A discrete chromatin loop in the mouse Tcra-Tcrd locus shapes the TCRδ and TCRα repertoires

Liang Chen, Zachary Carico, Han-Yu Shih & Michael S Krangel

The locus encoding the T cell antigen receptor (TCR) α-chain and δ-chain (Tcra-Tcrd) undergoes recombination of its variable-diversity-joining (V(D)J) segments in CD4+CD8− double-negative thymocytes and CD4+CD8+ double-positive thymocytes to generate diverse TCRδ repertoires and TCRα repertoires, respectively. Here we identify a chromatin-interaction network in the Tcra-Tcrd locus in double-negative thymocytes that was formed by interactions between binding elements for the transcription factor CTCF. Disruption of a discrete chromatin loop encompassing the D, J and constant (C) segments of Tcrd allowed a single V segment to frequently contact and rearrange to D and J segments and dominate the adult TCRδ repertoire. Disruption of this loop also narrowed the TCRα repertoire, which, we believe, followed as a consequence of the restricted TCRδ repertoire. Hence, a single CTCF-mediated chromatin loop directly regulated TCRδ diversity and indirectly regulated TCRα diversity.

Adaptive immunity depends on highly diverse repertoires of antigen receptors expressed by T lymphocytes and B lymphocytes. This diversity is generated by V(D)J recombination, in which variable (V), diversity (D) and joining (J) segments of genes encoding T cell antigen receptors (TCRs) and immunoglobulins are assembled during the early stages of the development of T lymphocytes and B lymphocytes, respectively. Initiation of this process requires the collaborative function of the recombinase proteins RAG1 and RAG2 (collectively called ‘RAG’ here)1. RAG is thought to bind to a recombination signal sequence (RSS) in a D or J segment within a recombination center and then capture a second RSS to form a synaptic complex2. Within this complex, RAG introduces precise double-strand breaks between gene segments and RSSs. Repair of double-strand breaks by non-homologous end joining results in assembly of antigen-receptor coding and signal joints1.

Diversification of antigen receptors must overcome daunting topological constraints to recruit gene segments for recombination that may be distributed across several megabases (Mb) of DNA. Various studies have shown that loci encoding antigen receptors undergo large-scale conformational changes during lymphocyte development and that these changes bring distant gene segments into proximity. For example, three-dimensional fluorescence in situ hybridization has shown that the immunoglobulin heavy-chain locus (Igh), the immunoglobulin κ-chain locus (Igk), and the Tcra and Tcra-Tcrd loci undergo contraction that coincides with the developmental stages during which V(D)J recombination occurs3-7. Conversely, loci can be extended to terminate V(D)J recombination, as has been documented for Igh and Tcrb3,4. Dynamic regulation of locus conformation ensures that V(D)J recombination occurs in a developmental stage–specific manner and provides the opportunity for distal V segments to compete with proximal V segments to ensure the assembly of diverse antigen-receptor repertoires.

Chromatin-conformation capture (3C) and 3C-based assays have shown that loci encoding antigen receptors are demarcated by chromatin loops that juxtapose distant segments of DNA. Although studies have indicated roles for the transcription factors Pax5 and YY1 in the organization of Igh loops, the chief mediator of chromatin looping at Igh, Igk, Tcra and Tcrb is the CCCTC-binding factor CTCF8-17. CTCF is a highly conserved, ubiquitously expressed, zinc finger–containing transcription factor that binds throughout the genome and mediates long-distance looping between CTCF-binding elements (CBEs)18. CTCF can block, or ‘insulate’, enhancer activity by creating DNA loops that separate enhancers from promoters, or it can facilitate gene expression by creating DNA loops that juxtapose enhancers and promoters. These two mechanisms account for the known roles of CTCF in V(D)J recombination at loci encoding antigen receptors.

At the Igh locus, IGCR1, an intergenic CBE between the variable heavy-chain (VH) and diversity heavy-chain (DH) arrays, insulates DH-proximal VH segments from the influence of the Igh enhancer (Ei)19. When IGCR1 is deleted, rearrangements are biased toward the hyperactive DH-proximal VH segments and become disordered and lineage nonspecific. Intergenic CBEs at the Igk locus similarly insulate proximal Vκ segments from Igk enhancers11,19. At the Tcra-Tcrd locus, CTCF marks many important cis-regulatory elements and, as a result, helps to target the Tera enhancer (Eα) to the Jα promoter (TEA; ‘T early-α’) and to the promoters of Jα-proximal Vα segments. These interactions promote transcription, accessibility and recombination of these Vα and Jα segments14. Emerging genome-wide studies also indicate that CTCF-mediated looping might serve a structural or organizing role rather than a direct gene-regulatory role20-23. The 1.6-Mb Tcra-Tcrd locus displays a complex organization of gene segments and an intricate program of V(D)J recombination that leads to the development of both γδ T lymphocytes and αβ T lymphocytes24. Approximately 100 V segments are distributed...
across 1.5 Mb, with Tcra D, J and constant (C) segments and Tcra J and C segments clustered within the final 0.1 Mb of the locus (called the ‘3’ end’ of the locus here). The majority of V segments rearrange to Iα segments in CD4⁺CD8⁺ double-positive (DP) thymocytes and contribute to the TCRα repertoire. However, only a few V segments rearrange to Dδ and Jδ segments in CD4⁺CD8⁻ double-negative (DN) thymocytes and contribute to the TCRδ repertoire. Several Vδ segments (Trdv1, Trdv2-2, Trdv4 and Trdv5) are positioned proximal to the DδJδ cluster and are thought to be used exclusively for Tcra rearrangement. Others (Trav1-4-dv7, Trav6-7-dv9, Trav4-4-dv10, Trav14d-3-dv8, Trav16d-dv11) and the ‘Trav15-dv6 family’ (our collective term for Trav15-1-dv6-1, Trav15-2-dv6-2, Trav15d-1-dv6d-1 and Trav15d-2-dv6d-2) are more distal, are interspersed among Vα segments and are used as both Vδ segments and Vα segments. How the locus produces a balanced and diverse TCRδ repertoire with representation of proximal and distal Vδ segments is unclear. Here we defined a CTCF-dependent chromatin-interaction network that extended across 0.5 Mb of the Tcra-Tcrd locus in DN thymocytes. We identified two intergenic CBEs, INT1 and INT2, with central roles in this interaction network. INT1 interacted broadly and dynamically across this region of chromatin. However, INT2 interacted specifically with the CBE associated with TEA, forming a high-frequency chromatin loop that segregated Tcra D, J and C segments from most Vδ segments. Mice with deletion of INT1 and INT2 on both alleles (called ‘INT1-2-deficient mice’ here) had a highly restricted TCRδ repertoire that was strongly biased toward Trdv2-2. This Vδ segment is normally segregated from Dδ segments by the INT2-TEA loop but was newly included within the Dδ-containing loop on alleles that lacked INT1 and INT2. Biased Vδ use resulted not from increased accessibility but from increased interactions between Trdv2-2 and Dδ segments. Of note, the TCRα repertoire was also altered in INT1-2-deficient mice, which indicated that heterogeneity of Tcra rearrangement was a diversifier of Tcra rearrangement. Our results indicate that a CTCF-dependent chromatin-interaction network creates diversity of the TCRδ and TCRα repertoires during T cell development.

