Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Nat Immunol. Author manuscript; available in PMC 2016 April 01.

Published in final edited form as:

Nat Immunol. 2015 October ; 16(10): 1085–1093. doi:10.1038/ni.3232.

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

Liang Chen, **Zachary Carico**, **Han-Yu Shih**, and **Michael S. Krangel**

Department of Immunology, Duke University Medical Center, Durham, NC 27710, USA

Abstract

The *Tcra-Tcrd* locus undergoes V(D)J recombination in $CD4^{-}CD8^{-}$ double-negative thymocytes and $CD4^{+}CD8^{+}$ double-positive thymocytes to generate diverse TCRδ and TCRα repertoires, respectively. Here we reveal a *Tcra-Tcrd* locus chromatin interaction network in double-negative thymocytes that was formed by interactions between CTCF-binding elements. Disruption of a discrete chromatin loop encompassing *Tcrd* diversity, joining and constant gene segments allowed a single variable gene segment to frequently contact and rearrange to diversity and joining gene segments and dominate the adult TCRδ repertoire. Disruption of this loop also narrowed the TCRα repertoire, which, we believe, follows as a consequence of the restricted TCRδ repertoire. Hence, a single CTCF-mediated chromatin loop directly regulates TCRδ diversity and indirectly regulates TCRα diversity.

Adaptive immunity depends on highly diverse repertoires of antigen receptors (AgRs) expressed by T and B lymphocytes. This diversity is generated by V(D)J recombination, in which variable (V), diversity (D) and joining (J) gene segments of T cell receptor (TCR) and immunoglobulin (Ig) genes are assembled during the early stages of T and B lymphocyte development, respectively. Initiation of this process requires the collaborative function of recombination activating gene 1 and 2 proteins (RAG1 and RAG2; hereafter, RAG)[1]. RAG is thought to bind to a D or J segment recombination signal sequence (RSS) within a recombination center and then capture a second RSS to form a synaptic complex[2]. Within this complex, RAG introduces precise double-strand breaks (DSBs) between gene segments and RSSs. Repair of DSBs by non-homologous end joining results in assembly of antigen receptor coding and signal joints[1].

AgR diversification must overcome daunting topological constraints to recruit gene segments for recombination that may be distributed across several megabases (Mb) of DNA.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence should be directed to M.S.K. (krang001@mc.duke.edu). Michael S. Krangel, Department of Immunology, PO Box 3010, Duke University Medical Center, Durham NC 27710; Tel: 919-684-4985; Fax: 919-684-8982.

AUTHOR CONTRIBUTIONS

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

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

ACCESSION CODES GEO: 4C data, GSE67442.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Multiple studies have shown that AgR loci undergo large-scale conformational changes during lymphocyte development, bringing distant gene segments into proximity. For example, 3D-fluorescence *in situ* hybridization has shown that the *Igh*, *Igk*, *Tcrb* and *Tcra-Tcrd* loci undergo contraction coinciding with the developmental stages during which V(D)J recombination occurs[3–7]. Conversely, loci can be extended to terminate V(D)J recombination, as has been documented for *Igh* and *Tcrb*[3,4]. 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 AgR repertoires.

Chromatin conformation capture (3C) and 3C-based assays have shown that AgR loci are demarcated by chromatin loops that juxtapose distant segments of DNA. Although studies have implicated roles for Pax5 and ying yang 1 (YY1) in *Igh* loop organization, the primary mediator of chromatin looping at *Igh*, *Igk*, *Tcra* and *Tcrb* is the CCCTC-factor binding factor (CTCF)[8–17]. 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 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 AgR loci. At the *Igh* locus, IGCR1, an intergenic CBE between the V_H and D_H arrays, insulates D_H-proximal V_H gene segments from the influence of the *Igh* enhancer (E_μ)[9]. With IGCR1 deleted, rearrangements are biased towards the hyperactive D_H-proximal V_H segments and become disordered and lineage-nonspecific. Intergenic CBEs at the *Igk* locus similarly insulate proximal V_κ gene segments from *Igk* enhancers[11,19]. At the *Tcra-Tcrd* locus, CTCF marks many important *cis*-regulatory elements and as a result helps to target the *Tcra* enhancer (E_α) to the J_α promoter, T-early-alpha (TEA), and to the promoters of J_α-proximal V_α gene segments. These interactions promote transcription, accessibility and recombination of these V_α and J_α gene segments[14]. Emerging genome-wide studies also indicate that CTCF-mediated looping may serve a structural or organizing role rather than a direct gene regulatory role[20–23].

