



# Variable Extent of Lineage-Specificity and Developmental Stage-Specificity of Cohesin and CCCTC-Binding Factor Binding Within the Immunoglobulin and T Cell Receptor Loci

Salvatore Loguercio<sup>1</sup>, E. Mauricio Barajas-Mora<sup>2</sup>, Han-Yu Shih<sup>3</sup>, Michael S. Krangel<sup>3</sup> and Ann J. Feeney<sup>2\*</sup>

<sup>1</sup> Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA, United States, <sup>2</sup> Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA, United States, <sup>3</sup> Department of Immunology, Duke University Medical Center, Durham, NC, United States

## OPEN ACCESS

### Edited by:

Barbara L. Kee,  
University of Chicago,  
United States

### Reviewed by:

Eugene Oltz,  
Washington University in St. Louis,  
United States  
Almudena R. Ramiro,  
Centro Nacional de Investigaciones  
Cardiovasculares (CNIC), Spain

### \*Correspondence:

Ann J. Feeney  
feeney@scripps.edu

### Specialty section:

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

**Received:** 23 November 2017

**Accepted:** 16 February 2018

**Published:** 08 March 2018

### Citation:

Loguercio S, Barajas-Mora EM, Shih H-Y, Krangel MS and Feeney AJ (2018) Variable Extent of Lineage-Specificity and Developmental Stage-Specificity of Cohesin and CCCTC-Binding Factor Binding Within the Immunoglobulin and T Cell Receptor Loci. *Front. Immunol.* 9:425. doi: 10.3389/fimmu.2018.00425

CCCTC-binding factor (CTCF) is largely responsible for the 3D architecture of the genome, in concert with the action of cohesin, through the creation of long-range chromatin loops. Cohesin is hypothesized to be the main driver of these long-range chromatin interactions by the process of loop extrusion. Here, we performed ChIP-seq for CTCF and cohesin in two stages each of T and B cell differentiation and examined the binding pattern in all six antigen receptor (AgR) loci in these lymphocyte progenitors and in mature T and B cells, ES cells, and fibroblasts. The four large AgR loci have many bound CTCF sites, most of which are only occupied in lymphocytes, while only the CTCF sites at the end of each locus near the enhancers or J genes tend to be bound in non-lymphoid cells also. However, despite the generalized lymphocyte restriction of CTCF binding in AgR loci, the Igκ locus is the only locus that also shows significant lineage-specificity (T vs. B cells) and developmental stage-specificity (pre-B vs. pro-B) in CTCF binding. We show that cohesin binding shows greater lineage- and stage-specificity than CTCF at most AgR loci, providing more specificity to the loops. We also show that the culture of pro-B cells in IL7, a common practice to expand the number of cells before ChIP-seq, results in a CTCF-binding pattern resembling pre-B cells, as well as other epigenetic and transcriptional characteristics of pre-B cells. Analysis of the orientation of the CTCF sites show that all sites within the large V portions of the IgH and TCRβ loci have the same orientation. This suggests either a lack of requirement for convergent CTCF sites creating loops, or indicates an absence of any loops between CTCF sites within the V region portion of those loci but only loops to the convergent sites at the D-J-enhancer end of each locus. The V region portions of the Igκ and TCRα/δ loci, by contrast, have CTCF sites in both orientations, providing many options for creating CTCF-mediated convergent loops throughout the loci. CTCF/cohesin loops, along with transcription factors, drives contraction of AgR loci to facilitate the creation of a diverse repertoire of antibodies and T cell receptors.

**Keywords:** CCCTC-binding factor, cohesin, long-range looping, immunoglobulin loci, T cell receptor loci, 3D chromatin topology

## INTRODUCTION

The evolutionarily conserved, ubiquitously expressed zinc finger protein CCCTC-binding factor (CTCF) plays many important roles in gene activation and/or gene repression (1). Many years ago, it was shown to be the main protein associated with insulator function, in that it insulates a transgene from variable expression due to integration position effects, and that it also blocks an enhancer from acting on a promoter when the insulator element is in between them (2, 3). Subsequently, it was appreciated that CTCF exerts essentially all these activities by virtue of its ability to create long-range loops, thus creating domains within which promoters and enhancers can interact (4–7). By contrast, interactions of promoters in one domain and enhancers in another seldom happen. More recently, it has been appreciated that the entire genome is organized into “topologically associating domains” (TAD), which are megabase-sized chromatin domains that are largely conserved across different cell types (8–10). The boundaries of almost all TADs are bound by CTCF, and thus CTCF is now well established as one of the major architectural proteins organizing the entire genome (8, 10–13). Since TADs are generally conserved among various cell types and even conserved across species, it is not surprising that the majority of CTCF sites are invariant across different cell types (8, 13, 14). However there are also CTCF sites that are cell type specific and which correlate with cell type-specific gene expression (14, 15).

Cohesin is often found bound at CTCF sites genome wide (16–18), and recent findings suggest a key role for cohesin in the long-range looping of CTCF to create TADs (19). According to the loop extrusion model, the cohesin complex binds to chromosomes and extrudes the chromatin, forming a loop, until it encounters a boundary element, predominantly a CTCF site (20–23). By extruding the chromatin until 2 CTCF sites in convergent orientation to each other are reached, long-range loops with convergent CTCF sites at the base of the loops are formed. Hence, cohesin plays an essential role in the formation of the CTCF loops at TAD boundaries. These loops are not fixed, and cohesin and CTCF will dissociate with time and then the cohesin complex will bind again, and begin the loop extrusion process again until a new CTCF-mediated loop is formed (20, 23, 24).

Lymphocyte differentiation is an ideal model system to study universal vs. lineage-specific vs. developmental stage-specific binding of CTCF since it is easy to isolate pure populations of the sequential stages of differentiation of the B and T lineage progenitors. The developmental pathway of lymphocytes is well worked out, and the rearrangement of antigen receptor (AgR) genes in T and B cells occurs at precisely defined stage of lymphocyte development (25). The process of V(D)J rearrangement creates a wide diversity of immunoglobulins (Igs) and T cell receptors (TCRs) to enable the immune system to combat a wide variety of pathogens and abnormal cells. Rearrangement at each of the B cell and T cell AgR loci occurs at precise stages of differentiation and only in the appropriate lineage of lymphocytes.

T and B cells both arise from a common progenitor cell type (26). The first stage of committed B cell development is the pro-B cell stage during which the Igh locus rearranges. Following successful productive rearrangement of an immunoglobulin heavy chain (Igh) gene, the pro-B cells differentiate into the pre-B cell stage during which the Igk and the Igl loci rearrange sequentially. T cell precursors migrate from the bone marrow (BM) to the thymus, and they first undergo rearrangement at the CD4<sup>+</sup>CD8<sup>-</sup> “double-negative” (DN) thymocyte stage, at which time the TCR $\beta$ , TCR $\gamma$ , and TCR $\delta$  rearrangements all take place. If both productive TCR $\gamma$  and TCR $\delta$  rearrangements occur, the cell differentiates into a  $\gamma\delta$  T cell. However, if a productive TCR $\beta$  rearrangement precedes productive rearrangement of both the TCR $\gamma$  and TCR $\delta$  loci, the DN thymocyte progenitor will differentiate into a double-positive (DP) thymocyte, at which stage the TCR $\alpha$  genes undergoes rearrangement and the DP thymocyte will then differentiate into an  $\alpha\beta$  T cell (27).

In each AgR locus, one of the many V gene segments joins to one of the many D or J gene segments to create an enormous diversity of receptors. However, some AgR loci are very large, up to 3 Mb, and the portion of each AgR locus which contains D and J genes covers only a very small fraction of each locus. Since the many V gene segments are spread so far apart, this raises the question of how V gene segments that are very distant from the J gene segments can manage to find the J genes to rearrange. Several years ago, it was shown by 3D fluorescence *in situ* hybridization (3D-FISH) that the Igh locus has a rosette-like structure made by multiple long-range interactions (28). This structure becomes even more compact at the pro-B cell stage of B cell development, the developmental stage when the Igh locus undergoes V(D)J rearrangement (28–30). This process of locus contraction brings the Vh genes, spread over 2.5 Mb, into closer proximity to the D and J genes to which one Vh will rearrange to create a functional VDJ exon encoding the variable antigen-binding part of the Igh protein. The other AgR loci were also shown to undergo locus contraction at or prior to the developmental stage when they undergo rearrangement (31–34).

We previously hypothesized that a protein such as CTCF, with its ability to make long-range loops, might be responsible for creating the rosette-like structure at the Igh and presumably at other AgR loci, and might also contribute to locus contraction (35). If this were a reasonable hypothesis, then there would need to be many CTCF and cohesin sites within the AgR loci, and if they contributed to locus contraction, CTCF binding might be increased in an AgR locus at the specific stage of B or T cell development at which that AgR locus undergoes rearrangement. In order to see if this were a viable hypothesis, we performed ChIP-chip, and subsequently ChIP-seq, for CTCF on pro-B cells and pre-B cells and indeed we found that there were many sites bound in the Igh and Ig kappa light chain loci (35, 36). However, it appeared from the ChIP-chip and from ChIP/qPCR that the CTCF binding at the Igh locus, although lymphoid specific, showed limited lineage- and stage-specificity (i.e., similar numbers in pro-B cells, pre-B cells, and thymocytes) (35). By contrast, we showed that CTCF binding demonstrated more stage-specificity at the Igk locus. Thus, CTCF binding, by itself,

**Abbreviations:** AgR loci, antigen receptor loci; Ig, immunoglobulin; BM, bone marrow; TAD, topologically associating domains.

cannot explain locus contraction, even though it contributes to the 3D conformation of the contracted *Igh* locus as determined by knockdown of CTCF in pro-B cells (36). However, we also performed ChIP-chip and ChIP/qPCR for Rad21, a component of the cohesin complex, and the pattern of Rad21 binding showed more developmental stage-specificity in both Ig loci (35).

In the current study, we greatly extend this analysis by presenting a detailed analysis of our ChIP-seq data of the pattern of CTCF and Rad21 binding in all of the Ig and TCR loci at the two stages of B cell development and two stages of T cell development when the various AgR loci undergo rearrangement. In addition, we also analyzed pre-pro-B cells, mature B and T cells, ES cells, and murine embryonic fibroblasts (MEF), in order to assess the extent of lineage- and stage-specificity of CTCF and cohesin binding at AgR loci. Although much of the CTCF binding in the large V region part of the AgR loci is lymphoid specific, our data presented here show that the extent of lineage (T vs. B lineage precursors) and developmental stage-specificity (pro-B vs. pre-B, DN vs. DP) of CTCF and cohesin binding varies significantly from locus to locus, with the *Igk* locus being the only one of the 6 AgR loci with clear lineage- and stage-specificity of the onset of most CTCF and cohesin binding. Also, we show that the pattern of CTCF and cohesin binding observed in the most mature B or T cell precursor population is maintained in mature B or T cells, respectively. MEFs show many fewer sites of binding while ES cells are intermediate. In contrast to the large V gene portion of each locus, the CTCF sites between the V and D/J genes and near enhancers tend to display conserved CTCF and cohesin binding among all of the cell types examined including non-lymphocytes. We hypothesize that the acquisition of binding of many CTCF/cohesin complexes within the V gene portion of each of the four large AgR loci establish many long-range loops, and provide a semi-contracted locus upon which transcription factors can act to direct full locus contraction so that V genes throughout each locus will have approximately equal opportunity to rearrange to create a diverse set of V(D)J rearrangements.

## MATERIALS AND METHODS

### Cell Types Being Analyzed

For all ChIP-seq analyses of AgR loci in lymphocyte progenitors, we use RAG1-deficient cell lines so that the AgR locus remains intact, unperturbed by V(D)J rearrangements that will result in unique deletions in each individual B or T cell precursor. RAG deficiency also provides a precise genetic block in differentiation, so the cell populations are free of contamination from later stages of differentiation. All CD19<sup>+</sup> cells from BM of RAG1<sup>-/-</sup> mice are pro-B cells. All CD19<sup>+</sup> cells from BM of RAG<sup>-/-</sup> mice bearing a rearranged *IgM* transgene are pre-B cells (37). Likewise, thymocytes from RAG<sup>-/-</sup> mice are all CD4<sup>-</sup>CD8<sup>-</sup> “double-negative” (DN) cells, and thymocytes from RAG<sup>-/-</sup> mice bearing a V $\beta$  transgene are all at the subsequent stage of “double-positive” (DP) cells. IL7-cultured pro-B cells were obtained from day 7 cultures of RAG1<sup>-/-</sup> pro-B cells cultured in media containing supernatant from J558 cells stably expressing IL7 as previously described (38). All mice were on the C57BL/6 background. All animal work is

approved by The Scripps Research Institute’s IACUC and follows their guidelines.

### ChIP/qPCR

ChIP was done with antibodies against H3K4me3 (Active Motif, Carlsbad, CA, USA), H3ac (EMD Millipore, Billerica, MA, USA), and H3K4me1 (Abcam) as previously described (39). Primers are listed in Table S1 in Supplementary Material.

### Gene Expression Analysis

RNA was extracted from cells using TRIzol (Life Technologies, Carlsbad, CA, USA), and cDNA was made with QuantiTect Reverse Transcription kit, and genomic DNA was eliminated using genomic DNA wipeout buffer (Qiagen, Valencia, CA, USA).

### ChIP-seq

ChIP-seq for CTCF and Rad21 were performed as previously described (36). CTCF and Rad21 ChIP-seq on pro-B cells were previously published (36), as were CTCF ChIP-seq on DN and DP (40) and CTCF and Rad21 on mature B and T cells (41, 42). Pre-pro-B cells were obtained from long-term culture of E2A<sup>-/-</sup> progenitors and the ChIP-seqs were from GEO GSM987801 and GSM987802 (43). The CTCF ChIP-seq for ES and MEF were GEO (GSM723004, GSM723008) (44). Rad21 ChIP-seq was only available for ES cells (GSM824847, GSM824848) (45), but not for MEF, so we also have included ChIP-seqs for another cohesin component, Smc1, for ES cells (GSM560341, GSM560342) and MEF (GSM560355, GSM560356) (46). All ChIP-seq were mapped to the reference genome of C57BL/6, mm9. All GEO accession numbers, including the new entries, are listed in Table S2 in Supplementary Material.