RESULTS

Mapping long-range interactions at the Tcra-Tcrd locus

Most CBEs at the Tcra-Tcrd locus are constitutively occupied by CTCF in B cells and DN and DP thymocytes14. A majority of these CBEs are associated with cis-regulatory elements, including V-segment promoters, TEA and Eκ. However, we noted two prominent intergenic CBEs, INT1 and INT2, in the ‘3’ portion of the Tcra-Tcrd locus (Fig. 1a). We sought to determine whether these CBEs are woven into a chromatin-interaction network that sets the stage for Tcra rearrangement in DN thymocytes. To map long-range interactions, we performed circular Hi-C experiments using INT1-2-deficient mice (INT1-2-deficient mice) and in INT1-2-deficient mice versus wild-type mice (Fig. 2). The frequency of Vγd use among γδ TCR⁺ thymocytes was threefold higher in INT1-2-deficient mice than in their wild-type counterparts, the development of αβTCR⁺ thymocytes was largely normal in INT1-2-deficient mice, on the basis of staining with antibodies specific to CD4 and CD8 (Fig. 3). DN thymocytes can be subcategorized into four successive developmental stages on the basis of their expression of CD44 and CD25: DN1 (CD25⁺CD44⁺), DN2 (CD25⁺CD44⁺), DN3 (CD3⁺CD44⁺) and DN4 (CD25⁻CD44⁺). The frequency of DN1–DN4 thymocytes was similar in wild-type and INT1-2-deficient mice (Fig. 3a). However, the frequency of γδ T cells in INT1-2-deficient mice was about half that in their wild-type littermates (Fig. 3c). Moreover, the frequency of Vγd use among γδ T cell development was determined by sequential Vγd usage (Supplementary Fig. 2). Therefore, INT1-2-deficient mice displayed defective γδ T cell development and a biased TCRδ repertoire that was heavily skewed toward Vγd. Vγd is a commonly used adult Vγ segment encoded by Travd2-2, the first functional Vγ segments upstream of the INT1-2 region (Fig. 1a). We sought to determine whether the skewed TCRδ repertoire in INT1-2-deficient mice could be attributed to dysregulated Tcra rearrangements by quantifying VDDJ coding joints in genomic DNA samples prepared from DN3 thymocytes. The frequency of Travd2-2 rearrangement was much higher in INT1-2-deficient thymocytes than in wild-type DN3 thymocytes (Fig. 4a,b). However, rearrangement of the related gene Travd2-1 was barely detectable in DN3 thymocytes of either genotype, a result confirmed by PCR with a shared Travd2 primer and a Travd1 primer: 11 of 11 clones of each genotype were Travd2-2 by sequencing (data not shown). Travd1, located 67 kb upstream of Travd2-2, was equally rearranged in wild-type and INT1-2-deficient thymocytes (Fig. 4a,b). However, all other Vδ segments tested, including Trav13-4-dv7, Trav6-7-dv9, Trav16d-dv11, Trav5 and two members of the Trav15-dv6 family (Trav15-1-dv6-1 and Trav15d-1-dv6d-1), were rearranged much less frequently in INT1-2-deficient thymocytes than in wild-type thymocytes (Fig. 4a,b). Therefore, Travd2-2 rearrangements predominated at the expense of other Vδ segments. This bias also extended to incomplete VD rearrangements. Tcra rearrangement is unusual, since it is unordered and VD, DD, and DJ rearrangements all occur. Of note, Travd2-2-to-Travd1-Travd2 rearrangement was
greater in INT1-2-deficient thymocytes than in wild-type thymocytes, whereas Trav15-dv6–to–Trdd1-Trdd2 and Travv5–to–Trdd1-Trdd2 rearrangements were less abundant in INT1-2-deficient thymocytes than in wild-type thymocytes (Fig. 4c). Trdv2-2–to–Trdd1-Trdd2 rearrangements were as frequent in INT1-2-deficient mice as they were in E3-deficient mice (Fig. 4c), in which partial rearrangements predominate31.

To better quantify dysregulation of Tcrd rearrangement in adult thymocytes, we analyzed retention of Tcrd genomic sequences in preparations of total thymocyte DNA. Deletional rearrangement of Tcrd segments in DN thymocytes places intervening signal-joint DNA onto extrachromosomal circles, which are diluted out and lost during pre-TCR–driven cell proliferation. In contrast, excised signal joints from Tcrd rearrangement are retained in DP thymocytes because they are generated after pre-TCR–driven proliferation (the Online Methods section includes further discussion of this point). Thus, retention of genomic DNA in total thymocytes can quantitatively ‘report’ the spectrum of Tcrd-rearrangement events. To ensure accurate quantification of DNA loss due to Tcrd rearrangement, we compared retention of Tcrd sequences to retention of TEA in wild-type thymocytes, because TEA is not excised by Tcrd rearrangements. By measuring the abundance of PCR amplicons situated immediately upstream of the gene segments assessed, we found that wild-type thymocytes had rearranged the Trdd1-Trdd2 and Trdd2-Trdj1 intervals on 96% and 90% of alleles, respectively (Fig. 4d). In addition, approximately 28% of alleles had rearranged V8 segments upstream of Trav1, approximately 42% had rearranged Travv2-2, and another 25% either had not undergone V-to-D rearrangement or had rearranged Trav5 by inversion (which would not result in deletion of the region upstream of Trav1) (Fig. 4d). In contrast, alleles with deletion of INT1-2 displayed impaired Travdd2-Trdj1 rearrangement but enhanced Trav2-2–to–Trdd1-Trdd2 rearrangement relative to that of wild-type alleles (Fig. 4d). Precocious Travv2-2–to–Trdd1-Trdd2 rearrangements might inhibit Travdd2–to–Trdj1 recombination events on alleles lacking INT1-2. Because the amplicon upstream of Travv2-2 was retained on 93% of mutated alleles, whereas that upstream of Travdd1 was retained on only 6% of mutated alleles (Fig. 4d), Travv2-2 seemed to undergo partial VDD rearrangement or complete VDDJ rearrangement on most alleles lacking INT1-2. Deletion of INT1-2 also caused increased rearrangement of Travv2-2 to the most 5’Tα segments in DP thymocytes (Supplementary Fig. 3a). This rearrangement must have occurred on alleles that had not undergone Travv2-2–to–D8 rearrangement in DN thymocytes and might explain slightly lower retention of TEA in INT1-2-deficient thymocytes than in wild-type thymocytes (Fig. 4d). However, we did not detect premature rearrangement of Travv2-2 or proximal V8 segments to Tα segments in INT1-2-deficient DN thymocytes (Supplementary Fig. 3b). Together, these data showed that the INT1-2 genomic region was essential for the generation of a diverse V8 repertoire in DN thymocytes.