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 and αβ T lymphocytes[24]. Approximately 100 V gene segments are distributed across 1.5 Mb, with *Tcrd* D, J, and constant (C) gene segments, and *Tcra* J and C gene segments clustered within the final 0.1 Mb of the locus (hereafter referred to as the 3′ end of the locus). The majority of V gene segments rearrange to J_α segments in $CD4^+CD8^+$ double-positive (DP) thymocytes and contribute to the TCRα repertoire. However, only a few V gene segments rearrange to D_δ and J_δ gene segments in $CD4^-CD8^-$ double-negative (DN) thymocytes and contribute to the TCRδ repertoire. Several V_δ gene segments (*Trdv1*, *Trdv2-2*, *Trdv4*, *Trdv5*) are positioned proximal to the $D_\delta J_\delta$ cluster and are thought to be used exclusively for *Tcrd* rearrangement. Others (*Trav21-dv12*, *Trav13-4-dv7*, *Trav6-7-dv9*, *Trav4-4-dv10*, *Trav14D-3-dv8*, *Trav16d-dv11* and the *Trav15-dv6* family) are more distal, are interspersed among V_α gene segments, and are used as both V_δ and V_α gene segments[25]. How the locus

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

produces a balanced and diverse TCRδ repertoire with representation of proximal and distal V_δ gene segments is unclear.

Here we defined a CTCF-dependent chromatin interaction network that extends across 0.5 Mb of the *Tcra-Tcrd* locus in DN thymocytes. We identified two intergenic CBEs, INT1 and INT2, that play central roles in this interaction network. INT1 interacts broadly and dynamically across this region of chromatin. However, INT2 specifically interacts with the CBE associated with the TEA promoter, forming a high frequency chromatin loop that segregates *Tcrd* D, J and C gene segments from most V_δ gene segments. Mice deleted for INT1 and INT2 on both alleles (hereafter referred to as INT1-2KO mice) had a highly restricted TCRδ repertoire, which was strongly biased towards *Trdv2-2*. This V_δ gene segment is normally segregated from D_δ gene segments by the INT2-TEA loop, but was newly included within the D_δ-containing loop with INT1-2 deleted. Biased V_δ usage resulted not from increased accessibility, but from increased interactions between *Trdv2-2* and D_δ gene segments. Of note, the TCRα repertoire was also altered in INT1-2KO mice, implicating heterogeneity of *Tcrd* rearrangement as a diversifier of *Tcra* rearrangement. Our results argue that a CTCF-dependent chromatin interaction network creates TCRδ and TCRα repertoire diversity 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 thymocytes[14]. A majority of these CBEs are associated with cis-regulatory elements, including V gene segment promoters, the TEA promoter and E_α. However, we noted two prominent intergenic CBEs, INT1 and INT2, in the 3′ portion of the locus. We asked whether these CBEs are weaved into a chromatin interaction network that sets the stage for *Tcrd* rearrangement in DN thymocytes. To map long-range interactions, we performed circular chromosome conformation capture sequencing (4C-seq), which assays genome-wide interactions with a single "viewpoint"[26]. We compared RAG2-deficient DN thymocytes to control splenic B cells; in both cell populations the *Tcra-Tcrd* locus remains in germline configuration. 4C libraries were prepared using *Hind*III for initial chromatin digestion and *Dpn*II for secondary digestion, with the results mapped to individual *Hind*III fragments. Data from viewpoints TEA, INT1, INT2 and E_δ are shown (Fig. 1a,b). In DN thymocytes, we found the TEA viewpoint to interact at high frequency with INT2; reciprocally, the INT2 viewpoint interacted frequently with TEA, forming a distinct 80 kb chromatin loop (Fig. 1a). This loop confined almost all additional contacts made by TEA and INT2, since both viewpoints interacted within the loop but rarely with regions outside the loop. Moreover, this loop segregated *Tcrd* D, J and C gene segments, as well as *Trdv4* and *Trdv5*, from other gene segments in the locus. Remarkably, although located only 4.7 kb upstream of INT2, INT1 participated in numerous low-frequency, long-range interactions extending across the 3′ 0.5 Mb of the *Tcra-Tcrd* locus (Fig. 1a); this suggests a dynamic loop organization. E_δ interacted almost exclusively with fragments within the TEA-INT2 loop (Fig. 1a), consistent with data showing that it only regulates transcription in the *Trdv4-Trdv5* interval[27]. The four interaction profiles were lineage-specific, since they were not

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

detected in B cells (Fig. 1b); nevertheless, CTCF binding to TEA, INT1 and INT2 was comparable in DN thymocytes and B cells (Fig. 1a,b). These CBEs appear to be key nodes in the *Tcra-Tcrd* locus interactome in DN thymocytes.

Genome-wide analyses recently revealed a strong preference for looping between convergently oriented CBEs[28,29]. Notably, the INT1 and INT2 CBEs share an orientation with the majority (87%) of *Tcra-Tcrd* locus CBEs, whereas the TEA CBE is in the reverse orientation and is convergent with INT2 (Supplementary Fig. 1). This may explain high frequency looping between these CBEs.