We utilized ChIP-seq performed in our lab over a 4-year period for pro-B cells, pre-B cells, DN, and DP thymocytes for CTCF and Rad21. The ChIP-seq for pre-pro-B cells, mature T and B cells, MEF and ES cells were obtained from GEO. In order to make the Genome Browser files from different labs and over different times and platforms comparable, we zoomed out to view 90 Mb regions of the genome on a few chromosomes, and adjusted the scale to make them equivalent in height (Figures S1A,B in Supplementary Material). This was done since we know that over half of CTCF sites are constant among different cell types, and that TADs, which are bounded by CTCF sites, are also relatively conserved across cell types (8, 15).

### Analysis of ChIP-seq Data

Preliminary quality control over raw sequence data was performed with FastQC 0.11 (47). Duplicate reads were removed before mapping, and TruSeq adapter sequences were removed with the HOMER trim tool (48). Experimental and input control fastq tags were aligned to the mouse reference genome (mm9) using Bowtie 1.1.2 (alignment parameters: -a -v 2 -m 3 —best —strata) (49).

Signal tracks normalized to 1M reads were generated from alignment (bam) files using MACS2 (50) according to the procedure described in Ref. (50) (parameters: macs2 callpeak -B —nomodel —extsize 147 —SPMR -g mm).



Confident peaks were called using MACS (v1.4.2) (50), with a false-discovery rate (FDR)  $\leq 1\%$  and the default  $p$ -value ( $1E-5$ ) as specified in Ref. (51). Downstream analysis and manipulation of the data, including annotation of peaks, motif finding, quantification of data at peaks/genomic regions, and overlap analysis, were performed with HOMER 4.8 (48) and R/Bioconductor (52). Custom code is available upon request.

In particular, peaks were annotated with genomic features obtained from RefSeq release 66 through [homer.ucsd.edu/homer/data/genomes/mm9.v5.7.zip](http://homer.ucsd.edu/homer/data/genomes/mm9.v5.7.zip). Peak overlaps were computed using the HOMER mergePeaks tool (parameter used: -d for literal overlaps).

The above pipeline was used to uniformly process all of our ChIP-seq data and raw sequencing data obtained from GEO. For the Rad21 and Smc1 (another cohesin subunit) ChIP-seq on ES and MEF, we used the called peaks provided by the authors.

## Analysis of Orientation for CTCF Sites

In order to assess directionality of CTCF-binding sites, we devised the following procedure:

- (1) Assemble a library of CTCF motifs. A total of nine CTCF motifs were considered—five motifs retrieved from InsulatorDB, and four motifs detected in pre-B and pro-B ChIP-seq with HOMER findMotifsGenome.pl. (three known motifs and one *de novo*) (Table S3 in Supplementary Material). InsulatorDB motifs were converted to motif matrices (input for subsequent HOMER processing steps) using HOMER seq2profile tool with 0 mismatches. Motif matrices for HOMER-detected motifs were downloaded directly from the motif finding report. Finally, a custom motif file was created following HOMER guidelines.
- (2) Run annotatePeaks.pl (HOMER) on confident ChIP-seq peaks in a specific region (bed file) against the above library of CTCF motifs (e.g., for Ig $\kappa$  it was CTCF pre-B ChIP-seq, limited to the Ig $\kappa$  locus). This will generate a peak annotation file with instances of CTCF motifs found within the confident peaks in the region.
- (3) For each peak (i.e., a row in the above annotation file) annotated to motif(s), assign a (+) (forward) or (-) (reverse) orientation to each motif, by checking the strand on which the motif is detected. If multiple motifs are detected, take the consensus sign. Ambiguous cases are discarded.
- (4) Using the data in the annotation file, assemble a bed file with strand (orientation) information.

## RESULTS

### Optimization of Mapping Parameters for AgR Loci

We have found that the common method of analyzing ChIP-seq data by mapping only unique reads is not appropriate for AgR loci which have arisen by extensive gene duplication of individual V genes. It is especially important for the TCR $\alpha/\delta$  locus, which has a relatively recent triplication of a large portion of the locus in C57BL/6 mice, the genome on which all of these data are

mapped (53). This became obvious when we mapped CTCF sites, and observed that the two more distal duplications (the “d” and “n” repeats) appeared not to have CTCF bound to them despite the fact that all three copies are extremely similar. We varied the parameters of the mapping, and found that the parameter v2m3, which allows retention and mapping of reads which map up to three different locations within the AgR locus, allows for accurate mapping (Figure S2A in Supplementary Material). This was verified by ChIP/qPCR using PCR primers that can distinguish the 3' proximal repeat from the d/n repeats, thus assaying the three triplicated regions. As can be seen in Figure S2B in Supplementary Material, both the 3' and d/n repeats bind CTCF. It can be seen that even in loci such as the Igh locus where the mapping difference between parameters are not as dramatic as at the TCR $\alpha/\delta$  locus, the number of reads decreased somewhat at v2m2 and more at v2m1. In particular, the distal half of the Igh locus, which harbors the large VhJ558 family, displays fewer reads at v2m1 (Figure S2C in Supplementary Material). Although other AgR loci tend to display greater V segment sequence divergence than in the TCR $\alpha/\delta$  locus, the v2m3 parameter is beneficial for all ChIP-seq analyses within multi-V gene AgR loci, and it was used for all of the analyses presented here.

### D-J-C and Enhancer-Proximal CTCF and Cohesin Binding Is Generally Not Lineage-Specific

To determine the lineage and stage-specificity of CTCF and cohesin binding, we performed ChIP-seq for CTCF and Rad21 (a subunit of cohesin) for two sequential stages in B lymphocyte differentiation (pro-B and pre-B) and two stages in thymocyte differentiation (DN and DP). We compared these data to published data for binding in ES cells and MEFs. The four largest AgR loci (Igh, Ig $\kappa$ , TCR $\beta$ , and TCR $\alpha/\delta$ ) have a similar organization with a large number of V genes spread over a major percentage of the loci, followed by the (D)J-C genes tightly clustered at one end of the locus. The much smaller Ig $\lambda$  and TCR $\gamma$  loci have a very different organization and many fewer V genes, and will be discussed later. All four large AgR loci have 1–4 CTCF bound sites between the V genes and the D and/or J genes, and the majority of these sites in the Igh, Ig $\kappa$ , and TCR $\alpha/\delta$  loci are bound by CTCF and cohesin not only in all four lymphocyte subsets but also in MEF and ES cells. Nevertheless, in some cases, such as Cer and Sis in the V $\kappa$ -J $\kappa$  intervening region (54, 55), only one of the two sites may be bound in all cell types (Table 1). By contrast, in the TCR $\beta$  locus, occupancy of CTCF and cohesin at the CTCF site called “5'PC” located 27.6 kb upstream of the D $\beta$ 1.1 gene (56), the weak CTCF site near the D $\beta$ 1.1 gene, and the CTCF site between E $\beta$  and V $\beta$ 31 is predominantly restricted to DP and DN thymocytes.

### Variable Lineage-Specificity of CTCF and Cohesin Binding in the V Gene Portions of AgR Loci

In contrast to the fairly uniform binding of CTCF/Rad21 in the D-J-C regions among diverse cell types, CTCF and Rad21 binding within the large V portion of these loci shows greater lymphocyte specificity (Figure 1). The Igh, TCR $\beta$ , and TCR $\alpha/\delta$  loci show

**TABLE 1** | CTCF and cohesin binding in D-J-C-enhancer regions: lineage- and developmental stage-specificity.

CTCF locus	Between V and D/J	Enhancer	# CTCF sites	Conserved in all lymphocyte subsets?	MEF	ES
Igh	IGCR1	3'RR	2	Yes	Low (site #2 only)	Yes
			9	Yes	Yes	Yes
Igx	Cer Sis	none	2	Yes but only site #1 in DN, DP	Yes	Yes
			2	Yes	Site #2 only	Site #2 only
			1	Yes	Yes	Yes
TCR $\alpha/\delta$	TEA	E $\alpha$	1	Yes	Yes	Yes
			2	Yes	Yes	Yes
TCR $\beta$	TCR $\beta$ 5' PC (between V and D) TCR $\beta$ 3 kb up from D $\beta$ 1.1	E $\beta/V\beta$ 31	1			
			1	DP, DN (tiny peak)	No	No
			1	DN, DP, pro-B only	No	Low
Cohesin locus	Between V and D/J	Enhancer	# CTCF sites	Conserved in all lymphocyte subsets?	MEF	ES
Igh	IGCR1	3'RR	2	Only #2 in DN, pre-pro-B	Low (site #2 only)	Yes
			9	Fewer in DP, DN	No	Yes
Igx	Cer Sis	none	2	Yes, but only site #1 in DN, DP; none pre-proB	No	Yes
			2	pro-B, pre-B, DP only	Very low	Yes
			1	Low in proB, preB, not in pre-pro-B	Yes	Yes
TCR $\alpha/\delta$	TEA	E $\alpha$	1	Yes	Yes	Yes
			2	Yes	Yes	Yes
TCR $\beta$	TCR $\beta$ 5' PC (between V and D) TCR $\beta$ 3 kb up from D $\beta$ 1.1	E $\beta/V\beta$ 31	1	DP, DN only	No	No
			1	DP, DN (tiny peak)	No	No
			1	DN, DP only	Yes	Yes

All peaks are ones identified by the peak-calling program MACS.

remarkably similar numbers of sites of CTCF binding throughout all stages of B and T progenitor populations, but many fewer sites in MEF and ES cells (Figure 1; Table S4 in Supplementary Material). Thus, they show lymphocyte-specificity, but among lymphocytes, lack lineage- and developmental stage-specificity. Pre-pro-B cells (E2A<sup>-/-</sup> early progenitors) show many fewer bound CTCF peaks, suggesting that entry into the B progenitor stage is important for the initiation of CTCF binding in Ig loci. Only the Ig $\kappa$  locus shows marked lineage- and stage-specificity for CTCF binding.

The number of V genes in each locus, the number of CTCF sites that are bound in the V gene portion of each locus at the stage of its rearrangement, and the location of those bound CTCF sites relative to the V genes coding regions are shown in Table 2. The number of bound sites does not directly correlate with the size of the locus. The number of CTCF or Rad21 bound peaks within the V region portion of the 2.8 Mb Igh and the 1.8 Mb TCR $\alpha/\delta$  loci are well over 100 (Figure 1; Table 2). The 3.2 Mb Ig $\kappa$  locus has only 65 bound CTCF sites, and the much smaller 666 kb TCR $\beta$  locus has 18 bound CTCF sites. A hypergeometric test shows significant enrichment of the number of bound sites in the AgR loci in the various cell types studied compared to genome-wide binding (Table S5 in Supplementary Material). 76–95% of peaks contain consensus motifs within each locus (Table 3).

With the exception of the TCR $\beta$  locus, the large loci show more lineage- and stage-specificity in binding of the cohesin subunit Rad21 (Figure 1). As with CTCF binding, the Ig $\kappa$  locus shows the greatest lineage- and stage-specificity, and again, MEF and ES cells show the least amount of Rad21 bound. However, it should be noted that our lymphocyte and non-lymphocyte data for Rad21 and Smc1 (another cohesin subunit) may not be directly comparable because we used MACS to calculate Rad21

peaks from all lymphocyte subsets whereas we plotted peak calls obtained by the authors for Rad21 and Smc1 in ES cells and MEFs. Nonetheless, cohesin binding appears to contribute greater lineage-specificity than CTCF binding at AgR loci.

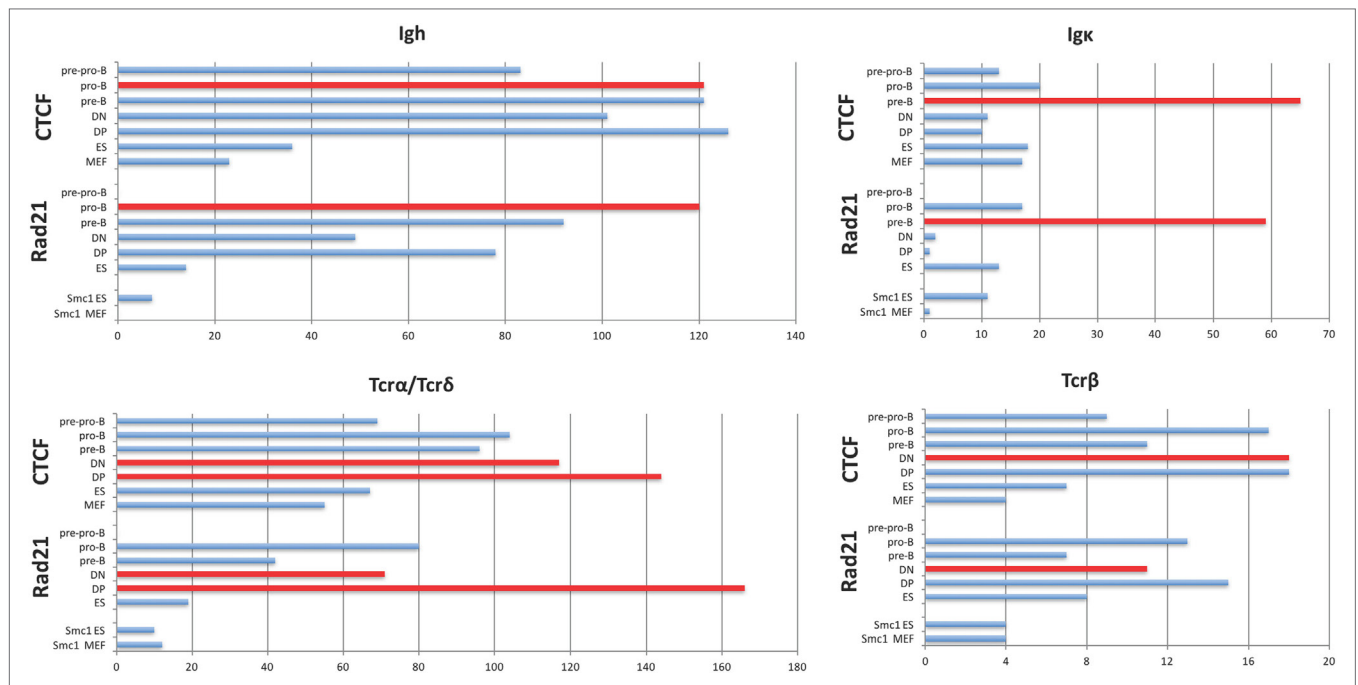
## Lineage- and Stage-Specificity of CTCF and Cohesin Binding Intensity in AgR Loci

Although the number of MACS called CTCF peaks in a locus is often similar among the different stages of lymphocyte differentiation, we observed substantial developmental stage-specificity in the abundance of reads (heights of peaks) of CTCF and cohesin binding at these sites. Often binding was highest in the developmental stage during which rearrangement normally occurs. The predominant reason for a higher peak at any specific location in one cell type vs. another is that a larger proportion of the cells used for the ChIP-seq lysate had CTCF (or Rad21) bound at that site. Within a locus, one reason for different heights of the various CTCF peaks could also be that a different proportion of cells had CTCF bound at the different sites, or it could reflect differences in the binding strength of CTCF to the individual sites, since the motif for CTCF is quite variable (57).

