR loci, there is a much higher density of CTCF/cohesin sites at rearranging loci than elsewhere in the genome. We hypothesize that the many CTCF/cohesin sites are necessary to create the multi-looped rosette-like structure that is the basic conformation of all AgR loci. This rosette structure makes it easier to compact various loci at the time of rearrangement. For some V genes, such as the V_H genes in domain 1 of the *Igh* locus, having a CTCF site near the RSS appears to be critical for a V_H gene to undergo rearrangement, but these V_H genes are rather poor in active histone marks and ncRNA. Thus, in lieu of these accessibility factors, being physically tethered to the recombination center, presumably by interactions with CTCF/DFL, is of great importance. In addition to the many CTCF sites throughout the large V gene portions of the AgR loci, CTCF/cohesin sites in between the V and J regions of the large AgR loci seem to be particularly important in regulating proper V gene rearrangements in a lineage- and developmental stage-specific manner (Figure 1). We also propose that ncRNA, or germline transcription, can directly facilitate *Igh* locus compaction if V_H genes or intergenic regions being transcribed are located in the same transcription factory as the $I\mu$ ncRNA. Since the DJ_H rearrangement is directly adjacent to the highly transcribed $I\mu$, transcription will place the DJ_H rearrangement very close to the transcription factory. We hypothesize that the structure of the *Igh* locus is very dynamic in pro-B cells, with different subsets of V_H genes being transcribed in each pro-B cell (Figure 2, bottom). Thus, we suggest that the dynamic and stochastic nature of germline transcription will physically move different parts of the V_H gene locus into proximity to the DJ_H rearrangement in each pro-B cell, and this will provide equal opportunity for V_H genes throughout the locus to come into proximity to the DJ_H rearrangement. Presumably, this

same activity could take place at the other AgR loci. In this way, the production of diverse repertoires of antibodies and TCR is assured.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health grants R01AI08218 and R21AI1007343.

REFERENCES

- Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science* (2002) **295**:1306–11. doi:10.1126/science.1067799
- Ohlsson R, Gondor A. The 4C technique: the ‘Rosetta stone’ for genome biology in 3D? *Curr Opin Cell Biol* (2007) **19**:321–5. doi:10.1016/j.ccb.2007.04.008
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragozcy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* (2009) **326**:289–93. doi:10.1126/science.1181369
- Guo C, Gerasimova T, Hao H, Ivanova I, Chakraborty T, Selimyan R, et al. Two forms of loops generate the chromatin conformation of the immunoglobulin heavy-chain gene locus. *Cell* (2011) **147**:332–43. doi:10.1016/j.cell.2011.08.049
- Kosak ST, Skok JA, Medina KL, Riblet R, Le Beau MM, Fisher AG, et al. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* (2002) **296**:158–62. doi:10.1126/science.1068768
- Kind J, van Steensel B. Genome-nuclear lamina interactions and gene regulation. *Curr Opin Cell Biol* (2010) **22**:320–5. doi:10.1016/j.ccb.2010.04.002
- Roldan E, Fuxa M, Chong W, Martinez D, Novatchkova M, Busslinger M, et al. Locus ‘decontraction’ and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nat Immunol* (2005) **6**:31–41. doi:10.1038/ni1150
- Shih HY, Krangel MS. Distinct contracted conformations of the *Tcrα/Tcrδ* locus during *Tcrα* and *Tcrδ* recombination. *J Exp Med* (2010) **207**:1835–41. doi:10.1084/jem.20100772
- Sayegh C, Jhunjhunwala S, Riblet R, Murre C. Visualization of looping involving the immunoglobulin heavy-chain locus in developing B cells. *Genes Dev* (2005) **19**:322–7. doi:10.1101/gad.1254305
- Skok JA, Gislis R, Novatchkova M, Farmer D, de Laat W, Busslinger M. Reversible contraction by looping of the *Tcrα* and *Tcrβ* loci in rearranging thymocytes. *Nat Immunol* (2007) **8**:378–87. doi:10.1038/ni1448
- Jhunjhunwala S, van Zelm MC, Peak MM, Cutchin S, Riblet R, van Dongen JJ, et al. The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. *Cell* (2008) **133**:265–79. doi:10.1016/j.cell.2008.03.024
- Liu H, Schmidt-Supprian M, Shi Y, Hobeika E, Barteneva N, Jumaa H, et al. Yin Yang 1 is a critical regulator of B-cell development. *Genes Dev* (2007) **21**:1179–89. doi:10.1101/gad.1529307
- Hesslein DG, Pflugh DL, Chowdhury D, Bothwell AL, Sen R, Schatz DG. Pax5 is required for recombination of transcribed, acetylated, 5' IgH V gene segments. *Genes Dev* (2003) **17**:37–42. doi:10.1101/gad.1031403
- Fuxa M, Skok J, Souabni A, Salvagiotto G, Roldan E, Busslinger M. Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes Dev* (2004) **18**:411–22. doi:10.1101/gad.291504
- Su IH, Basavaraj A, Krutchinsky AN, Hobert O, Ullrich A, Chait BT, et al. Ezh2 controls B cell development through histone H3 methylation and Igh rearrangement. *Nat Immunol* (2003) **4**:124–31. doi:10.1038/ni876
- Reynaud D, Demarco IA, Reddy KL, Schjervén H, Bertolino E, Chen Z, et al. Regulation of B cell fate commitment and immunoglobulin heavy-chain gene rearrangements by Ikaros. *Nat Immunol* (2008) **9**:927–36. doi:10.1038/ni.1626
- Phillips JE, Corces VG. CTCF: master weaver of the genome. *Cell* (2009) **137**:1194–211. doi:10.1016/j.cell.2009.06.001
- Wallace JA, Felsenfeld G. We gather together: insulators and genome organization. *Curr Opin Genet Dev* (2007) **17**:400–7. doi:10.1016/j.gde.2007.08.005
- Phillips-Cremins JE, Corces VG. Chromatin insulators: linking genome organization to cellular function. *Mol Cell* (2013) **50**:461–74. doi:10.1016/j.molcel.2013.04.018
- Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* (2012) **485**:376–80. doi:10.1038/nature11082