INT1-2KO mice have an altered TCRδ repertoire

To test a role for the INT1 and INT2 CBEs in *Tcrd* rearrangement, we generated an INT1-2-deleted allele in which the 5.8 kb DNA fragment containing INT1 and INT2 was eliminated in the mouse germline (Fig. 2a,b,c). Although the number of total thymocytes was mildly reduced in INT1-2KO mice, the development of $\alpha\beta TCR^+$ thymocytes was largely normal based on staining with antibodies specific to CD4 and CD8 (Fig. 3a,b). DN thymocytes can be subdivided into four successive developmental stages based on expression of CD44 and CD25: DN1 ($CD25^-CD44^+$), DN2 ($CD25^+CD44^+$), DN3 ($CD25^+CD44^-$), and DN4 ($CD25^-CD44^-$). Percentages of DN1-DN4 thymocytes were comparable between wild-type and INT1-2KO mice (Fig. 3a). However, the percentage of γδ T cells in INT1-2KO mice was about half that of control littermates (Fig. 3a,b). Moreover, the percentage of $V_\delta 4$ usage among γδ TCR^+ thymocytes increased by 3-fold in INT1-2KO mice (Fig. 3a,b); $V_\delta 6.3$ usage was, however, unchanged (Supplementary Fig. 2). Therefore, INT1-2KO mice display defective γδ T cell development and a biased TCRδ repertoire that is heavily skewed towards $V_\delta 4$.

$V_\delta 4$ is a commonly used adult V_δ[30] that is encoded by *Trdv2-2*, the first functional V_δ gene segment upstream of the INT1-2 region (Fig. 1a). We asked whether the skewed TCRδ repertoire in INT1-2KO mice could be attributed to dysregulated *Tcrd* rearrangements by quantifying VDDJ coding joints in genomic DNA samples prepared from DN3 thymocytes. The frequency of *Trdv2-2* rearrangement was markedly increased in INT1-2KO as compared to wild-type DN3 thymocytes (Fig. 4a,b). However, rearrangement of the related *Trdv2-1* was barely detectable in DN3 thymocytes of either genotype, a result confirmed by PCR using a shared *Trdv2* primer with a *Trdj1* primer; 11/11 clones of each genotype were *Trdv2-2* by sequencing. *Trdv1*, located 67 kb upstream of *Trdv2-2*, was equally rearranged in wild-type and INT1-2KO thymocytes (Fig. 4a,b). However, all other V_δ gene segments tested, including *Trav13-4-dv7*, *Trav6-7-dv9*, *Trav16d-dv11*, *Trdv5* and two *Trav15-dv6* family members, were substantially less frequently rearranged in INT1-2KO thymocytes (Fig. 4a,b). Therefore, *Trdv2-2* rearrangements predominated at the expense of other V_δ gene segments. This bias also extended to incomplete VD rearrangements. *Tcrd* rearrangement is unusual since it is unordered and VD, DD, and DJ rearrangements all occur. Of note, rearrangement of *Trdv2-2*-to-*Trdd1-Trdd2* increased, whereas *Trav15-dv6*- and *Trdv5*-to-*Trdd1-Trdd2* rearrangements decreased in INT1-2KO thymocytes (Fig. 4c). *Trdv2-2*-to-*Trdd1-Trdd2* rearrangements in INT1-2KO mice were as frequent as in E_δ-deficient (E_δKO) mice, in which partial rearrangements predominate[31].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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* gene segments in DN thymocytes places intervening signal joint DNAs onto extrachromosomal circles, which are diluted out and lost during pre-TCR-driven cell proliferation. In contrast, excised signal joints from *Tcra* rearrangement are retained in DP thymocytes because they are generated after pre-TCR-driven proliferation (see Methods for further discussion of this point). Thus, genomic DNA retention 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 indicated gene segments, 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 V_δ gene segments upstream of *Trdv1*, approximately 42% had rearranged *Trdv2-2*, and another 25% either had not undergone V-to-D rearrangement or had rearranged *Trdv5* by inversion (which would not delete the region upstream of *Trdd1*) (Fig. 4d). In contrast, INT1-2-deleted alleles displayed impaired *Trdd2-Trdj1* rearrangement, but increased *Trdv2-2*-to-*Trdd1-Trdd2* rearrangement (Fig. 4d). Precocious *Trdv2-2*-to-*Trdd1-Trdd2* rearrangements may inhibit *Trdd2*-to-*Trdj1* recombination events on INT1-2-deleted alleles. Because the amplicon upstream of *Trdv2-2* was retained on 93% of alleles whereas that upstream of *Trdd1* was retained on only 6% of alleles, *Trdv2-2* appears to undergo partial VDD or complete VDDJ rearrangement on most INT1-2-deleted alleles. INT1-2 deletion also caused increased rearrangement of *Trdv2-2* to the most 5′ J_α gene segments in DP thymocytes (Supplementary Fig. 3a). This rearrangement must occur on alleles that had not undergone *Trdv2-2*-to-D_δ rearrangement in DN thymocytes, and may explain slightly reduced retention of TEA in INT1-2KO as compared to wild-type thymocytes (Fig. 4d). However, we did not detect premature rearrangement of *Trdv2-2* or proximal V_α gene segments to J_α gene segments in INT1-2-deleted DN thymocytes (Supplementary Fig. 3b). Together, these data show that the INT1-2 genomic region is essential to generate a diverse V_δ repertoire in DN thymocytes.