We examine each of the four large AgR loci separately in the sections below. The analysis of the two smaller AgR loci that do not have the same V, D, and J gene organization pattern will be presented later.

### Igh Locus

The Igh locus rearranges in the pro-B cell stage. T cells display some Dh-Jh rearrangements which may have occurred before the B and T cell lineages have bifurcated from their common progenitor (59). The 121 CTCF sites within the 2.5 Mb of the V gene part of the Igh locus show a similar pattern of peak heights



**FIGURE 1** | Plots of the number of significant peaks within the V gene-containing portion of each AgR locus called by MACS from ChIP-seq of pre-pro-B cells, pro-B cells, pre-B cells, double-negative (DN) and double-positive (DP) thymocytes, and murine embryonic fibroblast (MEF) and ES cells. The Rad21 and Smc1 peaks for ES and MEF were obtained from GEO, and the number of called peaks were obtained from GEO files. The cell type which is rearranging the locus being analyzed is plotted in red. The TCRα/δ locus rearranges at two stages since the TCRδ genes rearrange in DN thymocytes, and the TCRα genes rearrange in DP thymocytes. Statistics are shown in Tables S4 and S5 in Supplementary Material.

**TABLE 2** | Bound CCCTC-binding factor (CTCF) in V gene portions of all AgR loci.

Locus	Total number of V genes	# Functional V genes	Size of entire locus	# Bound CTCF within V gene locus	Location of CTCF sites relative to V genes
IgH	195	113	2.8 Mb	121	Within 100 bp of the RSS for most genes in the Vh families: 7183, Q52, S107, X24, Vh11, Vh12, Vh10, J606; intergenic for all the other Vh families
Igκ	140	125	3.2 Mb	65	Most intergenic, only a few within 4–8 kb upstream of promoter
Igλ	3	3	233 kb	3	Upstream of each V gene
TCRβ	33	22	666 kb	18	Several at/near promoters; 2 near RSS, several have none
TCRα/δ	131	109	1.8 Mb	144	Many upstream of V gene promoter, some intergenic
TCRγ	7	7	174 kb	2	Flanking 2 of 3 V clusters

The number of CTCF sites bound is for the cell type in which that locus rearranges.

**TABLE 3** | CTCF peaks with consensus motifs within V gene portion of each locus.

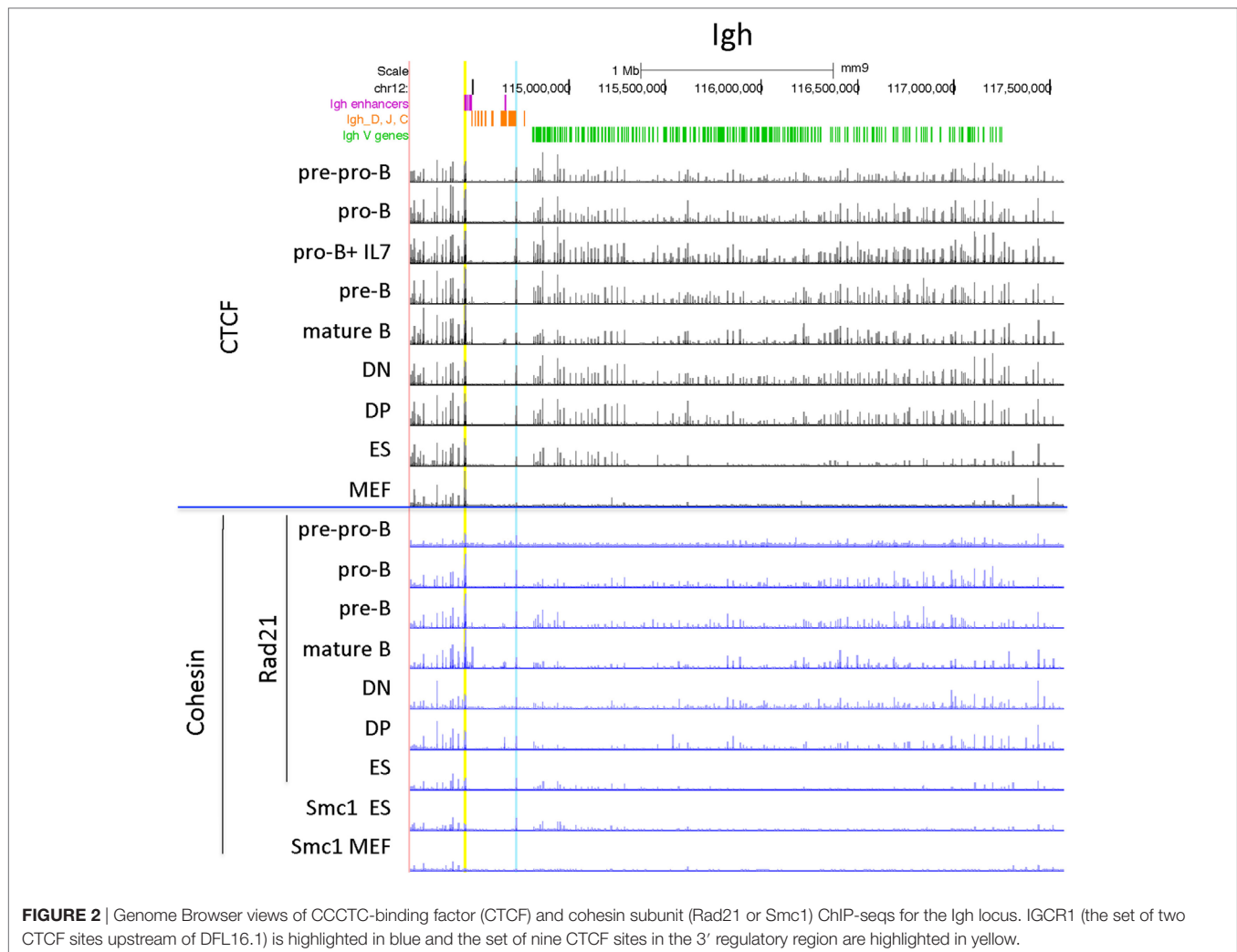
Region analyzed	ChIP-seq analyzed	Total # peaks (MACS)	# Peaks with one or more of the CTCF motifs	% of MACS peaks that matched at least one of nine CTCF PWM
genome wide	pre-B	42,250	34,560	81.8
Vh in IgH	pro-B	121	115	95.0
Vκ in Igκ	pre-B	65	50	76.9
Vβ in TCRβ	DN	18	14	77.8
Vα/δ in TCRαδ	DP	144	114	79.2

of CTCF binding throughout the locus for the lymphocyte subsets examined, although the peak heights are lower overall in the pre-pro-B cells (Figure 2). By contrast, the MEFs display

very little binding. The ES cells show a pattern of CTCF and cohesin binding similar to the lymphocytes at the Jh-proximal ~800 kb whereas the remaining Jh-distal ~1.7 Mb show few if any occupied CTCF sites in ES cells. The Rad21 binding pattern is similar for the pro-B, pre-B cells and mature B cells, but shows low binding in only in the most Jh-proximal part of the locus in MEF and ES cells. Pre-pro-B cells show little Rad21 binding anywhere within the locus.

### Igk Locus

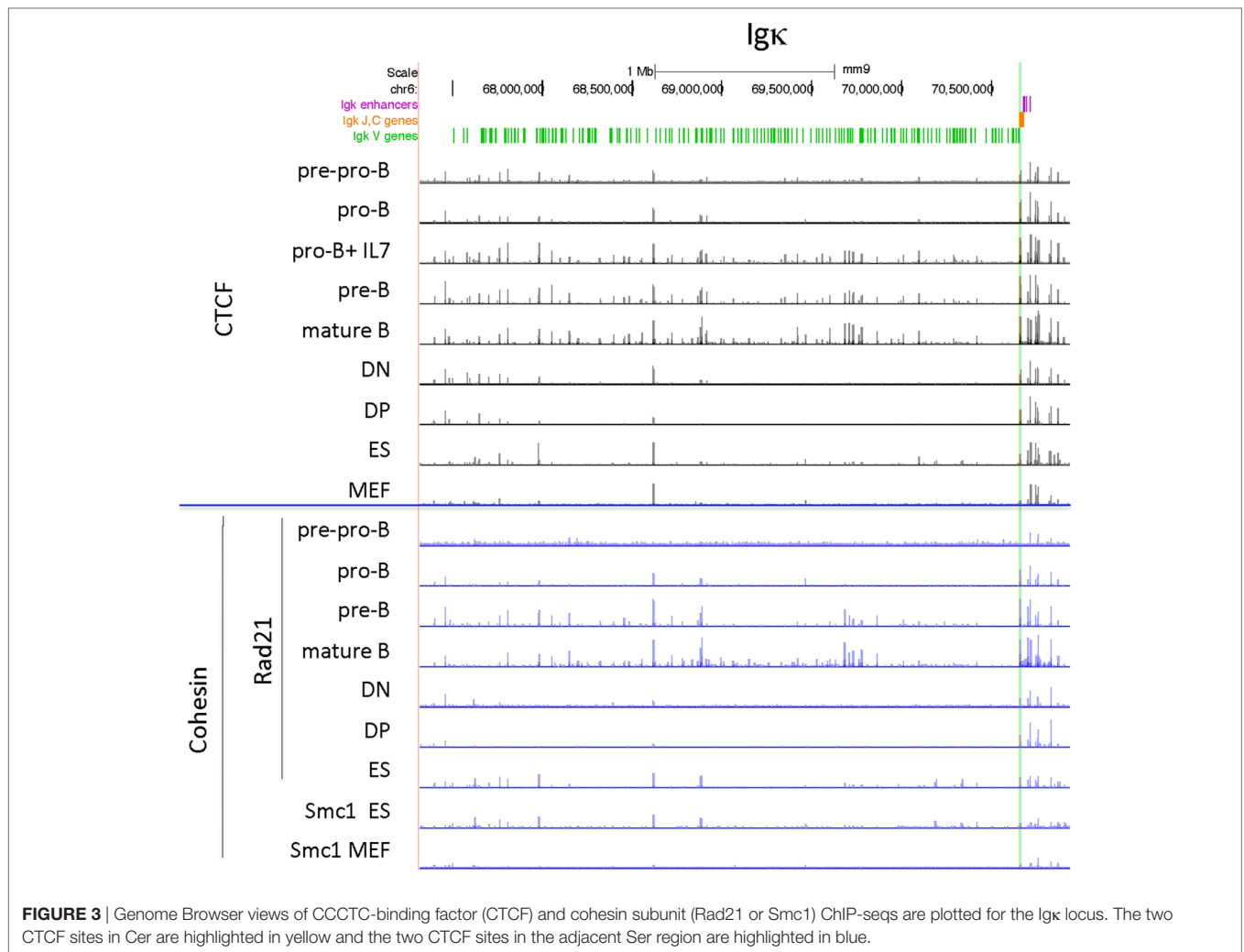
The Igk locus rearranges in pre-B cells and displays maximal CTCF and cohesin binding at that stage (Figure 3). Only one strong binding site 2 Mb away from the Jκ regions is present in all cell types examined, but the majority of the peaks show extensive specificity. Pro-B cells display lower level CTCF binding to several sites in the middle and Jκ-distal end of the locus, but very little



binding at all to the entire proximal half of the V portion of the locus. Other lymphocyte subsets and ES cells display only a few  $J\kappa$ -distal CTCF-binding peaks. There is little to no Rad21 binding in MEFs and pre-pro-B cells, and very few sites bound in DN and DP thymocytes. ES cells have some peaks for both Rad21 and Smc1 cohesin subunits, and pro-B cells have a few small peaks. Thus, the Igh locus shows greater lineage- and stage-specificity of CTCF and cohesin binding than any of the other AgR loci.

Culture of pro-B cells in IL7 for several days is often done to obtain large numbers of pro-B cells for ChIP or ChIP-seq (58, 60). However, we found that the CTCF-binding pattern of IL7-cultured RAG<sup>-/-</sup> pro-B cells resembles that of pre-B cells, not pro-B cells, in that all of the binding sites in the proximal half of the locus are now strongly binding CTCF (Figure 3). To further address the nature of these cells, we analyzed the epigenetic markers H3ac and H3K4me3 at select sites within the Igh and Igk loci by ChIP/qPCR, and also analyzed the production of ncRNA (also known as “germline transcription”) (Figure 4). Within the Igh locus, pro-B cells display extensive ncRNA at three PAIR elements in the distal part of the locus (61), and ChIP/qPCR shows a 10-fold reduction of this transcription in pre-B cells, and even greater

reduction in IL7-cultured pro-B cells (Figure 4B). Likewise the level of ncRNA over the VhJ558 genes (62) is substantially lower in both pre-B and IL7-cultured pro-B cells. The extent of H3K4me3 at PAIR4 in IL7-cultured pro-B cells is intermediate between that in pro-B and pre-B cells. ChIP-seq for H3K4me1, the epigenetic mark of enhancers, shows that although pro-B cells have very few sites of H3K4me1 in the V $\kappa$  locus in pro-B cells, IL7-cultured pro-B cells resemble pre-B cells in that they have many regions of H3K4me1 bound throughout the V $\kappa$  locus (Figure 4C). Transcription of the  $\kappa^o$  ncRNA (63), which runs through the J $\kappa$  and C $\kappa$  genes is greatly increased in pre-B cells, and is intermediate in the IL7-cultured pro-B cells compared to pro-B cells (Figure 4B). Likewise, the extent of H3ac at the J $\kappa$  genes is greatly increased in pre-B cells compared to pro-B cells and is intermediate in IL7-cultured pro-B cells (Figure 4A). There is a relatively high level of ncRNA through the V $\kappa$ 38-93 gene in pre-B cells but not pro-B cells, and the IL7-cultured pro-B cells resemble pre-B cells in both ncRNA transcription as well as in the H3K4me3 profile. Thus, for epigenetic marks, germline transcription, and CTCF binding, IL7-cultured pro-B cells resemble pre-B cells more than pro-B cells for many characteristics. Thus,



previous published data on IL7-cultured pro-B cells should be interpreted with caution.