21. Hou C, Li L, Qin ZS, Corces VG. Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol Cell* (2012) **48**:471–84. doi:10.1016/j.molcel.2012.08.031
22. Botta M, Haider S, Leung IX, Lio P, Mozziconacci J. Intra- and inter-chromosomal interactions correlate with CTCF binding genome wide. *Mol Syst Biol* (2010) **6**:426. doi:10.1038/msb.2010.79
23. Nasmyth K, Haering CH. Cohesin: its roles and mechanisms. *Annu Rev Genet* (2009) **43**:525–58. doi:10.1146/annurev-genet-102108-134233
24. Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, et al. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* (2008) **132**:422–33. doi:10.1016/j.cell.2008.01.011
25. Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* (2008) **451**:796–801. doi:10.1038/nature06634
26. Rubio ED, Reiss DJ, Welch PL, Distechi CM, Filippova GN, Baliga NS, et al. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A* (2008) **105**:8309–14. doi:10.1073/pnas.0801273105
27. Degner SC, Verma-Gaur J, Wong TP, Bossen C, Iverson GM, Torkamani A, et al. CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the *Igh* locus and antisense transcription in pro-B cells. *Proc Natl Acad Sci U S A* (2011) **108**:9566–71. doi:10.1073/pnas.1019391108
28. Degner SC, Wong TP, Jankevicius G, Feeney AJ. Cutting edge: developmental stage-specific recruitment of cohesin to CTCF sites throughout immunoglobulin loci during B lymphocyte development. *J Immunol* (2009) **182**:44–8.
29. Verma-Gaur J, Torkamani A, Schaffer L, Head SR, Schork NJ, Feeney AJ. Non-coding transcription within the *Igh* distal V_H region at PAIR elements affects the 3D structure of the *Igh* locus in pro-B cells. *Proc Natl Acad Sci U S A* (2012) **109**:17004–9. doi:10.1073/pnas.1208398109
30. Ebert A, McManus S, Tagoh H, Medvedovic J, Salvaggio G, Novatchkova M, et al. 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**:175–87. doi:10.1016/j.immuni.2011.02.005
31. Featherstone K, Wood AL, Bowen AJ, Corcoran AE. The mouse immunoglobulin heavy chain V-D intergenic sequence contains insulators that may regulate ordered V(D)J recombination. *J Biol Chem* (2010) **285**:9327–38. doi:10.1074/jbc.M109.098251
32. Guo C, Yoon HS, Franklin A, Jain S, Ebert A, Cheng HL, et al. CTCF-binding elements mediate control of V(D)J recombination. *Nature* (2011) **477**:424–30. doi:10.1038/nature10495
33. Giallourakis CC, Franklin A, Guo C, Cheng HL, Yoon HS, Gallagher M, et al. Elements between the *Igh* variable (V) and diversity (D) clusters influence antisense transcription and lineage-specific V(D)J recombination. *Proc Natl Acad Sci U S A* (2010) **107**:22207–12. doi:10.1073/pnas.1015954107
34. Feeney AJ, Verma-Gaur J. CTCF-cohesin complex: architect of chromatin structure regulates V(D)J rearrangement. *Cell Res* (2012) **22**:280–2. doi:10.1038/cr.2011.188
35. Liu Z, Widlak P, Zou Y, Xiao F, Oh M, Li S, et al. A recombination silencer that specifies heterochromatin positioning and Ikaros association in the immunoglobulin kappa locus. *Immunity* (2006) **24**:405–15. doi:10.1016/j.immuni.2006.02.001
36. Xiang Y, Zhou X, Hewitt SL, Skok JA, Garrard WT. A multifunctional element in the mouse *Igk* locus that specifies repertoire and *Ig* loci subnuclear location. *J Immunol* (2011) **186**:5356–66. doi:10.4049/jimmunol.1003794
37. Xiang Y, Park SK, Garrard WT. V_k gene repertoire and locus contraction are specified by critical DNase I hypersensitive sites within the V_k - κ intervening region. *J Immunol* (2013) **190**:1819–26. doi:10.4049/jimmunol.1203127
38. Ribeiro de Almeida C, Stadhouders R, de Bruijn MJ, Bergen IM, Thongjuea S, Lenhard B, et al. The DNA-binding protein CTCF limits proximal V_k recombination and restricts enhancer interactions to the immunoglobulin kappa light chain locus. *Immunity* (2011) **35**:501–13. doi:10.1016/j.immuni.2011.07.014
39. Seitan VC, Hao B, Tachibana-Konwalski K, Lavagnoli T, Mira-Bontenbal H, Brown KE, et al. A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature* (2011) **476**:467–71. doi:10.1038/nature10312