V_δ usage in early fetal thymocytes is distinct from that of adult since it is strongly biased towards *Trdv4* ($V_\delta 1$), which is proximal to D_δ gene segments and within the INT2-TEA loop[30]. We asked whether *Tcrd* rearrangement in the fetal thymus was dysregulated in INT1-2KO mice. *Trdv4* rearrangement was unchanged in INT1-2KO E15.5 fetal thymocytes (Fig. 4e). However, we detected substantially increased rearrangement of *Trdv2-2* (Fig. 4e), indicating that the INT1-2 region limits rearrangement of adult V_δ segments in the fetal thymus. In contrast, INT1-2 deletion caused no dysregulation of *Trdv4* rearrangement in adult thymocytes (Supplementary Fig. 3b).

INT1-2 does not regulate chromatin accessibility

The regulation of RSS accessibility to the recombinase represents a critical control point for V(D)J recombination[2]. Germline transcription creates accessibility by disrupting nucleosome structure and organization and depositing histone modifications that facilitate

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

RAG binding and RAG enzymatic activity[2,32,33]. We asked whether abnormal *Tcrd* recombination on INT1-2-deleted alleles reflected altered germline transcription of *Tcrd* gene segments. Germline transcripts were quantified in *Rag2*$^{-/-}$ DN thymocytes carrying wild-type or INT1-2-deleted alleles maintained in unrearranged configuration. We found no differences in germline transcription besides a modest increase at *Trdj1* on INT1-2-deleted alleles (Fig. 5a). Thus, promoter activities were largely unaffected by INT1-2 deletion. We also analyzed histone H3 acetylation (H3ac) by chromatin immunoprecipitation (ChIP). Enrichment of histone H3ac was comparable at all sites examined on wild-type and mutant alleles (Fig. 5b). Hence, INT1-2-deletion did not substantially impact chromatin accessibility of *Tcrd* gene segments in DN thymocytes, and in particular, had no effect on V_δ gene segments. Increased *Trdj1* transcription may influence J_δ usage, because *Trdv2-2* rearrangement was elevated selectively at *Trdj1* on INT1-2-deleted alleles (Supplementary Fig. 3c).

INT1-2 regulates chromatin loop organization

We used 3C-quantitative PCR (qPCR) to ask whether INT1-2-deletion generates an altered landscape of long-distance chromatin interactions. 3C libraries were prepared by *Hind*III digestion and *Hind*III fragments were assayed for interactions with the TEA viewpoint (Fig. 5c,d). Note that although the INT1-2-deleted allele lacks the INT1 *Hind*III fragment, it retains the portion of the INT2 *Hind*III fragment that includes the primer-binding site. Consistent with the 4C-seq data, the TEA viewpoint strongly interacted with INT2 (fragment xiv) on wild-type alleles (Fig. 5d). However, TEA interacted minimally with the residual INT2 fragment on INT1-2-deleted alleles, instead interacting frequently with another intergenic CBE, INT3 (fragment vi), located 49 kb upstream of *Trdv2-2* (Fig. 5d). This interaction occurred despite weak CTCF binding at INT3 on both wild-type and INT1-2-deleted alleles (Supplementary Fig. 4). As a consequence of INT3-TEA interaction, *Trdv2-2* was confined within a new 250 kb loop that included *Tcrd* D, J, and C gene segments, E_δ, *Trdv5*, several V_δ pseudogenes, and two V_δ gene segments (*Trdv4* and *Trdv2-1*) that rearrange minimally in adult DN thymocytes (Supplementary Fig. 3b)[30]. Moreover, within this loop, TEA interacted more frequently with the region encompassing *Trdv2-2* and a neighboring CBE (fragments ix to xii) (Fig. 5d). As expected, interaction between TEA and *Trdd1* (fragment xv) was unaffected by INT1-2 deletion (Fig. 5d).

To test whether this new loop organization facilitates contacts between *Trdv2-2* (fragment xii) and D_δ and J_δ gene segments, we used fragments D2J1 (containing *Trdd2* and *Trdj1*) and E_δ (containing *Trdj2* and E_δ) as viewpoints (Fig. 5e). Interactions of *Trdv2-2* with these viewpoints were substantially more frequent on INT1-2-deleted alleles as compared to wild-type alleles (Fig. 5e). However, as expected, interactions of D2J1 with *Trdv5* (fragment xvi) and of E_δ with *Trdd1* (fragment xv) were comparable on wild-type and INT1-2-deleted alleles (Fig. 5e). Therefore, INT1-2-deletion redefines the chromatin interaction landscape in a manner that facilitates contacts between the *Trdv2-2* and D_δ and J_δ RSSs (Supplementary Fig. 5).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Partial redundancy between INT1 and INT2