V8 use in early fetal thymocytes is distinct from that in adult thymocytes since it is strongly biased toward Travv4 (V81), which is proximal to D8 segments and within the INT2-TEA loop10. We sought to determine whether Tcrd rearrangement in the fetal thymus...
INT1-2 regulates chromatin loop organization

We used 3C and quantitative PCR to investigate whether the deletion of INT1-2 generated an altered landscape of long-distance chromatin interactions. We prepared 3C libraries by digestion with HindIII and then assayed HindIII fragments for interactions with the TEA viewpoint (Fig. 5c,d). Although the allele with deletion of INT1-2 lacked the INT1 HindIII fragment, it retained the portion of the INT2 HindIII fragment that included the primer-binding site. Consistent with the data obtained by 4C and sequencing, the TEA viewpoint strongly interacted with INT2 (fragment xi) on wild-type alleles (Fig. 5d). However, TEA interacted minimally with the residual INT2 fragment on alleles with deletion of INT1-2 and instead interacted frequently with another intergenic CBE, INT3 (fragment vi), located 49 kb upstream of INT1-2 (Supplementary Fig. 4). As a consequence of the INT3-TEA interaction, Trdv2-2 was confined within a new 250-kb loop that included TcrD, J and C segments, Eδ6, Trdv5, several Vδ pseudogenes, and two Vδ gene segments (Trdv4 and Trdv2-1) that rearranged minimally in adult DN thymocytes (Fig. 4b and Supplementary Fig. 3b). Moreover, within this loop, TEA interacted more frequently with the region encompassing Trdv2-2 and a neighboring CBE (fragments ix–xii) (Fig. 5c,d). As expected, interaction between TEA and Trdl1 (fragment xv) was unaffected by deletion of INT1-2 (Fig. 5d).

To investigate whether the new loop organization reported above facilitated contacts between Trdv2-2 (fragment xii) and Dδ3 and Iδ segments, we used fragments D2J1 (containing Trdl2 and Trdj1) and Eδ (containing Trdj2 and Eδ) as viewports (Fig. 5e). Interactions of Trdv2-2 with these viewports were substantially more frequent on alleles with deletion of INT1-2 than on wild-type alleles (Fig. 5e). However, as expected, interactions of D2J1 with Trdv5 (fragment xvi) and of Eδ with Trdl1 (fragment xv) were similar on wild-type alleles and alleles with deletion of INT1-2 (Fig. 5e). Therefore, deletion of...
INT1-2 redefined the chromatin-interaction landscape in a manner that facilitated contacts among the Trav15-dv6 and Dδ and Jβ RSSs (Supplementary Fig. 5).

Partial redundancy between INT1 and INT2
Because the INT1-2 deletion spanned 5.8 kb, we were unable to evaluate the specific contributions of INT1 and INT2 to the observed dysregulation of rearrangement and chromatin looping on the mutated allele. To specifically assess INT2 and the INT2-TEA chromatin loop, we generated an allele in which INT2 was replaced with a scrambled allele. To specifically assess INT2 and the INT2-TEA chromatin loop, regulation of rearrangement and chromatin looping on the mutated DNA sequence (called the INT2M allele here; Fig. 6a–c). Chromatin-immunoprecipitation analysis confirmed that CTCF did not bind to the mutated INT2 site (Fig. 6d). In contrast to results obtained for INT1-2-deficient mice, we observed no difference between wild-type mice and INT2M mice in terms of the number of total thymocytes or the frequency of γδ T cells (Fig. 7a). However, INT2M mice had twice as many Vγδ+ γδ T cells as wild-type mice had (Fig. 7a). Consistent with that result, Travδ2 rearrangement on INT2M alleles was 50% greater than that on wild-type alleles, whereas rearrangement of Travδ5 and that of Trav15-dv6 were each 50% lower on INT2M alleles than on wild-type alleles (Fig. 7b). Rearrangement of several other Vδ gene segments was unchanged on INT2M alleles relative to that on wild-type alleles (Fig. 7b). Therefore, INT2M mice partially recapitulated the phenotypic defects observed in INT1-2-deficient mice. We also measured chromatin interactions on INT2M alleles using TEA, Eδ and Dδ2Jβ1 as viewpoints. Perhaps unexpectedly, the interaction between

Figure 3 Thymocyte development in INT1-2-deficient mice. (a) Flow cytometry of thymocytes from wild-type and INT1-2-deficient littermates, stained for CD4 and CD8 (total thymocytes) (far left), for CD44 and CD25 (DN thymocyte populations depleted of CD4+ and CD8+ cells and pre-gated as 7-AAD−CD4−CD8−CD11b−Ter1 19−B220−Gr1−CD3ε−) (middle left), for γδ TCR (total thymocytes) (middle right), or for Vδ4 (pre-gated γδ TCR+ thymocytes) (far right). Numbers in quadrants (left) indicate percent cells in each; numbers adjacent to outlined areas (right) indicate percent γδ TCR+ cells (middle right) or Vδ4+ cells (far right) among the populations noted above. (b) Quantification of total thymocytes (left) and frequency of γδ TCR+ thymocytes among total thymocytes (middle) or Vδ4+ thymocytes among pre-gated γδ TCR+ thymocytes (right) in wild-type and INT1-2-deficient mice. Each symbol represents an individual mouse; small horizontal lines indicate the mean. *P < 0.05 and **P < 0.01 (unpaired Student’s t-test (left) or Mann-Whitney U-test (middle and right)). Data are representative of (a) or pooled from (b) three independent experiments.