40. Shih HY, Verma-Gaur J, Torkamani A, Feeney AJ, Galjart N, Krangel MS. $Tcr\alpha$ gene recombination is supported by a $Tcr\alpha$ enhancer- and CTCF-dependent chromatin hub. *Proc Natl Acad Sci U S A* (2012) **109**:E3493–502. doi:10.1073/pnas.1214131109
41. Villey I, Caillol D, Selz F, Ferrier P, de Villartay J-P. Defect in rearrangement of the most 5' $Tcr\alpha$ following targeted deletion of T early a (TEA): implications for $Tcr\alpha$ locus accessibility. *Immunity* (1996) **5**:331–42. doi:10.1016/S1074-7613(00)80259-9
42. Yancopoulos GD, Alt FW. Regulation of the assembly and expression of variable-region genes. *Ann Rev Immunol* (1986) **4**:339–68. doi:10.1146/annurev.iy.04.040186.002011
43. Lennon GG, Perry RP. C mu-containing transcripts initiate heterogeneously within the *IgH* enhancer region and contain a novel 5'-nontranslatable exon. *Nature* (1985) **318**:475–8. doi:10.1038/318475a0
44. Yancopoulos GD, Alt FW. Developmentally controlled and tissue-specific expression of unrearranged V_H gene segments. *Cell* (1985) **40**:271–81. doi:10.1016/0092-8674(85)90141-2
45. Bolland DJ, Wood AL, Johnston CM, Bunting SF, Morgan G, Chakalova L, et al. Antisense intergenic transcription in V(D)J recombination. *Nat Immunol* (2004) **5**:630–7. doi:10.1038/ni1068
46. Mitchell JA, Fraser P. Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev* (2008) **22**:20–5. doi:10.1101/gad.454008
47. Cook PR. A model for all genomes: the role of transcription factories. *J Mol Biol* (2010) **395**:1–10. doi:10.1016/j.jmb.2009.10.031
48. Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, et al. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet* (2010) **42**:53–61. doi:10.1038/ng.496
49. Matheson LS, Corcoran AE. Local and global epigenetic regulation of V(D)J recombination. *Curr Top Microbiol Immunol* (2012) **356**:65–89. doi:10.1007/82_2011_137
50. Osborne CS, Chakalova L, Mitchell JA, Horton A, Wood AL, Bolland DJ, et al. Myc dynamically and preferentially relocates to a transcription factory occupied by *Igh*. *PLoS Biol* (2007) **5**:e192. doi:10.1371/journal.pbio.0050192
51. Choi NM, Loguercio S, Verma-Gaur J, Degner SC, Torkamani A, Su AI, et al. Deep sequencing of the murine *Igh* repertoire reveals complex regulation of nonrandom V gene rearrangement frequencies. *J Immunol* (2013) **191**:2393–402. doi:10.4049/jimmunol.1301279
52. Gopalakrishnan S, Majumder K, Predeus A, Huang Y, Koues OI, Verma-Gaur J, et al. Unifying model for molecular determinants of the preselection $V\beta$ repertoire. *Proc Natl Acad Sci U S A* (2013) **110**:E3206–15. doi:10.1073/pnas.1304048110
53. Williams GS, Martinez A, Montalbano A, Tang A, Mauhar A, Ogwaro KM, et al. Unequal V_H gene rearrangement frequency within the large V_H7183 gene family is not due to RSS variation, and mapping of the genes shows a bias of rearrangement based on chromosomal location. *J Immunol* (2001) **167**:257–63.
54. Feeney AJ. Genetic and epigenetic control of V gene rearrangement frequency. *Adv Exp Med Biol* (2009) **650**:73–81. doi:10.1007/978-1-4419-0296-2_6
55. Feeney AJ, Atkinson MJ, Cowan MJ, Escuro G, Lugo G. A defective V_kA2 allele in Navajos which may play a role in increased susceptibility to *Haemophilus influenzae* type b disease. *J Clin Invest* (1996) **97**:2277–82. doi:10.1172/JCI118669

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 20 December 2013; paper pending published: 07 January 2014; accepted: 28 January 2014; published online: 11 February 2014.

Citation: Choi NM and Feeney AJ (2014) CTCF and ncRNA regulate the three-dimensional structure of antigen receptor loci to facilitate V(D)J recombination. *Front. Immunol.* **5**:49. doi: 10.3389/fimmu.2014.00049

This article was submitted to *B Cell Biology*, a section of the journal *Frontiers in Immunology*.

Copyright © 2014 Choi and Feeney. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.