Because the INT1-2 deletion spans 5.8 kb, we could not evaluate the specific contributions of the INT1 and INT2 CBEs to the observed dysregulation of rearrangement and chromatin looping on the mutant allele. To specifically test the INT2 CBE and the INT2-TEA chromatin loop, we generated an allele in which the INT2 CBE was replaced with a scrambled DNA sequence (hereafter referred to as the INT2M allele; Fig. 6a,b,c). CTCF chromatin immunoprecipitation (ChIP) confirmed that CTCF does not bind to the mutant INT2 site (Fig. 6d). In contrast to INT1-2KO mice, we observed no change in the number of total thymocytes or the percentage of γδ T cells (Fig. 7a). However, INT2M mice had twice as many $V_\delta4^+$ γδ T cells as wild-type mice (Fig. 7a). Consistent with this result, *Trdv2-2* rearrangement on INT2M alleles increased by 50% relative to wild-type, whereas rearrangement of *Trdv5* and *Trav15-dv6* were each reduced by 50% (Fig. 7b). Rearrangement of several other V_δ gene segments was unchanged (Fig. 7b). Therefore, INT2M mice partially recapitulate the phenotypic defects observed in INT1-2KO mice. We also measured chromatin interactions on the INT2M allele using TEA, E_δ and $D_\delta2J_\delta1$ as viewpoints. Perhaps surprisingly, interaction between TEA and INT2 only decreased modestly on the INT2M allele, whereas interaction between TEA and INT1 doubled (Fig. 7c). Elevated interaction with INT1 may explain the relatively modest reduction in TEA-INT2 interaction, given the resolution of 3C. INT2M alleles also displayed moderately increased interactions between TEA and sites upstream of INT1, including INT3 and *Trdv2-2*; similarly, *Trdv2-2* interacted more frequently with D2J1 and E_δ (Fig. 7c). However, none of these increases were as substantial as those on INT1-2-deleted alleles. These data suggest that with the INT2 CBE eliminated, the INT1 CBE partially subsumes its function by looping to TEA. However, additional looping to upstream sites allows communication between *Trdv2-2* and D_δ and J_δ segments, leading to increased *Trdv2-2* rearrangement. Together, these data implicate the INT2 CBE in the dysregulation on INT1-2-deleted alleles, and reveal that INT1 can partially compensate for INT2 when the latter is inactivated.

Altered TCRα repertoire in INT1-2KO mice

Large V_α and J_α arrays allow for multiple rounds of V_α-J_α rearrangement. Numerous studies support a model of sequential J_α usage in DP thymocytes, with primary rearrangements targeted to the most 5′ (*Trac*-distal) J_α segments made accessible by TEA promoter activity, and subsequent rearrangements targeted to progressively more 3′ J_α gene segments made accessible by V_α promoters introduced in prior rounds of recombination[24,34]. Accordingly, V_α usage must progress from J_α-proximal to J_α-distal on individual alleles. Numerous studies show that J_α-proximal V_α gene segments (*Trav19* and *Trav21-dv12*) rearrange almost exclusively to 5′ J_α gene segments[35-37]. This usage is consistent with 3C data indicating that these V_α and J_α gene segments are brought into contact by E_α on unrearranged alleles in DP thymocytes[14]. However, if primary rearrangement were always to initiate with the most proximal V_α gene segments, combinatorial diversity of the TCRα repertoire would be constrained. Although the most distal V_α gene segments rarely rearrange to 5′ J_α gene segments, members of centrally positioned V_α families often do[35-37]. We envisage that V_α-J_α combinatorial diversity can be facilitated by heterogeneous V_δ

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

rearrangement in DN thymocytes that variably truncates the V_α array, placing a range of more distal V_α segments in a J_α-proximal position prior to the onset of V_α-J_α recombination. This hypothesis predicts that if V_δ usage were limited to the most proximal V_δ gene segments, as on INT1-2-deleted alleles, combinatorial V_α-J_α diversity would be reduced. To understand the impact of INT1-2 deletion on the TCRα repertoire, we used qPCR to analyze V_α-J_α recombination in genomic DNA isolated from DP thymocytes of wild-type and INT1-2KO mice (Fig. 8). As expected, in wild-type DP thymocytes we found that the most proximal V_α gene segments (*Trav21-dv12*, *Trav19*, *Trav17*) rearranged almost exclusively to the most 5′ J_α gene segments (*Traj61*, *Traj58*, *Traj56*) (Fig. 8a,b). In contrast, central V_α families (*Trav12*, *Trav13*, *Trav14*) rearranged to broadly distributed J_α segments (Fig. 8c). Yet in INT1-2KO DP thymocytes, rearrangement of proximal V_α gene segments to 5′ J_α gene segments was markedly increased, whereas rearrangement of central V_α segments to 5′ J_α gene segments was strongly suppressed (Fig. 8b,c). We conclude that the rearrangement of broadly distributed V_δ gene segments in DN thymocytes provides an important mechanism to diversify the TCRα repertoire.