Figure 4 Restricted TCRβ repertoire in INT1-2-deficient mice. (a) Locus map identifying Vβ segments analyzed. (b,c) Quantitative PCR analysis of VBD-Trdv1 rearrangements (b) and VBD-Trdd2 rearrangements (c) in genomic DNA extracted from DN3 thymocytes from wild-type, INT1-2-deficient and Eδ-deficient (EδKO) mice, with results normalized to those of the control gene Cd14; PCR analysis of Trav15-dv6 detects Trav15-1-dv6-1 and Trav15d-1-dv6d-1, and PCR analysis of Trav16d-dv11 detects Trav16d-dv11 and Trav16. (d) Quantitative PCR analysis of rearrangements, assessed by measurement of the retention of chromosomal DNA in total thymocytes (with PCR amplicons upstream of gene segments identified); results were normalized to each other on the basis of the abundance of Eδ and are presented relative to retention of the TEA amplicon in wild-type cells, set as 100%. (e) Quantitative PCR analysis of VBD-Trdv1 rearrangements in genomic DNA extracted from wild-type and INT1-2-deficient thymi at embryonic day 15.5 (normalized as in b,c). *P < 0.05; **P < 0.01; ***P < 0.001 and ****P < 0.0001 (two-way analysis of variance (ANOVA) with Tukey’s (b,c) or Sidak’s (d,e) multiple-comparison test). Data are pooled from three independent experiments (b,c); mean and s.e.m. of n = 3 samples (wild-type or INT1-2-deficient) or 2–3 samples (Eδ-deficient), each pooled from two to three mice), two independent experiments (d), mean ± s.e.m. of n = 3 samples per genotype, each from one mouse) or two independent experiments (e); mean and s.e.m. of 3 samples (wild type) or 4 samples (INT1-2-deficient), each pooled from two to three mice).
Figure 5 Deletion of INT1-2 alters chromatin loop organization but not chromatin accessibility. (a) Germline transcription of various genes (horizontal axis) in wild-type and INT1-2-deficient DN thymocytes (all on a Rag2⁻/⁻ background); results were normalized to those of the control gene Hprt. (b) Abundance of acetylated histone H3 (H3ac) in wild-type and INT1-2-deficient DN thymocytes (all on a Rag2⁻/⁻ background); results (bound/input) were normalized to those obtained for B2m. (c) Map of fragments analyzed by 3C of long-distance interactions: gray ovals, CBEs, below, viewpoint (gray rectangles) and target (numbered black rectangles) HindIII fragments; V segments shaded gray are pseudogenes. (d) 3C analysis of interactions of HindIII fragments with the TEA viewpoint in wild-type and INT1-2-deficient DN thymocytes (all on a Rag2⁻/⁻ background); results were normalized to those of the TEA nearest-neighbor fragment. (e) 3C analysis as in (d), from D2J1 and Eδ viewpoints; results were normalized to the interaction between TEA and its neighbor fragment (D2J1 viewpoint) or to those of the Eδ nearest-neighbor fragment (Eδ viewpoint). *P < 0.05, **P < 0.001 and ***P < 0.0001 (two-way ANOVA with Sidak’s multiple-comparison test (a,b,d) or unpaired Student’s t-test with the Holm-Sidak correction for multiple comparisons (e)). Data are pooled from three independent experiments (a; mean and s.e.m. of n = 2–4 preparations (wild type) or 2–3 preparations (INT1-2-deficient), each pooled from two to three mice), three independent experiments (b; mean and s.e.m. of n = 3 preparations (wild type) or 3–4 preparations (INT1-2-deficient), each pooled from eight to ten mice), four independent experiments (d; mean and s.e.m. of n = 3–5 preparations (wild type) or 3–6 preparations (INT1-2-deficient), each pooled from eight to ten mice) or three independent experiments (e; mean and s.e.m. of n = 3 preparations per genotype (for D2J1), or n = 4 preparations (wild type) or 3–4 preparations (INT1-2-deficient) (for Eδ), each pooled from eight to ten mice).

TEA and INT2 was only slightly lower on INT2M alleles than on wild-type alleles, whereas the interaction between TEA and INT1 was twice as much on INT2M alleles as on wild-type alleles (Fig. 7c). More frequent interaction with INT1 might explain the relatively modest reduction in TEA-INT2 interaction, given the resolution of 3C. INT2M alleles also displayed moderately more frequent interactions between TEA and sites upstream of INT1, including INT3 and Trdv2-2, than did wild-type alleles; similarly, Trdv2-2 interacted more frequently with D2J1 and Eδ on INT2M alleles than on wild-type alleles (Fig. 7c). However, none of these increases on INT2M alleles were as substantial as those on alleles lacking INT1-2. These data suggested that when INT2 was eliminated, INT1 partially subsumed its function by looping to TEA. However, additional looping to upstream sites allowed communication between Trdv2-2 and D3 and Iγ segments, which led to increased Trdv2-2 rearrangement. Together these data indicated involvement of INT2 in the dysregulation on alleles with deletion of INT1-2 and revealed that INT1 was able to partially compensate for INT2 when the latter was inactivated.

Figure 6 Generation of INT2M mice. (a) The wild-type 129/SvJ Tcra-Tcrδ allele, INT2M targeting construct, INT2M allele containing a neo° cassette, and INT2M allele with deletion of the neo° cassette (presented as in Fig. 2b). (b) Southern blot analysis of genomic DNA from wild-type embryonic stem cells (WT) and INT2M embryonic stem cells with targeting of the neo° cassette (INT1-2M neo°); the 6.3-kb HindIII fragment indicates a correctly targeted allele. (c) Genotyping PCR of wild-type mice and their homozygous INT2M (INT1-2KO) or heterozygous INT2M (INT2M het) littermates. (d) Chromatin-immunoprecipitation analysis of the binding of CTCF to INT1, INT2 or TEA (horizontal axis) in wild-type Tcra-Tcrδ alleles and INT2M Tcra-Tcrδ alleles of Rag2⁻/⁻ thymocytes and INT2M Rag2⁻/⁻ thymocytes, respectively; Trdv4 serves as a negative control. *P < 0.01 (unpaired Student’s t-test with the Holm-Sidak correction for multiple comparisons). Data are representative of two experiments (b) or more than three experiments (c) or are pooled from two independent experiments (d; mean and s.e.m. of n = 3 samples (wild type) or 2 samples (INT2M), each pooled from two to three mice).
Altered TCRα repertoire in INT1-2-deficient mice

Large Vα and Jα arrays allow multiple rounds of Vα-Jα rearrangement. Numerous studies have supported a model of sequential Jα use in DP thymocytes, with primary rearrangements targeted to the most 5′ (Trac-distal) Jα segments made accessible by activity of TEA, and subsequent rearrangements targeted to progressively more 3′ Jα segments made accessible by Vα promoters introduced in prior rounds of recombination.24,34 Accordingly, Vα use must progress from a Jα-proximal location to a Jα-distal location on individual alleles. Many studies have shown that Jα-proximal Vα segments (Trav19 and Trav21-dv12) rearrange almost exclusively to 5′ Jα segments35–37. This use is consistent with 3C data indicating that these Vα segments rarely rearrange to 5′ Jα segments, members of centrally positioned Vα families often do35–37. We envisaged that Vα-Jα combinatorial diversity might be facilitated by heterogeneous Vβ rearrangement in DN thymocytes that variably truncates the Vα array, placing a range of more distal Vα segments in a Jα-proximal position before the onset of Vα-Jα recombination. With this hypothesis, we predicted that if Vβ use were limited to the most proximal Vβ segments, as on alleles lacking INT1-2, combinatorial Vα-Jα diversity would be reduced. To understand the effect of deletion of INT1-2 on the TCRα repertoire, we used quantitative PCR to analyze Vα-Jα rearrangement in genomic DNA isolated from DP thymocytes of wild-type and INT1-2-deficient mice (Fig. 8). As expected, we found that in wild-type DP thymocytes, the most proximal Vα segments (Trav21-dv12, Trav19 and Trav17) rearranged almost exclusively to the most 5′ Jα segments (Traj61, Traj58 and Traj56) (Fig. 8a,b). In contrast, central Vα families (Trav12, Trav13 and Trav14) rearranged to broadly distributed Jα segments (Fig. 8c). However, rearrangement of proximal Vα segments to 5′ Jα segments was markedly greater in INT1-2-deficient DP thymocytes than in wild-type cells, whereas rearrangement of

Figure 7 Partial redundancy between INT1 and INT2. (a) Quantification of total thymocytes (left), and frequency of γδ TCR+ thymocytes among total thymocytes (middle) and of Vγ4+ thymocytes among pre-gated γδ TCR+ thymocytes (right) in wild-type and INT2M mice. Each symbol represents an individual mouse; small horizontal lines indicate the mean. (b) Quantitative PCR analysis of VDD-Trd1 rearrangement in genomic DNA extracted from DN3 thymocytes from 3- to 4-week-old wild-type and INT2M mice (normalized as in Fig. 4b,c). (c) 3C analysis of long-distance interactions in wild-type, INT1-2-deficient and INT2M DN thymocytes (all on a Rag2−/− background), from the TEA viewpoint (data for wild-type and INT1-2-deficient sites vi, x and xiv are identical to those in Fig. 5d). Data are pooled from three independent experiments (a,b; mean and s.e.m. of n = 3 preparations per genotype, each a pool of two to three mice, in b) or four independent experiments (c; mean and s.e.m. of n = 3–5 preparations per genotype, each from one mouse).