### TCR $\alpha/\delta$ Locus

The organization of this locus is by far the most complex since it contains all the genes for both the alpha and delta chains of the TCR with TCR $\delta$  rearrangement occurring at the DN stage of thymocyte development, and TCR $\alpha$  rearrangement in the subsequent DP stage of development. However the pattern of CTCF binding at the TCR $\alpha/\delta$  locus is very similar in all lymphocyte subsets, with the exception of particularly low level binding in pre-pro-B cells (**Figure 5**). ES cells also have many sites bound by CTCF, and only MEFs have noticeably smaller CTCF peaks. By contrast, high level Rad21 binding is restricted to the two thymocyte subsets and mature T cells, and there is almost no cohesin bound in MEF or pre-pro-B cells.

### TCR $\beta$ Locus

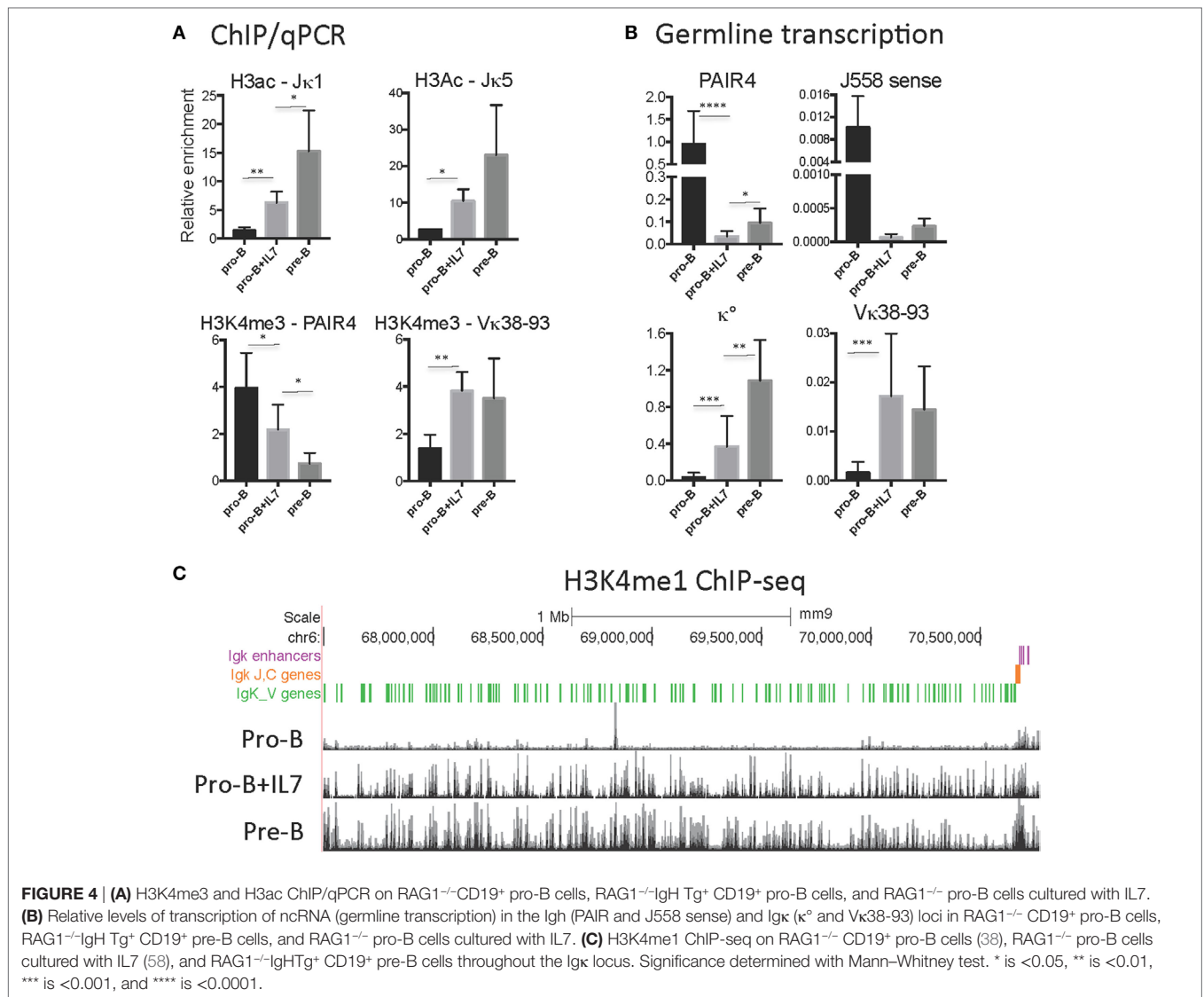
The TCR $\beta$  locus is the smallest of the 4 large AgR loci, with all but 2 of its 22 functional V genes located in a compact 235 kb region. CTCF binding throughout the locus is similar in all lymphocyte

subsets, but much less binding is found in ES cells and MEFs (**Figure 6**). Rad21 binding is highest in T cell subsets, but there is low level binding in B cell progenitors as ES cells.

### CTCF Orientation within the AgR Loci

It has been documented that 65–90% of long-range CTCF–CTCF interactions take place between CTCF sites present in convergent orientation (facing each other) (10, 64). The boundaries of TADs have a very high percentage of convergently oriented CTCF loops, probably due to the process of loop extrusion (20). However, there is more variable orientation of pairs of CTCF sites within TADs (13, 65). Since the many CTCF sites within the AgR loci could serve to create the rosette structure that the AgR loci are thought to adopt, and since they could also aid in bringing the many V genes, spread out over Mb in some cases, into close proximity to D-J genes, the orientation of CTCF sites within the AgR loci is of critical importance for predicting the role and location of CTCF–CTCF loops (28). We, therefore, compiled 9 CTCF motifs as described in Section “Materials And Methods,” and determined the orientation of all CTCF sites identified by any of those motifs.





Two of the large AgR loci, Igh and TCR $\beta$ , have rather simple CTCF orientation patterns, with all bound CTCF sites in the large V gene portion of the loci pointing toward the D-J-C regions, and most of the CTCF sites in the D-J-C-enhancer regions oriented toward the V region CTCF sites (66) (**Figures 7 and 8A**). By contrast, the two other large AgR loci, Igk and TCR $\alpha/\delta$ , have complex patterns of CTCF orientation, with many of the sites within the V portions of the loci pointing away from the D-J-C regions as well as many oriented toward the D-J-C regions, as described below.

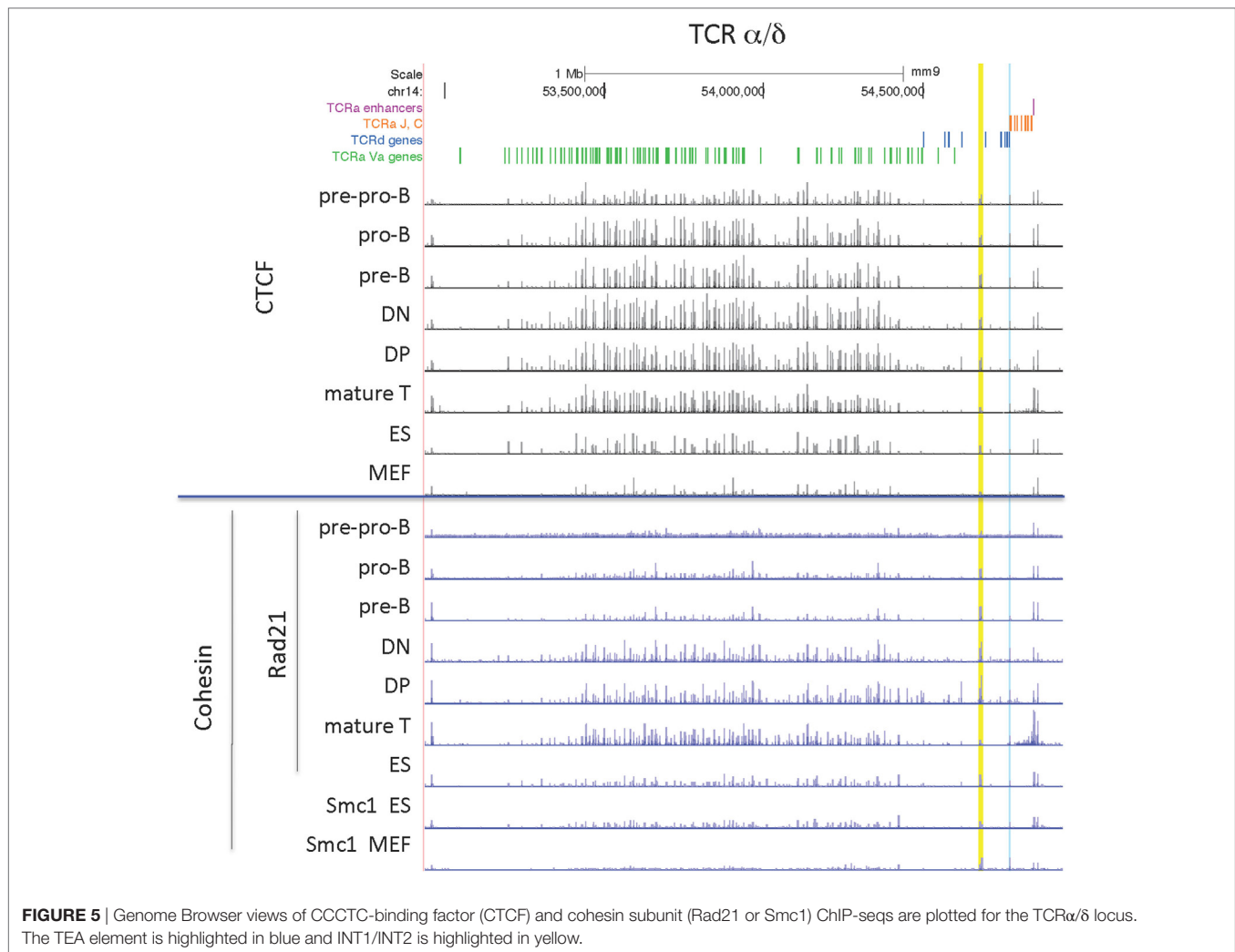
### TCR $\alpha/\delta$ Locus

CTCF sites are located in both orientations in the V part of the locus, although the majority of sites (129/144) are pointing toward the J-C regions (**Figure 8B**) as previously observed (67). The CTCF site at the TEA promoter is an important regulatory element located upstream of the 61 J $\alpha$  gene segments (68, 69), and it is oriented toward the V genes. The next CTCF site with the same orientation is just upstream of the V gene Trav17. This

CTCF near Trav17 is oriented to make long-range loops with any of the many upstream CTCF sites throughout the V region part of the locus. Two prominent CTCF sites upstream of Trdv2-2 are oriented toward TEA, and have been demonstrated interact with TEA (67). Deletion of those two CTCF sites, called INT1 and INT2, results in altered TCR $\delta$  and TCR $\alpha$  V gene usage (67).

### Igk Locus

The Igk locus has four CTCF sites in between the V genes and the J genes. The set of two sites closest to the V genes, called Cer, plays an important role in locus conformation and regulation, and both CTCF sites are oriented toward the V genes (55). The set of two CTCF sites closest to the J $\kappa$  genes, called Sis, has a less prominent role in locus regulation (54). One of those CTCF sites in Sis is orientated toward the V genes and the other is in the opposite orientation. The CTCF sites in the 3.1 Mb V region portion of the locus are present in both orientations (**Figure 9**). The majority of sites in the J-proximal half of the locus are pointed away from the CTCF sites in the V-J intervening region. Thus,



these CTCF sites are not oriented to participate in looping to convergently oriented CTCF sites that would bring proximal V genes closer to the J genes. By contrast, the majority of CTCF sites in the distal half of the locus are oriented toward the J genes, and thus could potentially interact in convergent orientation with the CTCF sites in Cer and Sis in the V-J intervening region. Unlike all other AgR loci, approximately half of the V $\kappa$  genes are in the opposite transcriptional orientation to the J  $\kappa$  segments and so rearrange by inversion. However, there is no correlation between the orientation of CTCF sites and the transcriptional orientation of the V $\kappa$  genes.