DISCUSSION

Our work has defined a CTCF-dependent chromatin interaction network that organizes the 3'portion of the *Tcra-Tcrd* locus in DN thymocytes. We identified two CBEs, INT1 and INT2, as key players in this interactome that play critical roles in diversifying the TCRδ and TCRα repertoires. Eliminating INT1 and INT2 from the *Tcra-Tcrd* locus redefined the chromatin interaction network, generating a new loop organization that facilitated rearrangement of *Trdv2-2*, while discouraging rearrangement of other V_δ gene segments. Abnormally homogeneous V_δ usage subsequently restricted V_α usage during primary V_α-J_α rearrangement in DP thymocytes. As such, our work has demonstrated an important and previously unappreciated link between TCRδ and TCRα repertoire diversification.

As defined by 3D-fluorescence *in situ* hybridization, the *Tcra-Tcrd* locus adopts a highly contracted configuration in DN thymocytes[7]. Within this compact structure, our 4C analysis identified a high-frequency (and thus relatively stable) chromatin loop between the TEA and INT2 CBEs, as well as multiple low frequency (and presumably more dynamic) chromatin loops between the INT1 CBE and other sites in the 3' portion of the locus. The TEA, INT1 and INT2 CBEs are all located in transcriptionally silent regions of the locus in DN thymocytes. Although these CBEs interact with transcriptionally active regions (eg., D_δ and J_δ gene segments), we presume that formation of the INT2-TEA loop is transcription-independent. In that sense, looping in this portion of the locus in DN thymocytes is different than in DP thymocytes, which involves E_α and its target promoters and is associated with transcriptional activation of those promoters[14]. Thus, we view the INT2-TEA loop to be primarily structural in nature, setting the stage for *Tcrd* recombination in DN thymocytes. Remarkably, the chromatin loop landscape of DN thymocytes is absent in the decontracted *Tcra-Tcrd* locus in B cells, even though the relevant CBEs are occupied by CTCF in these cells. What instigates CBE-mediated looping is unknown.

INT1-2-deleted *Tcra-Tcrd* alleles behave like IGCR1-deficient *Igh* alleles, in the sense that both display dominant contributions of immediately upstream V gene segments to the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

respective repertoires[9]. However, from a mechanistic perspective, the INT1-2 deletion is distinct, since dysregulated usage of upstream V_H gene segments on IGCR1-deficient alleles was associated with increased V_H transcription and accessibility[9]. Based on this, the IGCR1 CBEs function, at least in part, as a transcriptional insulator that protects proximal V_H gene segments from E_μ. The distinct roles of the INT1 and INT2 CBEs may reflect the distinct properties of E_δ and E_μ. Unlike E_μ, which is capable of long-distance interactions and distal V_H activation[10,15,16], E_δ may be unable to activate transcription over long distances[27,38]. Therefore, rather than functioning as a transcriptional insulator, CTCF-mediated loops appear to regulate the TCRδ repertoire by serving as a rheostat that determines the frequency with which V_δ and D_δ RSSs come into contact and can undergo synapsis. Apparently, the highly accessible *Trdv2-2* gene segment must be physically segregated from D_δ and J_δ segments by the INT2-TEA loop on wild-type alleles, whereas *Trdv5*, an intrinsically less accessible V_δ, 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 V_δ gene segments), even though the weakly accessible *Trdv5* contacts D_δ and J_δ segments more frequently. A contact mechanism was also invoked to explain the effects of an ectopic CBE insertion into *Tcrb* locus[39]. Whether endogenous *Tcrb* CBEs function similarly is not known.

Remarkably, although only 4.7 kb apart, the INT1 and INT2 CBEs have very different interactomes. Our data suggests that INT2 normally outcompetes INT1 for the convergently oriented TEA CBE, 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[16,40], suggesting that the INT1 CBE samples heterogeneous V_δ gene 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 suggests 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 may be compromised. Nevertheless, INT1 cannot fully assume the stable looping function of INT2, because INT2M alleles display elevated looping between TEA and upstream sites (eg., INT3, which is also convergent with the TEA CBE). Repurposing of one CBE due to loss of another was recently demonstrated at the *Tcrb* locus[17].

The defect in γδ T cell production in INT1-2KO mice seems unlikely to reflect a reduction in complete V_δ-D_δ-J_δ rearrangements. Rather, reduced γδ T cell numbers may be secondary to the restricted TCRδ repertoire in INT1-2KO thymocytes. Cells bearing $V_\delta4$ (*Trdv2-2*), $V_\delta5$ (*Trdv5*) and $V_\delta6$ (*Trav15-dv6*) are differentially selected in the thymus[41]. Defective γδ production may therefore reflect constraints on the selection of $V_\delta4^+$ γδ cells. In fact, although INT1-2KO DN3 thymocytes displayed dramatically reduced *Trav15-dv6* rearrangements, $V_\delta6.3^+$ cells still represented 15% of total γδ T cells in these mice. Thus, the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

contribution of $V_\delta4^+$ cells to the γδ T cell repertoire of INT1-2KO mice (55%) may underestimate the extent to which *Trdv2-2* rearrangements predominate in DN3 thymocytes.