Figure 8 Restricted TCRα repertoire in INT1-2-deficient mice. (a) Partial locus map, showing Tcra segments (lines extending upward) and Tcρ segments (lines extending downward); lines positioned identically (‘co-linear’) indicate segments with both Tcra and Tcρ designations: black, D and J segments; colors (key), selected Tcra families; gray, Tcρ segments. (b,c) Quantitative PCR analysis of the rearrangement of Jα-proximal Vα segments (b) and Jα-distal Vα segments (c) to various Jα segments (horizontal axes) in genomic DNA extracted from DP thymocytes from wild-type and INT1-2-deficient mice (normalized as in Fig. 4b,c). ND, not detected. *P < 0.05, **P < 0.01, ***P < 0.001 and ****P < 0.0001 (two-way ANOVA with Sidak’s multiple-comparison test). Data are pooled from three independent experiments (b,c; mean and s.e.m. of n = 3 preparations per genotype, each from one mouse).
central Vα segments to 5′ Jα gene segments was strongly suppressed in INT1-2-deficient DP thymocytes relative to that in wild-type cells (Fig. 8b,c). These data indicated that the rearrangement of broadly distributed Vα segments in DN thymocytes provided an important mechanism for diversification of the TCRα repertoire.

**DISCUSSION**

Our work has defined a CTCF-dependent chromatin-interaction network that organizes the 3′ portion of the Tera-Tcrd locus in DN thymocytes. We identified two CBEs, INT1 and INT2, as key participants in this interactome with critical roles in diversifying the TCRδ and TCRα repertoires. Eliminating INT1 and INT2 from the Tera-Tcrd locus redefined the chromatin-interaction network, generating a new loop organization that facilitated rearrangement of Trdv2-2 while discouraging rearrangement of other Vδ segments. Abnormally homogenous Vδ use subsequently restricted Vα use during primary Vα-Jα rearrangement in DP thymocytes. Thus, our work has demonstrated an important and previously unappreciated link between the repertoire diversification of TCRδ and that of TCRα.

As defined by three-dimensional fluorescence in situ hybridization, the Tera-Tcrd locus adopts a highly contracted configuration in DN thymocytes. Our 4C analysis identified within this compact structure a high-frequency (and thus relatively stable) chromatin loop between TEA and INT2, as well as multiple low-frequency (and presumably more dynamic) chromatin loops between INT1 and other sites in the 3′ portion of the locus. TEA, INT1 and INT2 are all located in transcriptionally silent regions of the locus in DN thymocytes. Although these CBEs interacted with transcriptionally active regions (e.g., Dδ and Jδ segments), we presume that formation of the INT2-TEA loop is transcription independent. In that sense, looping in this portion of the locus is different in DN thymocytes than it is in DP thymocytes, in which Eδ interacts with and transcriptionally activates its target promoters. Thus, we view the INT2-TEA loop to be mainly structural in nature, setting the stage for Tcrd recombination in DN thymocytes. Notably, the chromatin-loop landscape of DN thymocytes was absent in the decontracted Tera-Tcrd locus in B cells, even though the relevant CBEs were occupied by CTCF in these cells. What instigates CBE-mediated looping is unknown.

Tera-Tcrd alleles lacking INT1-2 acted like IγR1-deficient Igh alleles, in the sense that both display dominant contributions of immediately upstream V segments to the respective repertoires. However, from a mechanistic perspective, the deletion of INT1-2 was distinct, since dysregulated use of upstream VH segments on IγR1-deficient alleles is associated with increased VH transcription and accessibility. On the basis of those findings, IγR1 functions, at least in part, as a transcriptional insulator that protects proximal VH segments from Eγ. The distinct roles of INT1 and INT2 might reflect the distinct properties of Eδ and Eγ. Unlike Eγ, which is capable of long-distance interactions and distal VH activation, Eδ might be unable to activate transcription over long distances. Therefore, rather than functioning as a transcriptional insulator, CTCF-mediated loops seem to regulate the TCRδ repertoire by serving as a ‘rehostat’ that determines the frequency with which Vδ and Dδ RSSs come into contact and can undergo synapsis. Apparently, the highly accessible Trdv2-2 segment must be physically segregated from Dδ and Jδ segments by the INT2-TEA loop on wild-type alleles, whereas Trdv5, an intrinsically less accessible Vδ segment, does not require such segregation. In this view, recombination frequency depends on several factors, including accessibility and contact frequency. With Trdv2-2 incorporated into the same loop as Trdv5, the combination of high accessibility and increased contact frequency must provide Trdv2-2 a recombination advantage over Trdv5 (and other Vδ segments), even though the weakly accessible Trdv5 contacts Dδ and Jδ segments more frequently. A contact mechanism has also been invoked to explain the effects of insertion of an ectopic CBE into the Tcrb locus. Whether endogenous Tcrb CBEs function similarly is not known.

Notably, although they are only 4.7 kb apart, INT1 and INT2 have very different interactomes. Our data would suggest that INT2 normally outcompetes INT1 for the convergently oriented CBE associated with TEA, with looping between INT1 and TEA facilitated only by mutation of INT2. INT1, normally excluded from looping to TEA, displayed a diverse array of low-frequency interactions with similarly oriented CBEs and other elements. This looping is presumably heterogeneous at the single-cell level, which would suggest that INT1 samples heterogeneous Vδ segments and brings them into proximity of the INT2-TEA loop to promote repertoire diversity. Although we did not selectively mutate INT1, the intermediate phenotype of INT2M mice would suggest that INT1 and INT2 are both required for a normal TCRδ repertoire. To the extent that INT1 can assume the role of INT2 on INT2M alleles, the dynamic tethering function of INT1 might be compromised. Nevertheless, INT1 was not able to fully assume the stable looping function of INT2, because INT2M alleles displayed more looping between TEA and upstream sites (e.g., INT3, which was also convergent with TEA) than did wild-type alleles. Repurposing of one CBE due to loss of another has been demonstrated at the Tcra locus.

The defect in γδ T cell production in INT1-2-deficient mice probably did not reflect a reduction in complete Vδ-Dδ-Jδ rearrangements. Instead, the lower number of γδ T cells might have been secondary to the restricted TCRδ repertoire in INT1-2-deficient thymocytes. Cells bearing Vδ4 (Trdv2-2), Vδ5 (Trdv5) and Vδ6 (Trav15-δvδ6) are selected in the thymus. Defective γδ T cell production might therefore reflect constraints on the selection of Vδ4+ γδ cells. In fact, although INT1-2-deficient DN3 thymocytes displayed considerably fewer Trav15-δvδ6 rearrangements, Vδ6+3 cells still represented 15% of total γδ T cells in these mice. Thus, the contribution of Vδ4+ γδ cells to the γδ T cell repertoire of INT1-2-deficient mice (55%) might represent an underestimation of the extent to which Trdv2-2 rearrangements predominate in DN3 thymocytes.