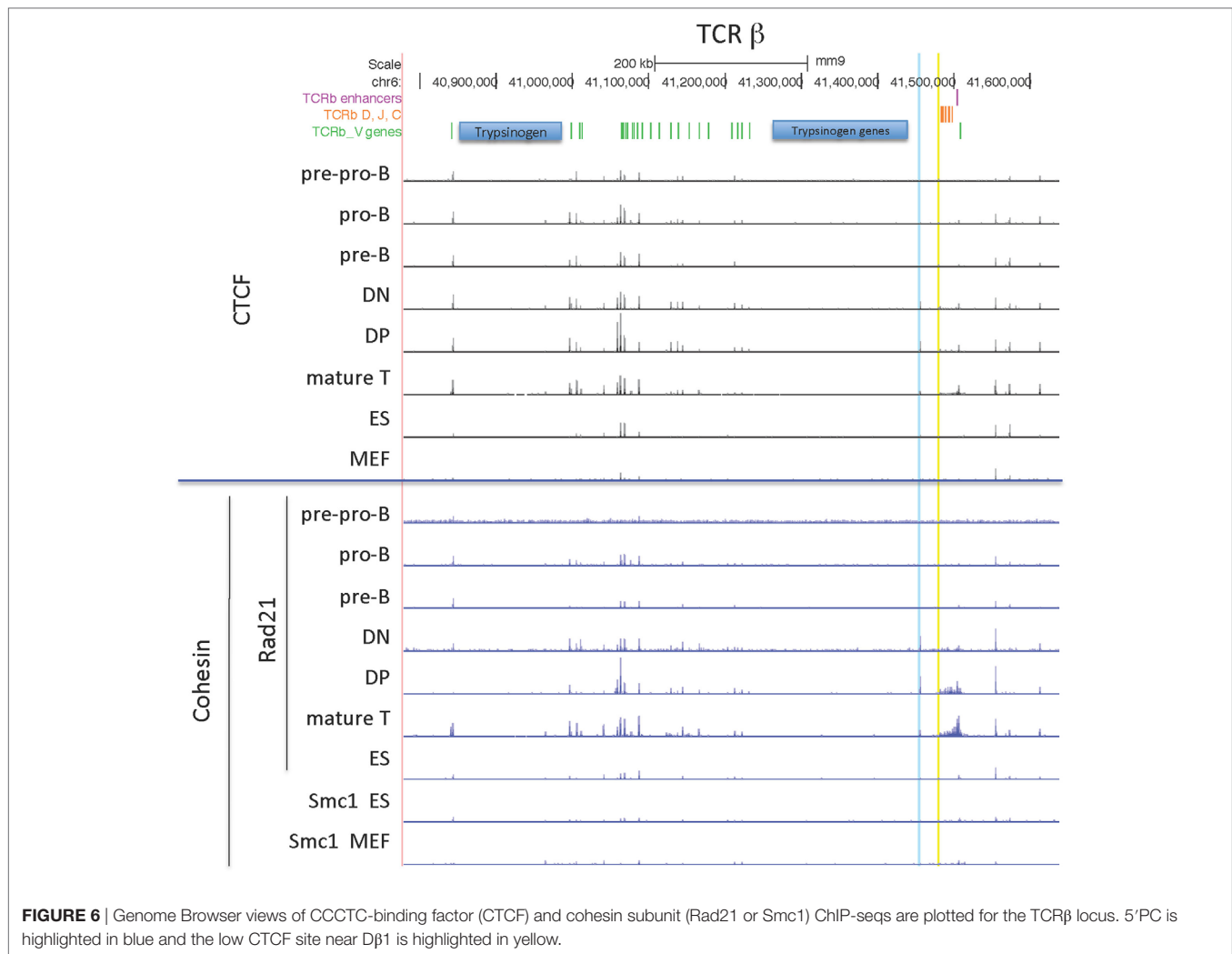
### TCR $\gamma$ Locus

The TCR $\gamma$  locus is a small locus, encompassing 200 kb, and is composed of three clusters of V-J-C genes, one of which contains only pseudogenes (**Figure 10**). The main cluster of functional genes contains 4 V $\gamma$  genes that are used sequentially during fetal and neonatal life (70). This cluster has one J $\gamma$  gene, one C $\gamma$  gene, and one enhancer and the entire cluster covers only 41 kb. This main functional cluster is flanked by CTCF sites oriented toward each other that could create a domain *via* a long-range loop

between the 2 CTCF sites as indicated in **Figure 10**. Furthermore, those two CTCF sites have dominant peaks of cohesin as well as CTCF sites bound in all lymphocyte subsets and in ES cells. However, there is some lineage and stage-specificity in that there is low level cohesin binding to this cluster in DN cells, the stage at which this locus rearranges. The TCR $\gamma$  locus is flanked by relatively conserved CTCF sites, but only the main cluster has convergently oriented sites.

### Ig Lambda Locus

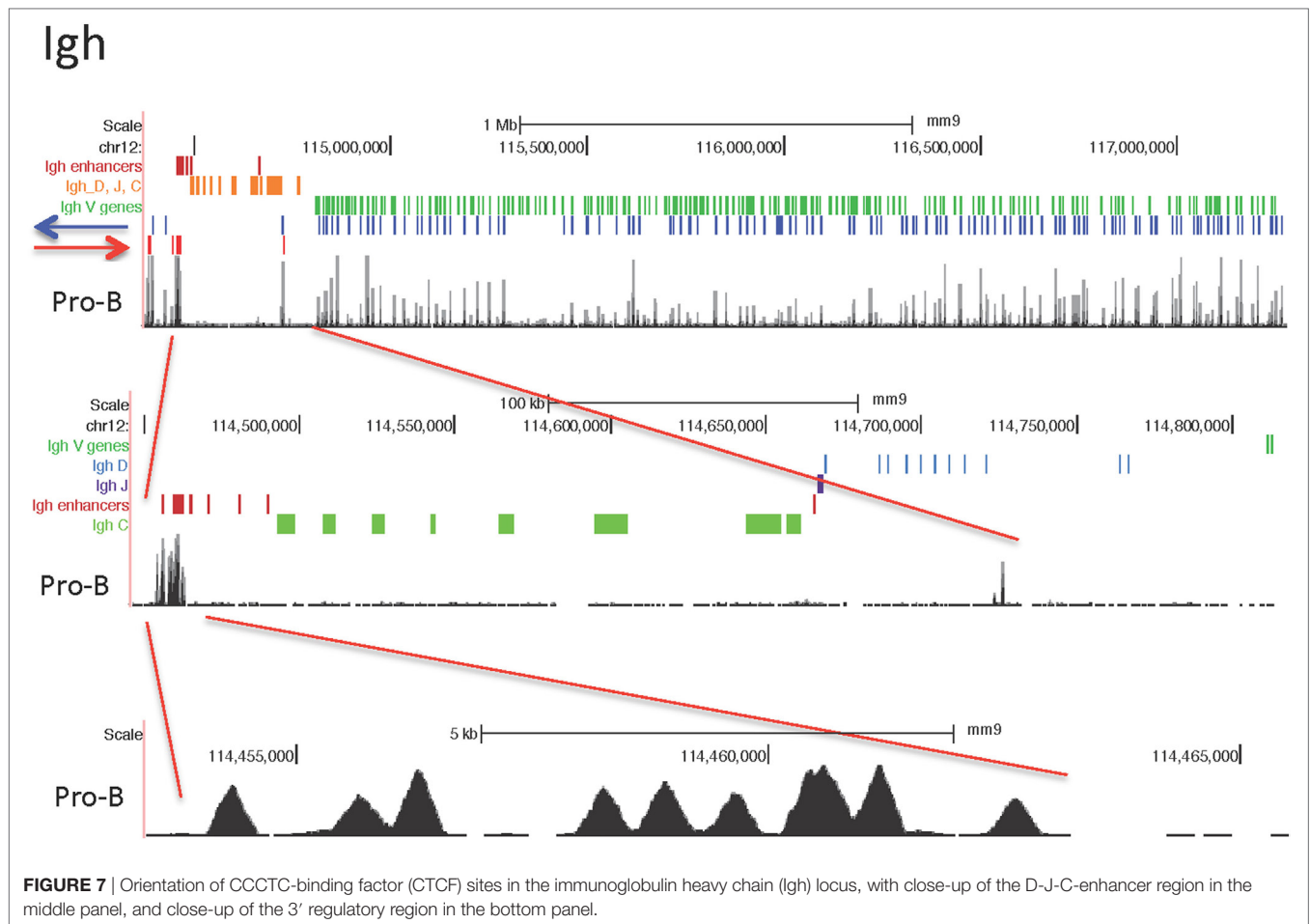
The Ig $\lambda$  locus, like the TCR $\gamma$  locus, is organized in clusters, and V genes only rearrange to J genes within their cluster (**Figure 11**) (71). The first cluster encodes the gene segments encoding the light chains  $\lambda 1$  (V $\lambda 1$  + J $\lambda 1$  + C $\lambda 1$ ) and  $\lambda 3$  (V $\lambda 1$  + J $\lambda 3$  + C $\lambda 3$ ), with  $\lambda 1$  being the most commonly used  $\lambda$  gene. This cluster is flanked by convergent CTCF sites, in which the outside downstream CTCF site is universally bound, but the upstream CTCF site flanking V $\lambda 1$  is only bound in B lineage cells. Likewise, the other cluster, encoding the  $\lambda 2$  and  $\lambda x$  light chains, is also flanked by CTCF sites, but in this case, there is a CTCF site just external to each of the V genes, V $\lambda x$  and V $\lambda 2$ , as well as a site downstream of the cluster.



As with the first cluster, the site downstream of the enhancer is bound in all cell types, whereas the sites flanking the  $\lambda 2$  gene are bound only in B lineage cells, and the site flanking the  $V\lambda x$  is bound only in pre-B cells, the stage at which lambda rearrangement occurs. Occupancy of all these sites is maintained in mature B cells. The CTCF sites flanking the V genes in this second cluster are each in convergent orientation with respect to the CTCF site downstream of the enhancer.  $V\lambda 2$  is the second most frequently rearranging  $V\lambda$  gene, and  $V\lambda x$  is seldom rearranged. This might be because only the CTCF site flanking the  $V\lambda 2$  gene is bound by CTCF and cohesin as cells enter the pre-B cell stage, creating a loop that would put the  $V\lambda 2$  gene close to the enhancer end of the cluster. In pro-B cells, the Igh intronic enhancer is in contact with the CTCF sites flanking DFL16.1 and in the 3' regulatory region (3'RR) (35). If a similar loop is formed here in the second cluster between the flanking CTCF sites and the enhancer that would give preference for rearrangement of the  $V\lambda 2$  gene rather than the  $V\lambda x$  gene, which is what is actually observed in the repertoire. Cohesin binding in B lineage cells mirrors the CTCF-binding pattern, but there is no cohesin binding in thymocytes, MEFs, or pre-pro-B cells.

## DISCUSSION

The ChIP-seq data presented here show that the CTCF and cohesin bound sites within the four largest AgR loci can be divided into two categories. All four large AgR loci have CTCF sites between V genes and D-J genes, and some have CTCF sites near the enhancers that are usually located downstream of J or C genes and, thus clustered at one end of the locus. With some exceptions, the CTCF sites in these very small D-J-C-enhancer regions of each locus tend to be bound by CTCF in ES cells, and sometimes in MEF, the two non-lymphoid cell types we examined. Some of these CTCF sites have been reported to have insulator function, preventing the spread of active chromatin and keeping germline transcription of proximal V genes low, and deletion of these sites results in preferential rearrangement of proximal V genes (72, 73). Despite having important roles in creating a diverse repertoire of AgR using the full range of functional V genes throughout the AgR locus, these CTCF sites show little lymphocyte specificity in binding. However, not all of these sites have cohesin bound, and thus it is unlikely that these sites without cohesin bound would make effective long-range loops.

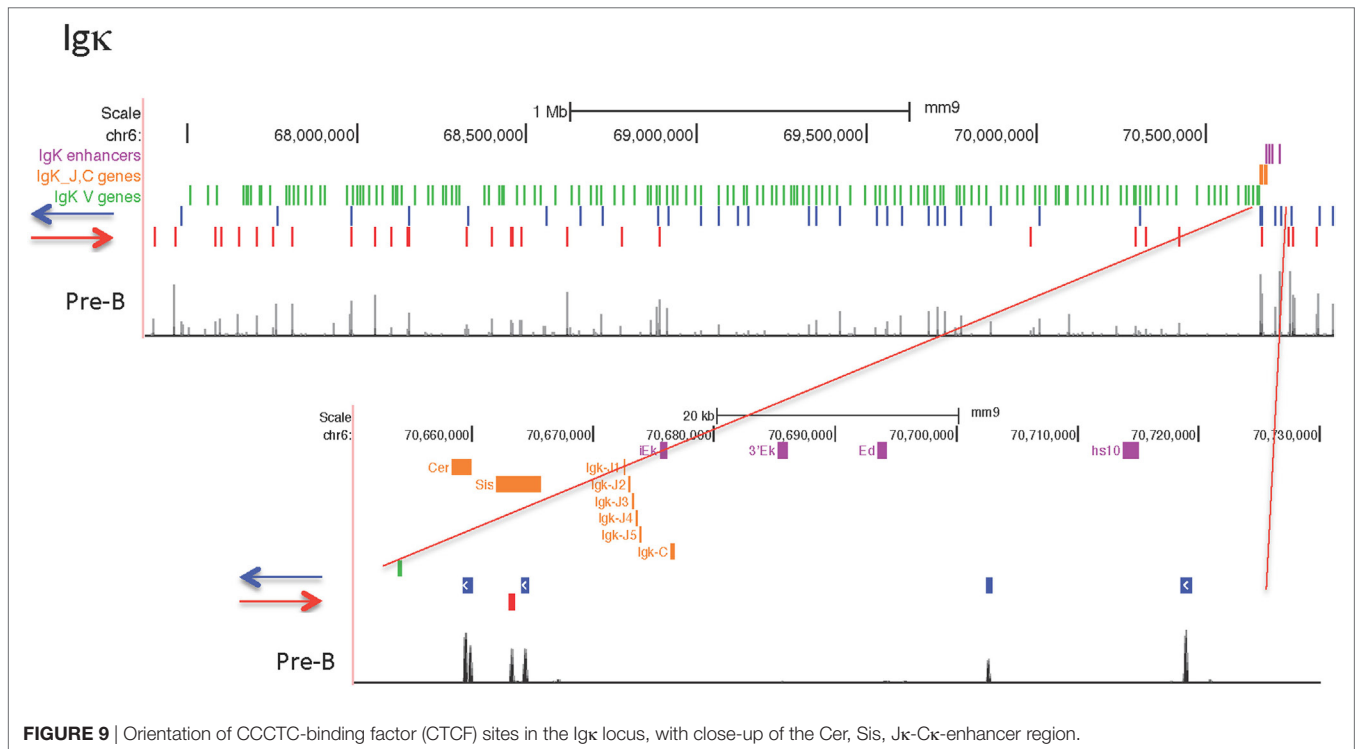
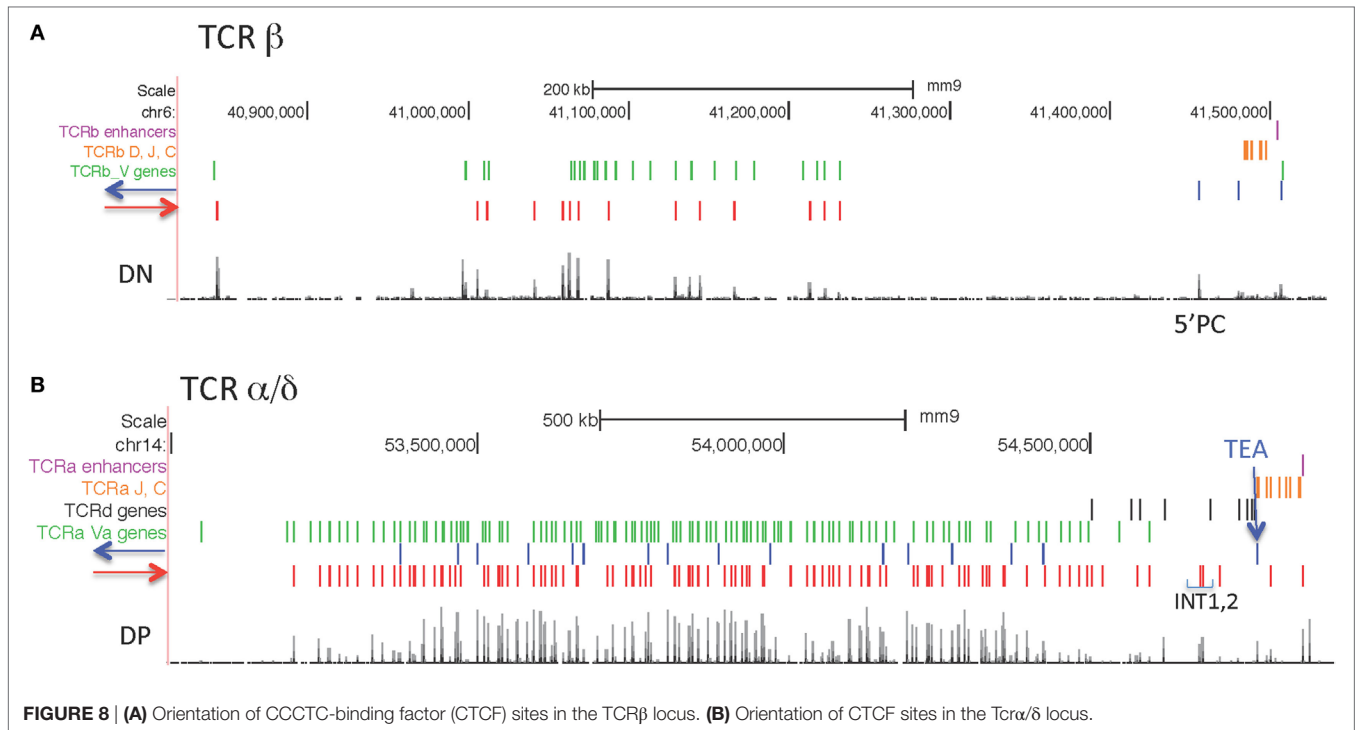