Our data indicate that *Tcra* combinatorial diversity is enhanced by the INT1 and INT2 CBEs. This regulation is unlikely to be direct, since the INT1 and INT2 CBEs would normally be deleted from over 70% of alleles by *Tcrd* rearrangements in DN thymocytes[14]. 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 *Trav12* family members lie distal to the INT1 interactome, even though among the distal V_α families tested, *Trav12* was most dependent on INT1 and INT2 for primary rearrangement to 5' J_α segments.

With these considerations in mind, we believe that the altered TCRα repertoire in INT1-2KO mice is an indirect result of perturbed *Tcrd* rearrangement. Our data are consistent with prior studies indicating that in wild-type mice, both J_α-proximal and J_α-distal V_α gene segments may participate in primary rearrangements (to the most 5' J_α gene segments)[35-37]. However, we conclude that use of J_α-distal V_α gene segments depends heavily on heterogeneous *Tcrd* rearrangements involving J_α-distal V_δ gene segments, which would variably truncate the V_α-V_δ array before *Tcra* rearrangement begins. With *Tcrd* rearrangements strongly biased to *Trdv2-2* in INT1-2KO DN thymocytes, the V_α-V_δ array would remain largely intact, preserving proximal V_α gene segments for primary *Tcra* rearrangement. Additionally, TEA-INT3 looping on these alleles could hold proximal V_α gene segments near 5' J_α gene segments, facilitating assembly of an E_α-dependent network of interactions involving TEA and proximal V_α promoters[14]. In these ways, homogeneous and proximally biased *Tcrd* rearrangements would favor proximally biased primary *Tcra* rearrangements. Collectively, our data argue that during primary *Tcra* recombination, 5' J_α gene segments rearrange to the most proximal of the available V_α gene segments. Whether this bias is strictly maintained through subsequent rounds of recombination is uncertain[37]. Nevertheless, our data emphasize that *Tcrd* rearrangement is an important diversifier of the TCRα repertoire, suggesting a rationale for the nested organization of *Tcrd* and *Tcra* gene segments in a single locus.

METHODS

Generation and maintenance of INT1-2KO and INT2M mice

Homology arms were generated by PCR using Phusion High Fidelity DNA Polymerase (Thermo Scientific) and were sequenced to confirm PCR fidelity. To generate INT1-2KO mice, the long homology arm extended from nucleotide 1,497,612 to 1,503,426 and the short homology arm extended from nucleotide 1,509,115 to 1,510,716 of *Tcra-Tcrd* locus NCBI Reference Sequence NT_039614.1. To generate INT2M mice, the long homology arm extended from the nucleotide 1,503,427 to 1,509,114, with nucleotides 1,509,043 to 1,509,062 (5′-GAACACTAGGGGGCAATGC-3′) replaced with a scrambled sequence (5′-CGACGAGAAGCTAGCAGTG-3′)[9]. The short-arm extended from the nucleotide 1,509,115 to 1,510,716. Homology arms were cloned into the pGKneoF2L2DTA targeting vector containing a phosphoglycerate kinase promoter-driven neomycin resistance (*neo*^{r}) cassette and diphtheria toxin A (DTA) selection marker (a gift from Y.-W. He, Duke

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

University). *Eco*RV-linearized targeting constructs were used to electroporate the TC1 129S6/SvEvTAc embryonic stem (ES) cell line. Neomycin resistant ES cell clones were first screened by PCR and then verified by Southern blot. Verified ES cells 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^{tg} female mice (Jackson Laboratory) to delete the *lox*P-flanked *neo*^{r} cassette and obtain germline transmission. Gene-targeted mice were bred to eliminate the CMV-Cre transgene and were of mixed C57BL/6 and 129 genetic background. Breeding schemes of *Rag*-sufficient mice ensured that littermate controls always segregated WT strain 129 *Tcra-Tcrd* alleles. Experiments analyzing mutant alleles on a *Rag2*$^{-/-}$ background used *Rag2*$^{-/-}$ mice on a 129 genetic background as controls. Mice were sacrifice at 3-4 weeks of age to harvest adult thymocytes. Fetal thymocytes were harvested from timed-pregnant female mice, with the day of detection of a vaginal plug designated E0.5. All mice were used in accordance with protocols approved by the Duke University Institutional Animal Care and Use Committee.