Our data indicated that Tcra combinatorial diversity was enhanced by INT1 and INT2. Such regulation would probably not be direct, since INT1 and INT2 would normally be deleted from over 70% of alleles by Tcrd rearrangements in DN thymocytes. A direct influence on Tcra repertoire diversity emerging from the remaining 30% of alleles could still be envisaged, perhaps reflecting the tethering function of INT1. However, all functional members of the Trav12 family are distal to the INT1 interactome, even though among the central Vα families tested, Trav12 was most dependent on INT1 and INT2 for primary rearrangement to 5′ Jα segments.

With the considerations noted above in mind, we propose that the altered TCRα repertoire in INT1-2-deficient mice was an indirect result of perturbed Tcra rearrangement. Our data are consistent with published studies indicating that in wild-type mice, both Jα-proximal and Jα-distal Vα segments may participate in primary rearrangements (to the most 5′ Jα gene segments). However, we conclude that the use of Jα-distal Vα segments depends heavily on heterogeneous Tcra rearrangements involving Jα-disulf Vδ segments, which would variably truncate the Vα-Vδ array before Tcra rearrangement begins. Given the finding that Tcra rearrangements were strongly biased to Trdv2-2 in INT1-2-deficient DN thymocytes, the Vα-Vδ array would remain largely intact, preserving proximal Vδ segments for primary Tcra rearrangement. Additionally, looping of TEA to INT3 on
these alleles could hold proximal $V_a$ segments near $5' J_a$ segments, facilitating the assembly of an $E_{R}^{-}$-dependent network of interactions involving TEA and proximal $V_a$ promoters. In these ways, homogenous and proximally biased $Tcrd$ rearrangements would favor proximally biased primary $Tcra$ rearrangements. Collectively, our data indicate that during primary $Tcra$ recombination, $5' J_a$ segments rearrange to the most proximal of the available $V_a$ segments. Whether this bias is strictly maintained through subsequent rounds of recombination is uncertain. Nevertheless, our data emphasize that $Tcrd$ rearrangement is an important diversifier of the TCRα repertoire, which suggests a rationale for the nested organization of $Tcrd$ and $Tcra$ gene segments in a single locus.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. GEO: 4C data, GSE67442.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank C. Bock for the production of mice with gene targeting; Y.-W. He (Duke University) for the targeting vector pGKneoF2L2DTA; O. Fedrigo for sequencing of 4C samples; D. Corcoran for advice on the processing of 4C sequencing data, L. Martinek, N. Martin and M. Cook for help with cell sorting and analysis; J. Liang for technical advice; Z. Huang for technical support; and E. Oltz and Y. Zhuang for comments on the manuscript. Supported by the US National Institutes of Health (R37 GM41052 to M.S.K.).

AUTHOR CONTRIBUTIONS

L.C. and M.S.K. planned the study; L.C., Z.C. and H.-Y. S. designed and performed the experiments; and L.C., Z.C. and M.S.K. analyzed the experiments and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Schatz, D.G. & Swanson, P.C. V(D)J recombination: mechanisms of initiation. Annu. Rev. Genet. 45, 167–202 (2011).
2. Schatz, D.G. & Ji, Y. Recombination centres and the orchestration of V(D)J recombination. Nat. Rev. Immunol. 11, 251–263 (2011).
3. Roldán, E. et al. Locus ‘decontraction’ and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. Nat. Immunol. 6, 31–41 (2005).
4. Skok, J.A. et al. Reversible contraction by looping of the Tcra and Tcrb loci in rearranging thymocytes. Nat. Immunol. 8, 378–387 (2007).
5. Jhungjunwala, S., van Zeilm, M.D., Peak, M.M. & Murre, C. Chromatin architecture and the generation of antigen receptor diversity. Cell 138, 435–448 (2009).
6. Shih, H.Y. & Krangel, M.S. Distinct contracted conformations of the Tcra/Tcrd locus during $Tcra$ and $Tcrd$ recombination. J. Exp. Med. 207, 1835–1841 (2010).
7. Degner, S.C. et al. CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the IgH locus and antisense transcription in pro-B cells. Proc. Natl. Acad. Sci. USA 108, 9566–9571 (2011).
8. Gao, C. et al. CTCF-binding elements mediate control of V(D)J recombination. Nature 477, 424–430 (2011).
9. Gao, C. et al. Two forms of loops generate the chromatin configuration of the immunoglobulin heavy-chain gene locus. Cell 147, 332–343 (2011).
10. Ribeiro de Almeida, C. et al. The DNA-binding protein CTCF limits proximal $V_a$ recombination and restricts $X$ enhancer interactions to the immunoglobulin $\alpha$ light chain locus. Immunity 35, 501–513 (2011).
11. Shih, H.Y. et al. $Tcra$ gene recombination is supported by a $Tcra$ enhancer- and CTCF-dependent chromatin hub. Proc. Natl. Acad. Sci. USA 109, E3493–E3502 (2012).
12. Verma-Gaur, J. et al. Noncoding transcription within the IgH distal $V_a$ region at PAIR elements affects the 3D structure of the IgH locus in pro-B cells. Proc. Natl. Acad. Sci. USA 109, 17004–17009 (2012).
13. Medvedovic, J. et al. Flexible long-range loops in the IgH gene region of the IgH locus facilitate the generation of a diverse antibody repertoire. Immunity 39, 229–244 (2013).
14. Majumder, K. et al. Lineage-specific compaction of $Tcb$ requires a chromatin barrier to protect the function of a long-range tethering element. J. Exp. Med. 212, 107–120 (2015).
15. Ding, C.T. & Corces, V.G. CTCF: an architectural protein bridging genome topology and the generation of antigen receptor diversity. Nat. Rev. Genet. 10, 401–414 (2009).
16. Hao, B. & Krangel, M.S. Long-distance regulation of fetal $V_a$ genes and a severely restricted repertoire. J. Immunol. 193, 3746–3754 (2014).
17. Ji, Y. et al. The 3D structure of the immunoglobulin heavy-chain locus: A 3D map of the human genome at kilobase resolution reveals principles of chromosome looping. Cell 159, 1665–1680 (2014).
18. de Wit, J., Wilbert, O.M., Kabizel, D. & Arden, B. $V_a$ repertoire during thymic ontogeny suggests three novel waves of $\gamma\delta$ TCR expression. J. Immunol. 164, 1000–1012 (2000).
19. Monroe, R.J. et al. Developmental regulation of $Tcrb$ locus accessibility and expression by the $Tcrb$ enhancer. Immunity 10, 503–513 (1999).
20. Krangel, M.S. et al. A 3D map of the human gamma/delta TCR locus shows that $\gamma\delta$ V11.2 does not require the $\gamma\delta$ J1.1. J. Immunol. 176, 2484–2491 (2011).
21. Varshavsky, A., Cao, J., Liang, L. & Dekker, J. The long-range interaction landscape of gene promoters. Nature 489, 109–113 (2012).
22. Jhungjunwala, S. et al. A map of the cis-regulatory sequences in the mouse genome. Nature 488, 116–120 (2012).
23. Phillips-Cremins, J.E. et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell 153, 1281–1295 (2013).
24. Symmons, O. et al. Functional and topological characteristics of mammalian regulatory domains. Genome Res. 24, 390–400 (2014).
25. Krangel, M.S., Carabana, J., Abarrategui, I., Schlimgen, R. & Hawwari, A. Forcing order within a complex locus: current perspectives on the control of VDJ recombination at the murine T-cell receptor $\alpha$ locus. Immunol. Rev. 200, 224–232 (2004).
26. Boss, N. & Lefranc, M.P. The mouse (Mus musculus) T cell receptor $\alpha$ (TRA) and $\delta$ (TRD) variable genes. Dev. Comp. Immunol. 27, 465–497 (2003).
27. Hao, B. & Krangel, M.S. Long-distance regulation of fetal $V_a$ gene segment TRDV4 by the Tcrd enhancer. J. Immunol. 187, 2484–2491 (2011).
28. Goedert, P.J. & Green, P.D. V(D)J recombination and restricted diversity at the TCR $\delta$ locus. Curr. Opin. Immunol. 23, 2598–2603 (2003).
29. Hao, B. et al. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep. 10, 1297–1309 (2015).
30. Hao, B. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromosome looping. Cell 159, 1665–1680 (2014).
31. Tao, X. et al. Developmental regulation of TCR$\alpha$ locus accessibility and expression by the TCR$\alpha$ enhancer. Immunity 10, 503–513 (1999).
32. Abarrategui, I. & Krangel, M.S. Noncoding transcription controls downstream promoters to regulate T-cell receptor $\alpha$ recombination. EMBO J. 26, 4380–4390 (2007).
33. Ji, Y. et al. Promoters, enhancers, and transcription target RAG1 binding during V(D)J recombination. J. Exp. Med. 207, 2809–2816 (2010).
34. Hawwari, A. et al. $V\gamma$ locus enhancer identity and position in secondary T cell receptor $\alpha$ recombination. Proc. Natl. Acad. Sci. USA 104, 903–907 (2007).
35. Peterson, N. et al. Quantitative and qualitative changes in V-J-$\gamma\delta$ V-J-$\gamma\delta$ rearrangements. Nat. Immunol. 13, 501–513 (2012).
36. Giroigorti, L. et al. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. Cell 157, 950–963 (2014).
37. Ribeiro de Almeida, C. et al. Intrathymic $\delta$ selection events in $\gamma\delta$ cell development. Immunity 7, 83–95 (1997).
ONLINE METHODS