By contrast, the majority of the CTCF sites within the large V gene portion of each of those loci, which comprise >85% of the total size of each locus, show lymphocyte specificity in that very few sites are bound in ES cells or MEFs. However, only the Igh locus shows both lineage (T vs. B cell) and also developmental stage-specificity (pre-B vs. pro-B) for the majority of its bound CTCF sites. This enhanced specificity of CTCF binding may be because, of the four large loci, this is the only one that does not undergo any rearrangement at the first stage of T or B cell differentiation (DN and pro-B, respectively), and thus pre-B cells are characteristic of a later stage of lymphocyte differentiation. ES cells, although having few CTCF sites bound in the AgR loci, often show more binding than the end-stage differentiated MEFs, suggesting that, just as ES cells often have low level transcription of a wide variety of cell type-specific genes, their CTCF binding, although limited, becomes more restricted as ES cells differentiate into non-lymphoid cells. Since DNA methylation has been reported to inhibit CTCF binding, it is possible that the CTCF sites in the AgR loci might be CpG methylated in non-lymphocytes, and only become demethylated in an early lymphoid progenitor stage (74–76). However, the majority of the CTCF sites in the Ig loci do not have CpGs at critical sites within their CTCF motif so this is unlikely to be the reason for the increase in binding with differentiation.

Importantly, we show that Rad21, a component of the cohesin complex, shows greater lineage and developmental stage-specificity in general than CTCF for most of the large AgR loci analyzed for both the V portions of the loci as well as for the several of the CTCF sites in the D-J-C-enhancer regions. Since the loop extrusion model proposes that cohesin is critical to the creation of CTCF-mediated loops, the greater developmental specificity of Rad21 binding to AgR loci suggests substantial developmental specificity in AgR looping (56, 67, 77).

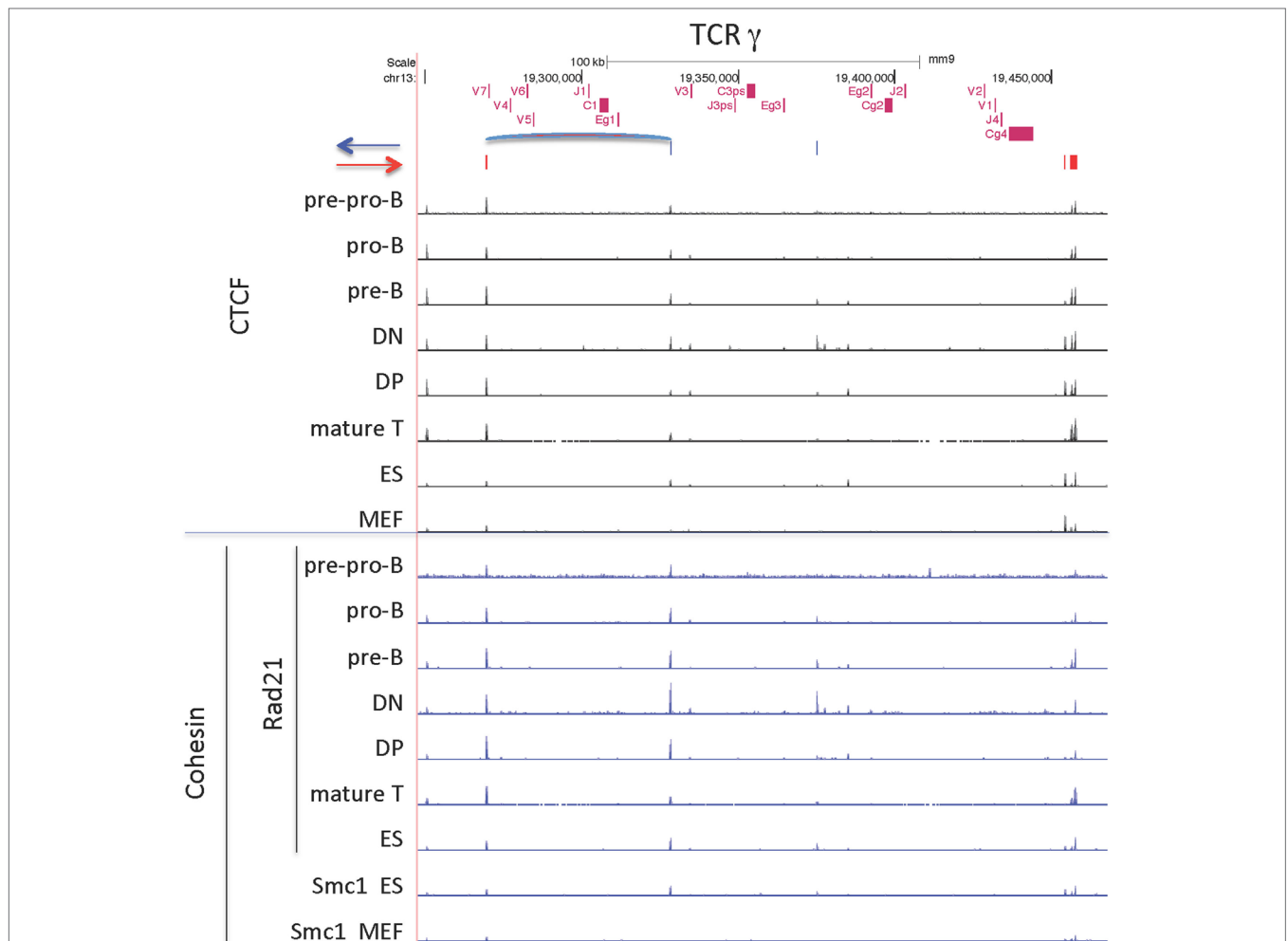
Over the past 3 years, it has become clear that the orientation of the non-palindromic CTCF motif is important in the creation of the long-range loops that mediate most of CTCF's architectural functions as well as its insulator and enhancer-blocking functions (10, 13, 64, 78). The genome is organized into TADs, and these topological domains are bounded in almost all cases by convergently oriented (facing each other) CTCF sites. The loop extrusion model posits a critical role for cohesin as a loop extruder in the creation of these TAD boundaries (20, 23). Within TADs, subTADs can also be formed which may or may not have CTCF at their boundaries. In many cases, genome-wide CTCF sites are found near enhancers and may help stabilize enhancer–promoter interactions (12). The orientation of CTCF sites forming subTADs is less clear, with more examples of tandem orientation loops (64). The Ig loci,





and likely the TCR loci, are present within a single TAD, but the subTAD boundaries are less well defined (43, 66, 79). The Igh locus has at least 2–3 subTADs, and the Igk locus has a complex 3D chromatin structure as well (43, 66, 79). Within the AgR loci, although it might be hypothesized that the orientation

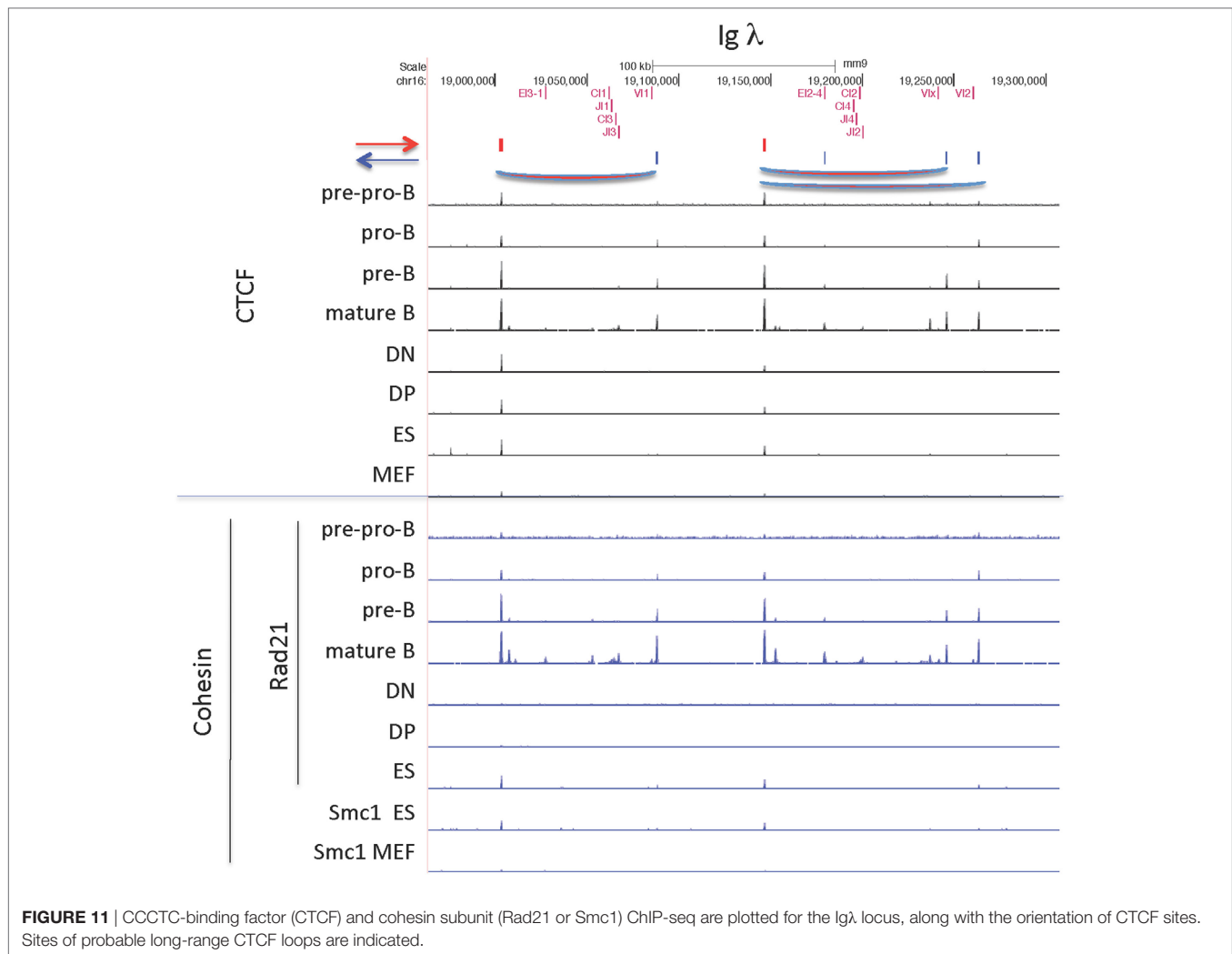
of CTCF sites creating long-range loops may be likely to be convergent, there are already documented exceptions. An ectopic CTCF site was knocked-in in between TCR $\delta$  D-J genes and the downstream Trdv5 gene (80). Surprisingly, this ectopic site was oriented toward the convergent TEA CTCF site, but



**FIGURE 10** | CCCTC-binding factor (CTCF) and cohesin subunit (Rad21 or Smc1) ChIP-seq are plotted for the TCR $\gamma$  locus, along with the orientation of CTCF sites. Sites of the probable long-range CTCF loop is indicated.

it showed preferential interactions with the tandemly oriented INT2 CTCF sites. Similarly, the CTCF site adjacent to E $\alpha$  at the far end of the TCR $\alpha/\delta$  locus is oriented away from the rest of the locus, yet it interacts with many sites within the TCR $\alpha/\delta$  locus (40, 67). Also, in the TCR $\beta$  locus, the CTCF site near E $\beta$  has been reported to interact very strongly with the poorly occupied CTCF site just upstream of the promoter of D $\beta$ 1, PD $\beta$ 1, even though those CTCF sites are in tandem orientation. However, the CTCF site near PD $\beta$ 1 is in the same HindIII fragment as PD $\beta$ 1, so it is not clear if those 3C experiments are reveal CTCF–CTCF interactions or CTCF interactions with the PD $\beta$ 1 promoter itself (81). We have previously documented that E $\mu$  is recruited to the interacting CTCF sites at IGCR1 and 3'RR, and the same could be happening here (36). Taken together, it appears that there are several examples of non-convergent CTCF-mediated loops within AgR loci, although it cannot be ruled out that some of these exceptions are mediated in part by enhancers or promoters rather than by CTCF–CTCF loops.