Generation and maintenance of INT1-2-deficient and INT2M mice. Homology arms were generated by PCR with Phusion High Fidelity DNA Polymerase (Thermo Scientific) and were sequenced to confirm PCR fidelity. For the generation of INT1-2-deficient mice, the long homology arm extended from nucleotide 1,497,612 to nucleotide 1,503,426 and the short homology arm extended from nucleotide 1,509,115 to nucleotide 1,510,716 of the Tcrδ-Tcrα locus (NCBI RefSeq database accession code NT_039614.1). For the generation of INT2M mice, the long homology arm extended from the nucleotide 1,503,427 to nucleotide 1,509,114, with nucleotides 1,509,043−1,509,062 (5′-GAACACTAGGGGGCAATGC-3′) replaced with a scrambled sequence (5′-CGACGAGAAGACTGAGCTTG-3′). The short-arm extended from the nucleotide 1,509,115 to nucleotide 1,510,716. Homology arms were cloned into the targeting vector pGKneoF2L2DTA, which contains a phosphoglycerate kinase promoter–driven neomycin-resistance (neo) cassette and diptheria toxin A (DTA) selection marker (a gift from Y.-W. He). EcoRV-linearized targeting constructs were used for electroporation of the TCI 129S6/SvEvTAC embryonic stem cell line. Neomycin-resistant embryonic stem cell clones were first screened by PCR and then their genotype was confirmed by Southern blot. Those that had been verified in that way were microinjected into C57BL/6J blastocysts, which were then implanted into pseudo-pregnant C57BL/6J females. Targeting constructs were used for electroporation of the TC1 129S6/SvEvTAC embryonic stem cell line. Neomycin-resistant embryonic stem cell clones were first screened by PCR and then their genotype was confirmed by Southern blot. Those that had been verified in that way were microinjected into C57BL/6J blastocysts, which were then implanted into pseudo-pregnant C57BL/6J female mice. Chimeric male founder mice were crossed with CMV-Cre® female mice (Jackson Laboratory) for deletion of the loxp-flanked neo cassette and to obtain germ line transmission. Mice with the appropriate gene targeting were bred for elimination of the CMV-Cre transgene and were of mixed C57BL/6J and 129 genetic background. Breeding schemes of Rag-deficient mice ensured that littermate controls always segregated wild-type strain 129 Tcrδ-Tcrα alleles. Experiments analyzing mutant alleles on a Rag2−/− background used Rag2−/− mice on a 129 genetic background as controls. Mice were killed at 3–4 weeks of age for collection of adult thymocytes. Fetal thymocytes were obtained from female mice with timed pregnancy, with the day of detection of a vaginal plug designated embryonic day 0.5. All animals were used in accordance with protocols approved by the Duke University Institutional Animal Care and Use Committee.

Flow cytometry and cell sorting. All reagents were from Biolegend, unless stated otherwise. Antibody to (anti-)TCRβ (GL3), -β2m (GK1.5), and -Vγ6.3/2 (8F4H7B7; BD Pharmingen), anti-CD4 (GK1.5) and anti-CD8 (53-6.7), followed by negative selection with sheep antibody to rat immunoglobulin G (IgG) Dynabeads (Life Technologies). DN cell samples that had been depleted of B cells, splenocytes were stained with 7-AAD, APC–anti-B220 (RA3-6B2), and PE–anti-CD3 (145-2C11), anti-Ter-119 (for erythroid cells) (TER-119), anti-CD11b (M1/70), and FITC–anti-CD4 (53-6.7/2 (8F4H7B7; BD Pharmingen), anti-CD4 (GK1.5) and anti-CD8 (53-6.7), followed by negative selection with sheep antibody to rat immunoglobulin G (IgG) Dynabeads (Life Technologies). DN cell samples that had been depleted of B cells, splenocytes were stained with 7-AAD, APC–anti-CD4 (RA3-6B2), and PE–anti-CD8 (53-6.7) and PE–CD11b (M1/70), with sorting for 7-AAD−B220+CD3−CD4−CD8− cells.

3C. 3C assays were performed essentially as described42, starting with 8 × 10^6 to 10 × 10^6 thymocytes that had undergone cross-linkage for 10 min at 25°C in 8 ml RPMI medium containing 10% FBS and 2% paraformaldehyde. The restriction enzyme HindIII (NEB) was used for digestion of chromatin. 3C products were quantified by Taqman–based quantitative real-time PCR as described44 (probe and PCR primer sequences, Supplementary Table 1). For the generation of control PCR templates, bacterial artificial chromosomes bMQ-440L6 and bMQ-441F17 (Source BioScience) were mixed in equimolar amounts, then were digested and religated. bMQ-440L6 spans proximal V gene segments from Trav19 to downstream of Trav2-2, whereas bMQ-441F17 spans from INT1-2 to the central Igα segments. This control template mixture was used for the generation of standard curves for all 3C-quantitative PCR assays.

4C and sequencing. Thymocytes were pooled from litters of C57BL/6 background Rag2−/− mice and splenic B cells were obtained from C57BL/6 mice. 3C libraries were generated from 1 × 10^6 cells as described34, by digestion with HindIII. Following generation of the 3C library, secondary digestion and re-ligation were performed as described45, with modifications. 3C libraries were digested overnight at 37°C with 200 U of DpnII, and reactions were supplemented with an additional 200 U of DpnII, followed by incubation for 6 h at 37°C. The digested libraries were purified by phenol–chloroform extraction, then underwent precipitation with 2.5 vol ethanol, and rehydration in 4 ml 30 mM Tris·HCl, pH 8.0, 10 mM MgCl2, 1 mM DTT and 0.1 mM ATP. T4 DNA ligase (200 U; NEB) was added, followed by incubation overnight at 16°C. The reaction was then supplemented with an additional 200 U T4 DNA ligase, followed by incubation for a minimum of 6 h at 16°C. 4C libraries were then purified by phenol–chloroform extraction, then underwent precipitation with 2.5 vol ethanol and rehydration in 200 μl 10 mM Tris·HCl, pH 8.0, 0.1 mM EDTA. Inverse PCR was performed for the TEA, INT1, INT2 and Eγ viewpoints for the generation of libraries for high-throughput sequencing. All PCR used Phusion polymerase in 1× Phusion HiFi buffer (NEB). For TEA, INT2 and Eγ, two separate PCRs were used for the generation of libraries. First-round products were generated with primers TEA-F (5′-TGCTGCAATCTCTTCTGGGTGTC-3′) and TEA-R (5′-CAATAACGTAACACCGCAAGC-3′), INT2-F (5′-TCCCTTATCTACAAGAAGTCTCG-3′) and INT2-R (5′-TGGTCCAGATCACAAAGAGCTATCT-3′), and Eγ-F (5′-GGAAATTACAGTGTCTGTCAAGC-3′) and Eγ-R (5′-CCACAATCTCTTCTGGGATGAC-3′). PCR conditions for TEA and Eγ were as follows: 30 s at 98°C, followed by 20 cycles of 10 s at 98°C, 30 s at 60°C and 2 min at 72°C, with a final extension for 10 min at 72°C. PCR conditions for INT2 were identical to that except that the annealing was at 55°C. Products from the first PCR were purified with QiaQuick PCR purification reagents (Qiagen) and UPrep spin columns (Genesee), and were subjected to a second round of PCR with versions of the forward (F) and reverse (R) primers noted above, with the addition of Illumina T5 adaptors to their 5′ ends, as follows:Adapter 1–TEA-F and Adapter 2–TEA-R; Adapter 2–INT2-F and Adapter 1–INT2-R; and Adapter 2–Eγ-F and Adapter 1–Eγ-R, where Adapter 1 is 5′-AA TTAGATCGGGCCAAAAAAGCCTATCTTCCCTTCTACAGACGACGCTCTCCG ATCT-3′ and Adapter 2 is 5′-CAAGACAGAGAGGGCGCAGTA-3′. PCR conditions for TEA and Eγ were as follows: 30 s at 98°C, followed by 10 cycles of 10 s at 98°C, 30 s at 65°C and 2 min at 72°C, with a final extension for 10 min at 72°C. PCR conditions for INT2 were identical to that, except that the annealing was at 58°C. For the INT1 viewpoint, one PCR of 30 cycles was performed with primers INT1-F (Adapter 2–5′-AGAAGGGGAAAGAAGCTGTG-3′) and INT1-R (Adapter 1–5′-ACTGACAGGACAGAAGGAC-3′), with annealing at 58°C. For both rounds of PCR, 5 μl of individual PCRs were run and results were pooled for each viewpoint. After the second round of PCR, products were purified as described above and amplification of libraries was verified by gel electrophoresis.

PCR products for each viewpoint from a given 4C library were quantified with Picogreen (Life Technologies), were ‘multiplexed’ by pooling in equimolar ratios, and were supplemented by the addition of a 15% or 30% ‘spike’ of PhiX control library (Illumina). Before being sequenced, the quality of the pooled libraries was assessed with the Bioanalyzer platform (Agilent). ‘Multiplexed’ libraries were then subjected to 50-bp single-end sequencing with the Illumina HiSeq 2000 platform.

Sequencing data were analyzed with a workflow modified from one that has been published43. FASTQ files containing raw ‘multiplexed’ data were split with viewpoint-specific primer sequences TEA-F, INT1-R, INT2-R and Eγ-R. The first 16 bp representing viewpoint sequence was excised, and the remaining 34 bp of each read were aligned to the mouse genome assembly version mm9 with Bowtie software, with allowance for 0 mismatches and sequences repeated up to ten times for alignment (Bowtie parameters -v 0 -m 10 -all -best -strata). A map of genomic HindIII digestion fragments was generated, and reads per HindIII fragment were quantified with Python scripts as described44 and were visualized using the UCSC Genome Browser. Data are presented as reads per million mapped sequence reads.

Tcrδ and Tcrα recombination and transcription. Genomic DNA was isolated from sorted DN3 thymocytes or DP thymocytes for analysis of Tcrδ recombination or Tcrα recombination, respectively. Rearrangements of Tcrδ and
CBE-orientation analysis. FASTA sequences corresponding to called peaks from ChIP-seq analysis of CTCF in DN thymocytes were obtained with the UCSC Genome Browser. These sequences were input into the MEME-Chip web-based motif analysis software suite (http://meme-suite.org/tools/meme-chip) with default parameters to scan both strands for one or zero occurrences of a particular 6- to 30-bp motif per input sequence. The motif with the top score matched the previously defined CTCF-binding motif from nucleotides 5–20 (ref. 45). Individual sequences were then manually curated for elimination of those that corresponded to a very minor CTCF peak, did not align to the center of a CTCF peak or were ambiguous with respect to orientation.

Statistical methods. Data were analyzed by two-way ANOVA or by the unpaired two-tailed Student’s t-test with corrections for multiple comparisons, as appropriate, with Graphpad Prism 6 software. P values of less than 0.05 were considered statistically significant. Sample sizes were estimated on the basis of initial experiments and measurements, rather than being predetermined on the basis of expected effect sizes. No data were excluded from analysis. There was no randomization of mice or ‘blinding’ of researchers to experimental groups.

42. Hagège, H. et al. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat. Protoc. 2, 1722–1733 (2007).
43. Stadhouders, R. et al. Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. Nat. Protoc. 8, 509–524 (2013).
44. Livak, F. & Schatz, D.G. T-cell receptor α locus V(D)J recombination by-products are abundant in thymocytes and mature T cells. Mol. Cell. Biol. 16, 609–618 (1996).
45. Nakahashi, H. et al. A genome-wide map of CTCF multivalency redefines the CTCF code. Cell Rep. 3, 1678–1689 (2013).