The Igh locus is the subject of the most detailed 3D-FISH analyses. The Igh locus has been demonstrated to be in a rosette-like structure, with three rosettes present in pre–pro-B cells that further contract into one in pro-B cells, the stage at which rearrangement of that locus takes place (28). It is likely that the other AgR loci also are composed of many long-range loops, and it is known that all the large AgR loci adopt a more contracted structure before they undergo rearrangement, although not precisely at the initiation time of rearrangement, and in some cases, not throughout the entire locus (30, 31, 33, 34, 82). We and others had proposed that the many CTCF sites within the Igh locus could create this rosette structure (35, 36, 83). We now know that all of the bound CTCF sites within the 2.5 Mb Vh region portion of the locus are oriented toward the D–J–C regions, and at the far 3' end of the locus are a cluster of nine CTCF sites known as the 3' regulatory region (3'RR), all of which point toward the V regions (Figure 8) (66, 84). The only other CTCF sites in the Igh locus are a pair of CTCF sites located ~3.1 and 5.7 kb 5' of the most J-distal functional D gene, DFL16.1 (35). One of these sites



is oriented toward the V genes, while the other one is oriented toward the 3'RR (72, 84). We had proposed several years ago that a loop could be made between these CTCF sites adjacent to DFL16 with those of the 3'RR to create a looped domain containing all of the D and J gene segments but excluding all of the V genes (35). We further proposed that this could allow D to J joining without V to DJ joining, and thus would facilitate ordered rearrangement. Our 3C studies and those of others confirmed the existence of this CTCF loop (36, 72). Since all of the bound CTCF sites within the 2.8 Mb V region are oriented toward the nine sites at the 3'RR, it has more recently been proposed that the rosette structure at the  $Igh$  locus is formed by CTCF sites throughout the  $Vh$  portion of the locus each directly interacting with the 3'RR, which has been termed a superanchor (66). Therefore, in this model, there are no CTCF-mediated loops between pairs of CTCF sites within the 2.8 Mb V region portion of the  $Igh$  locus itself. However, 4C analysis by Medvedovic et al. indicated that there is a continuum of loops across the entire  $Igh$  locus, while the 5C studies of Montefiori et al. give defined boundaries to the sub-domains of the  $Igh$  locus (60, 79). Some of these loops could be created by elements other than CTCF, but

to the extent that the loops in either study are CTCF-mediated, the role of convergent vs. tandem CTCF loops remains unclear for the  $Igh$  locus.

The much smaller  $TCR\beta$  locus, such as the  $Igh$  locus, has all of the CTCF sites within the V region part of the locus oriented toward the D-J gene segments (Figure 9). There are three CTCF sites in the D-J portion of the locus, all facing the many CTCF sites in the V portion. The tallest peak of the three (called 5'PC) is 27 kb upstream of D $\beta$ 1.1, and was described as being essential for long-range interactions of the D genes with distal V genes and for rearrangements of those distal V genes (56). A much lower CTCF peak is located ~3 kb upstream of D $\beta$ 1.1, and its deletion results in the spread of the active chromatin marks up to 5'PC, and also prevents 5'PC from making long-range interactions with distal V $\beta$  genes (56). However, in wild-type DN cells, the active histone marks extend ~6 kb upstream of D $\beta$ 1.1, stopping at an long-terminal repeat (LTR) element which appears to be the barrier element (85), so the relationship of the LTR with the weak CTCF site 3 kb downstream of the LTR for the barrier function is not clear. Nonetheless, it is clear that these two CTCF sites, but not the E $\beta$  enhancer, are required

for the long-range looping of distal half of the V locus to the J regions.

Igk locus has many CTCF sites within the 3.1 Mb V region portion of the locus oriented in both directions, although most of the sites in the proximal half of the locus are oriented away from the CTCF sites in the V-J intervening region, and most of those are weakly if at all bound in pro-B cells, although they are bound in IL7-cultured pro-B cells. The orientation pattern would predict that the CTCF sites at Cer would predominantly interact with sites in the distal half of the locus if convergent orientation were important, and we know that Cer is essential for locus contraction (55). Since the Igk locus is already contracted, as measured by 3D-FISH, even in non-IL7-cultured pro-B cells, the CTCF sites that are bound in pro-B cells in the distal half of the locus that are also oriented toward the CTCF sites in Cer may well be critical for creating long-range loops facilitating locus contraction (32). Hi-C analysis of IL7-cultured RAG<sup>-/-</sup> pro-B cells shows many interactions within the Igk locus, but there is a paucity of interactions from Cer/Sis/iEκ region to a large portion of the proximal half of the locus, from ~Vκ12-46 to Vκ4-83. Although CTCF sites throughout the locus are bound in IL7-cultured pro-B cells, this portion of the locus with the fewest Hi-C interactions within the proximal part of the locus lack any CTCF sites oriented toward Cer/Sis (43). However, the Hi-C does show interactions of Cer/Sis/iEκ region with the region containing the first 41 Vκ genes, and there are bound CTCF sites oriented in the proper convergent orientation in that region. By contrast, 4C analysis of fresh *ex vivo* RAG<sup>-/-</sup> pro-B cells show many fewer sites of interaction with Sis, iEκ, and 3'Eκ at the 3' half of the locus despite the presence of some CTCF sites oriented toward Cer, consistent with the minimal binding of CTCF to the proximal half of the locus in *ex vivo* pro-B cells (77). Our data indicating that IL7 culture of pro-B cells results in increased binding of CTCF and cohesin to the proximal half of the locus, and that it induces epigenetic and transcriptional characteristics of pre-B cells (Figure 4), may partially explain that discrepancy.

Long-range looping interactions are not all mediated by CTCF and cohesin. It is well established that enhancers and promoters make looping interactions, and indeed gene activation and transcription are critical in forming chromosomal compartments, structures which are not dependent upon CTCF (46, 86–88). AgR loci undergo robust non-coding RNA transcription, called “germline transcription,” through the J and constant regions at the particular stage of differentiation when that locus is undergoing V(D)J recombination (89). In addition, there is both sense and antisense, genic and intergenic, non-coding transcription happening, usually at low levels, throughout the large V region portions of each AgR locus (61, 90, 91). Thus, it is possible that germline transcription can result in a change in the topology of the AgR loci since multiple transcribing regions within an AgR loci may well be located in a transcription factory where other regions of the AgR loci, especially the J-C regions which undergo robust germline transcription, may also be located (61, 92). Also, non-coding RNA has been shown to be capable of making large scale alterations in chromatin structure by allowing the recruitment of CTCF

to sites which otherwise did not have it bound, creating long-range CTCF-mediated interactions (93). Thus, CTCF and its partner cohesin are not the only factors that are competent to produce long-range looping and to contribute to 3D chromatin conformation in the AgR loci.

In order to create a diverse repertoire of antibodies or TCRs, the AgR loci become more compact at the time of rearrangement to bring the many V genes, spread over large genomic space, into contact with the (D)J genes to which one V gene will rearrange. We have previously shown that CTCF is important for the contracted structure found at the Igh locus in pro-B cells, since knock-down of CTCF shows a more extended Igh locus (36). However, it is known that the transcription factors Pax5 and YY1 are essential for making long-range loops essential for locus contraction (79, 94) and we have data for the influence of novel enhancer elements in creating long-range loops within the Igk locus (EB-M and AF, unpublished data). Furthermore, once CTCF and cohesin become bound at any of the AgR loci, we show here that the pattern of binding remains stable from pre-B to mature B or from DP thymocytes to mature T cells even though at least for the Igh and TCRβ loci, it has been demonstrated that those loci de-compact at the next stage of differentiation. Hence, we hypothesize that the onset of lymphocyte-specific CTCF/cohesin binding throughout the V region portion of the AgR loci in lymphocytes can set up an early level of 3D structure for a locus. However, it is likely that transcription factors, possibly bound to novel or traditional enhancers, will ultimately determine the presence or absence of a contracted AgR locus, since it is clearly not the case that the CTCF/cohesin binding pattern reverts to that of pre-pro-B cells or of non-lineage-specific binding after each locus has finished V(D)J rearrangement.

In order to efficiently utilize the large number of V genes in each loci, it seems reasonable that there should be heterogeneity in the long-range looping interactions that bring the V genes close to the (D)J gene such that different V genes are closest to the (D)J gene in the population of precursor lymphocytes undergoing V(D)J recombination. This need for heterogeneity could suggest that long-range loops are constantly forming and reassembling in each precursor lymphocyte to create a structure in which different V genes are closest to the (D)J genes at different times, and there is evidence that CTCF binding, and by inference chromatin looping, is very dynamic (24). Alternatively, it could be that there is extensive heterogeneity within the population of lymphocyte precursors of long-range loops that might be somewhat more stable, as was suggested in studies on long-term cultured pre-B cells (95). The fact that the height of most of the bound CTCF peaks in the AgR loci are low might suggest that only a subset of the millions of cells analyzed in a ChIP-seq have CTCF bound at those individual sites at any given time, possibly due to weaker affinity for specific CTCF sites. If the majority of CTCF sites within the AgR loci had lower affinity, this could serve an important function in allowing the random loose creation and dissolution of many CTCF-mediated loops, providing the opportunity for a dynamic and changing heterogeneous set of loops in each precursor lymphocyte, which would help facilitate the creation



of diverse immune B and T cell repertoires, utilizing V genes throughout the large AgR loci.

## ETHICS STATEMENT

This study was carried out under approval of our protocol by The Scripps Research Institute's IACUC.

## AUTHOR CONTRIBUTIONS

SL performed all of the bioinformatic analyses. SL, EB-M, and AF analyzed all of the data. EB-M performed some of the ChIP. H-YS and MK provided the data in Figure S2B in Supplementary Material. AF and EB-M prepared the figures. AF, EB-M, and MK wrote the manuscript with input from all the authors, especially SL.

## REFERENCES

- Filippova GN, Fagerlie S, Klenova EM, Myers C, Dehner Y, Goodwin G, et al. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol Cell Biol* (1996) 16:2802–13. doi:10.1128/MCB.16.6.2802
- Bell AC, West AG, Felsenfeld G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* (1999) 98:387–96. doi:10.1016/S0092-8674(00)81967-4
- Gerasimova TI, Corces VG. Boundary and insulator elements in chromosomes. *Curr Opin Genet Dev* (1996) 6:185–92. doi:10.1016/S0959-437X(96)80049-9
- Yusufzai TM, Tagami H, Nakatani Y, Felsenfeld G. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol Cell* (2004) 13:291–8. doi:10.1016/S1097-2765(04)00029-2
- Gerasimova TI, Byrd K, Corces VG. A chromatin insulator determines the nuclear localization of DNA. *Mol Cell* (2000) 6:1025–35. doi:10.1016/S1097-2765(00)00101-5
- Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, et al. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev* (2006) 20:2349–54. doi:10.1101/gad.399506
- Hou C, Zhao H, Tanimoto K, Dean A. CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proc Natl Acad Sci U S A* (2008) 105:20398–403. doi:10.1073/pnas.0808506106
- 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
- Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* (2012) 485:381–5. doi:10.1038/nature11049
- Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* (2014) 159:1665–80. doi:10.1016/j.cell.2014.11.021
- Phillips JE, Corces VG. CTCF: master weaver of the genome. *Cell* (2009) 137:1194–211. doi:10.1016/j.cell.2009.06.001
- Ong CT, Corces VG. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet* (2014) 15:234–46. doi:10.1038/nrg3663
- Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, et al. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep* (2015) 10:1297–309. doi:10.1016/j.celrep.2015.02.004
- Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res* (2009) 19(1):24–32. doi:10.1101/gr.082800.108

## ACKNOWLEDGMENTS

We acknowledge the expert work of previous lab members, including Jiyoti Verma-Gaur, Stephanie Degner, Nancy Choi, Amy Baumgart, and Tim Wong.

## FUNDING

This work was supported by NIH grants R56AI119092 and R01 AI082918 (to AF) and R01 GM41052 (to MK).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://www.frontiersin.org/articles/10.3389/fimmu.2018.00425/full#supplementary-material>.

- Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, et al. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* (2007) 128:1231–45. doi:10.1016/j.cell.2006.12.048
- Parelho V, Hadjir 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
- Rubio ED, Reiss DJ, Welch PL, Distchele CM, Philippova 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
- 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
- Rao SS, Huang SC, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon KR, et al. Cohesin loss eliminates all loop domains. *Cell* (2017) 171:305–320e324. doi:10.1016/j.cell.2017.09.026
- Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of chromosomal domains by loop extrusion. *Cell Rep* (2016) 15:2038–49. doi:10.1016/j.celrep.2016.04.085
- Sanborn AL, Rao SS, Huang SC, Durand NC, Huntley MH, Jewett AI, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A* (2015) 112:E6456–65. doi:10.1073/pnas.1518552112
- Nasmyth K. Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annu Rev Genet* (2001) 35:673–745. doi:10.1146/annurev.genet.35.102401.091334
- Goloborodko A, Marko JF, Mirny LA. Chromosome compaction by active loop extrusion. *Biophys J* (2016) 110:2162–8. doi:10.1016/j.bpj.2016.02.041
- Hansen AS, Pustova I, Cattoglio C, Tjian R, Darzacq X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* (2017) 6. doi:10.7554/eLife.25776
- Alt FW, Yancopoulos GD, Blackwell TK, Wood C, Thomas E, Boss M, et al. Ordered rearrangement of immunoglobulin heavy chain variable region segments. *EMBO J* (1984) 3:1209–19.
- de Pooter RF, Kee BL. E proteins and the regulation of early lymphocyte development. *Immunol Rev* (2010) 238:93–109. doi:10.1111/j.1600-065X.2010.00957.x
- Xiong N, Raulet DH. Development and selection of gammadelta T cells. *Immunol Rev* (2007) 215:15–31. doi:10.1111/j.1600-065X.2006.00478.x
- 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
- 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

30. 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
31. Shih HY, Krangel MS. Distinct contracted conformations of the Tcra/Tcrd locus during Tcra and Tcrd recombination. *J Exp Med* (2010) 207:1835–41. doi:10.1084/jem.20100772
32. Rother MB, Palstra RJ, Jhunjhunwala S, van Kester KA, van IWF, Hendriks RW, et al. Nuclear positioning rather than contraction controls ordered rearrangements of immunoglobulin loci. *Nucleic Acids Res* (2016) 44:175–86. doi:10.1093/nar/gkv928
33. 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
34. Carico ZM, Roy Choudhury K, Zhang B, Zhuang Y, Krangel MS. Tcrd rearrangement redirects a processive Tcra recombination program to expand the Tcra repertoire. *Cell Rep* (2017) 19:2157–73. doi:10.1016/j.celrep.2017.05.045
35. 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. doi:10.4049/jimmunol.182.1.44
36. 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
37. Nussenzweig MC, Shaw AC, Sinn E, Danner DB, Holmes KL, Morse HC III, et al. Allelic exclusion in transgenic mice that express the membrane form of immunoglobulin mu. *Science* (1987) 236:816–9. doi:10.1126/science.3107126
38. Kleiman E, Jia H, Loguercio S, Su AI, Feeney AJ. YY1 plays an essential role at all stages of B-cell differentiation. *Proc Natl Acad Sci U S A* (2016) 113:E3911–20. doi:10.1073/pnas.1606297113
39. 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
40. Shih HY, Verma-Gaur J, Torkamani A, Feeney AJ, Galjart N, Krangel MS. Tcra gene recombination is supported by a Tcra enhancer- and CTCF-dependent chromatin hub. *Proc Natl Acad Sci U S A* (2012) 109:E3493–502. doi:10.1073/pnas.1214131109
41. Canela A, Maman Y, Jung S, Wong N, Callen E, Day A, et al. Genome organization drives chromosome fragility. *Cell* (2017) 170:507–521e518. doi:10.1016/j.cell.2017.06.034
42. Kieffer-Kwon KR, Nimura K, Rao SSP, Xu J, Jung S, Pekowska A, et al. Myc regulates chromatin decompaction and nuclear architecture during B cell activation. *Mol Cell* (2017) 67:566–578e510. doi:10.1016/j.molcel.2017.07.013
43. Lin YC, Benner C, Mansson R, Heinz S, Miyazaki K, Miyazaki M, et al. Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nat Immunol* (2012) 13:1196–204. doi:10.1038/ni.2432
44. Shen Y, Yue F, McCleary DF, Ye Z, Eadsall L, Kuan S, et al. A map of the cis-regulatory sequences in the mouse genome. *Nature* (2012) 488:116–20. doi:10.1038/nature11243
45. Downen JM, Bilodeau S, Orlando DA, Hubner MR, Abraham BJ, Spector DL, et al. Multiple structural maintenance of chromosome complexes at transcriptional regulatory elements. *Stem Cell Reports* (2013) 1:371–8. doi:10.1016/j.stemcr.2013.09.002
46. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* (2010) 467:430–5. doi:10.1038/nature09380
47. Andrews S. *FastQC: a quality control tool for high throughput sequence data*. (2015). Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
48. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* (2010) 38:576–89. doi:10.1016/j.molcel.2010.05.004
49. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* (2009) 10:R25. doi:10.1186/gb-2009-10-3-r25
50. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* (2008) 9:R137. doi:10.1186/gb-2008-9-9-r137
51. Bardet AF, He Q, Zeitlinger J, Stark A. A computational pipeline for comparative ChIP-seq analyses. *Nat Protoc* (2011) 7:45–61. doi:10.1038/nprot.2011.420
52. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* (2008) 5:R80. doi:10.1186/gb-2004-5-10-r80
53. Carico Z, Krangel MS. Chromatin dynamics and the development of the TCRalpha and TCRdelta repertoires. *Adv Immunol* (2015) 128:307–61. doi:10.1016/bs.ai.2015.07.005
54. Xiang Y, Zhou X, Hewitt SL, Skok JA, Garrard WT. A multifunctional element in the mouse Igkappa locus that specifies repertoire and Ig loci subnuclear location. *J Immunol* (2011) 186:5356–66. doi:10.4049/jimmunol.1003794
55. Xiang Y, Park SK, Garrard WT. V kappa gene repertoire and locus contraction are specified by critical DNase I hypersensitive sites within the V kappa-J kappa intervening region. *J Immunol* (2013) 190:1819–26. doi:10.4049/jimmunol.1203127
56. Majumder K, Koues OI, Chan EA, Kyle KE, Horowitz JE, Yang-Iott K, et al. Lineage-specific compaction of Tcrb requires a chromatin barrier to protect the function of a long-range tethering element. *J Exp Med* (2015) 212:107–20. doi:10.1084/jem.20141479
57. Plasschaert RN, Vigneau S, Tempera I, Gupta R, Maksimoska J, Everett L, et al. CTCF binding site sequence differences are associated with unique regulatory and functional trends during embryonic stem cell differentiation. *Nucleic Acids Res* (2014) 42:774–89. doi:10.1093/nar/gkt910
58. Lin YC, Jhunjhunwala S, Benner C, Heinz S, Welinder E, Mansson R, et al. A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat Immunol* (2010) 11:635–43. doi:10.1038/ni.1891
59. Allman D, Sambandam A, Kim S, Miller JB, Pagan A, Well D, et al. Thymopoiesis independent of common lymphoid progenitors. *Nat Immunol* (2003) 4:168–74. doi:10.1038/ni878
60. Medvedovic J, Ebert A, Tagoh H, Tamir IM, Schwickert TA, Novatchkova M, et al. Flexible long-range loops in the VH gene region of the Igh locus facilitate the generation of a diverse antibody repertoire. *Immunity* (2013) 39:229–44. doi:10.1016/j.immuni.2013.08.011
61. Verma-Gaur J, Torkamani A, Schaffer L, Head SR, Schork NJ, Feeney AJ. Noncoding transcription within the Igh distal VH 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
62. Yancopoulos GD, Malynn BA, Alt FW. Developmentally regulated and strain-specific expression of murine V<sub>H</sub> gene families. *J Exp Med* (1988) 168:417–35. doi:10.1084/jem.168.1.417
63. Van Ness BG, Weigert M, Coleclough C, Mather EL, Kelley DW, Perry RP. Transcription of the unrearranged mouse C<sub>k</sub> locus: sequence of the initiation region and comparison of activity with a rearranged V<sub>k</sub>-C<sub>k</sub> gene. *Cell* (1981) 27:593–602. doi:10.1016/0092-8674(81)90401-3
64. Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, et al. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* (2015) 163:1611–27. doi:10.1016/j.cell.2015.11.024
65. Phillips-Cremmins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* (2013) 153:1281–95. doi:10.1016/j.cell.2013.04.053
66. Benner C, Isoda T, Murre C. New roles for DNA cytosine modification, eRNA, anchors, and superanchors in developing B cell progenitors. *Proc Natl Acad Sci U S A* (2015) 112:12776–81. doi:10.1073/pnas.1512995112
67. Chen L, Carico Z, Shih HY, Krangel MS. A discrete chromatin loop in the mouse Tcra-Tcrd locus shapes the TCRdelta and TCRalpha repertoires. *Nat Immunol* (2015) 16:1085–93. doi:10.1038/ni.3232
68. Villey I, Caillol D, Selz F, Ferrier P, de Villartay J-P. Defect in rearrangement of the most 5' TCR-Ja following targeted deletion of T early a (TEA): implications for TCR a locus accessibility. *Immunity* (1996) 5:331–42. doi:10.1016/S1074-7613(00)80259-9

69. Hawwari A, Bock C, Krangel MS. Regulation of T cell receptor alpha gene assembly by a complex hierarchy of germline Jalpha promoters. *Nat Immunol* (2005) 6:481–9. doi:10.1038/ni1189
70. Goldman JB, Spencer DM, Raulet DH. Ordered rearrangement of variable region genes of the T cell receptor g locus correlates with transcription of the unrearranged genes. *J Exp Med* (1993) 177:729–39. doi:10.1084/jem.177.3.729
71. Sanchez P, Nadel B, Cazenave PA. V lambda-J lambda rearrangements are restricted within a V-J-C recombination unit in the mouse. *Eur J Immunol* (1991) 21:907–11. doi:10.1002/eji.1830210408
72. 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
73. Xiang Y, Park SK, Garrard WT. A major deletion in the V kappa-J kappa intervening region results in hyperelevated transcription of proximal V kappa genes and a severely restricted repertoire. *J Immunol* (2014) 193:3746–54. doi:10.4049/jimmunol.1401574
74. Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res* (2012) 22:1680–8. doi:10.1101/gr.136101.111
75. Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* (2000) 405:482–5. doi:10.1038/35013100
76. Filippova GN, Thienes CP, Penn BH, Cho DH, Hu YJ, Moore JM, et al. CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. *Nat Genet* (2001) 28:335–43. doi:10.1038/ng570
77. Stadhouders R, de Bruijn MJ, Rother MB, Yuvaraj S, Ribeiro de Almeida C, Kolovos P, et al. Pre-B cell receptor signaling induces immunoglobulin kappa locus accessibility by functional redistribution of enhancer-mediated chromatin interactions. *PLoS Biol* (2014) 12:e1001791. doi:10.1371/journal.pbio.1001791
78. 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
79. Montefiori L, Wuerffel R, Roqueiro D, Lajoie B, Guo C, Gerasimova T, et al. Extremely long-range chromatin loops link topological domains to facilitate a diverse antibody repertoire. *Cell Rep* (2016) 14:896–906. doi:10.1016/j.celrep.2015.12.083
80. Chen L, Zhao L, Alt FW, Krangel MS. An ectopic CTCF binding element inhibits Tcrd rearrangement by limiting contact between Vdelta and Ddelta gene segments. *J Immunol* (2016) 197:3188–97. doi:10.4049/jimmunol.1601124
81. Rawat P, Jalan M, Sadhu A, Kanaujia A, Srivastava M. Chromatin domain organization of the TCRB locus and its perturbation by ectopic CTCF binding. *Mol Cell Biol* (2017) 37. doi:10.1128/MCB.00557-16
82. Skok JA, Gisler R, Novatchkova M, Farmer D, de Laat W, Busslinger M. Reversible contraction by looping of the Tcra and Tcrb loci in rearranging thymocytes. *Nat Immunol* (2007) 8:378–87. doi:10.1038/ni1448
83. Bossen C, Mansson R, Murre C. Chromatin topology and the regulation of antigen receptor assembly. *Annu Rev Immunol* (2012) 30:337–56. doi:10.1146/annurev-immunol-020711-075003
84. Lin SG, Guo C, Su A, Zhang Y, Alt FW. CTCF-binding elements 1 and 2 in the Igh intergenic control region cooperatively regulate V(D)J recombination. *Proc Natl Acad Sci U S A* (2015) 112:1815–20. doi:10.1073/pnas.1424936112
85. Carabana J, Watanabe A, Hao B, Krangel MS. A barrier-type insulator forms a boundary between active and inactive chromatin at the murine TCRbeta locus. *J Immunol* (2011) 186:3556–62. doi:10.4049/jimmunol.1003164
86. Rowley MJ, Nichols MH, Lyu X, Ando-Kuri M, Rivera ISM, Hermetz K, et al. Evolutionarily conserved principles predict 3D chromatin organization. *Mol Cell* (2017) 67:837–852e837. doi:10.1016/j.molcel.2017.07.022
87. Schoenfelder S, Clay I, Fraser P. The transcriptional interactome: gene expression in 3D. *Curr Opin Genet Dev* (2010) 20:127–33. doi:10.1016/j.gde.2010.02.002
88. Nora EP, Goloborodko A, Valton AL, Gibcus JH, Ueberohrn A, Abdennur N, et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* (2017) 169:930–944e922. doi:10.1016/j.cell.2017.05.004
89. Yancopoulos GD, Alt FW. Regulation of the assembly and expression of variable-region genes. *Annu Rev Immunol* (1986) 4:339–68. doi:10.1146/annurev.iy.04.040186.002011
90. Yancopoulos GD, Alt FW. Developmentally controlled and tissue-specific expression of unrearranged V<sub>H</sub> gene segments. *Cell* (1985) 40:271–81. doi:10.1016/0092-8674(85)90141-2
91. 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
92. 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
93. Isoda T, Moore AJ, He Z, Chandra V, Aida M, Denholtz M, et al. Non-coding transcription instructs chromatin folding and compartmentalization to dictate enhancer-promoter communication and T cell fate. *Cell* (2017) 171:103–119e118. doi:10.1016/j.cell.2017.09.001
94. Gerasimova T, Guo C, Ghosh A, Qiu X, Montefiori L, Verma-Gaur J, et al. A structural hierarchy mediated by multiple nuclear factors establishes IgH locus conformation. *Genes Dev* (2015) 29:1683–95. doi:10.1101/gad.263871.115
95. Levin-Klein R, Fraenkel S, Lichtenstein M, Matheson LS, Bartok O, Nevo Y, et al. Clonally stable V kappa allelic choice instructs Ig kappa repertoire. *Nat Commun* (2017) 8:15575. doi:10.1038/ncomms15575

**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.

Copyright © 2018 Loguercio, Barajas-Mora, Shih, Krangel 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) and the copyright owner 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.