Flow cytometry and cell sorting

Anti-mouse TCRγ/δ (GL3), anti-mouse TCR $V_{\delta}4$ (GL2), anti-mouse $V_{\delta}6.3/2$ (8F4H7B7, BD Pharmingen), anti-CD4 (GK1.5) and anti-CD8 (53-6.7) were used to stain total thymocytes. To analyze or sort DN3 thymocytes, total thymocytes were stained with anti-CD4 (GK1.5) and anti-CD8 (53-6.7), followed by negative selection with sheep anti-rat IgG Dynabeads (Life Technologies). Bead-depleted DN cells were then stained with 7AAD, APC-anti-CD44 (IM7), FITC-anti-CD25 (PC61), and PE-Cy5-conjugated lineage markers, including anti-Gr-1 (RB6-8C5), anti-CD3ε (145-2C11), anti-Ter-119/Erythroid Cells (TER-119), anti-CD11b (M1/70) and anti-B220 (RA3-6B2), with sorting for $7AAD^{-}CD25^{+}CD44^{-}Lineage^{-}$. To obtain DP thymocytes, total thymocytes were stained with 7AAD, anti-CD4 (GK1.5) and anti-CD8 (53-6.7) with sorting for $7AAD^{-}CD4^{+}CD8^{+}$. To obtain B cells, splenocytes were stained with 7AAD, APC-anti-B220 and PE-Cy5-conjugated anti-CD3ε, anti-CD4 (GK1.5), anti-CD8 and anti-CD11b, with sorting for $7AAD^{-}B220^{+}CD3^{-}CD4^{-}CD8^{-}CD11b^{-}$. All antibodies were purchased from Biolegend, unless otherwise stated.

Chromatin conformation capture (3C)

3C assays were performed essentially as described[42], starting with 8–10 × 10^{6} thymocytes cross-linked in 8 ml RPMI containing 10% fetal bovine serum and 2% paraformaldehyde for 10 min at 25 °C. *Hind*III (NEB) was used to digest chromatin. 3C products were quantified by Taqman-based quantitative real-time PCR as described[14]. The sequences of probes and PCR primers are shown in Supplementary Table 1. To generate control PCR templates, bacterial artificial chromosomes bMQ-440L6 and bMQ-464f17 (Source BioScience) were mixed in equimolar amounts, and were digested and religated. bMQ-440L6 spans proximal V gene segments from *Trav19* to downstream of *Trdv2-2*, whereas bMQ-464f17 spans from INT1-2 to the central J_{α} segments. This control template mixture was used to generate standard curves for all 3C-qPCR assays.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Circular chromosome conformation capture-sequencing (4C-seq)

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 10^7 cells as described[14], using a *Hind*III restriction digest. Following generation of the 3C library, secondary digestion and re-ligation were performed as described[43], with modifications. 3C libraries were digested with 200 U of *Dpn*II overnight at 37 °C and reactions were supplemented with an additional 200 U of *Dpn*II for 6 h at 37 °C. The digested libraries were purified by phenol/chloroform extraction, precipitated with 2.5 vol ethanol, and rehydrated in 4 ml 30 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 1 mM DTT and 0.1 mM ATP. 200 U T4 DNA ligase (NEB) were added and libraries were incubated overnight at 16 °C. The reaction was then supplemented with an additional 200 U T4 DNA ligase for a minimum of 6 h at 16 °C. 4C libraries were then purified using phenol/chloroform extraction, precipitated with 2.5 vol ethanol, and rehydrated in 200 μl 10mM Tris-HCl, pH 8.0, 0.1 mM EDTA. Inverse PCR was then performed for the TEA, INT1, INT2, and E_δ viewpoints to generate libraries for high-throughput sequencing. All PCR reactions used Phusion polymerase in 1x Phusion HiFi buffer (NEB). For TEA, INT2, and E_δ, two separate PCR reactions were used to generate libraries. First round PCR was conducted with primers TEA-F (5′-TGCCATCTCTTACTGGGATC-3′) and TEA-R (5′-CATAACAGTAACCCAGCAAGC-3′), INT2-F (5′-TCCCTTATCTACAAGAGTCTGC-3′) and INT2-R (5′-TAGTCCATCACAAAGTAAGCTT-3′), and E_δ-F (5′-GGAAGTACAGTGCTGTCAAGC-3′) and E_δ-R (5′-CCACAATCTTCTTGGATGATC-3′). PCR conditions for TEA and E_δ were: 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 final extension for 10 min at 72 °C. PCR conditions for INT2 were identical except that annealing was at 55 °C. Products from the first PCR were purified using QiaQuick PCR purification reagents (Qiagen) and UPrep spin columns (Genesee), and subjected to second round PCR with versions of the F and R primers that added to their 5′ ends Illumina T5 adaptors: Adaptor 1-TEA-F and Adaptor 2-TEA-R; Adaptor 2-INT2-F and Adaptor 1- INT2-R; and Adaptor 2-E_δ-F and Adaptor 1-E_δ-R, where adaptor 1 is 5′-AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCT -3′ and adaptor 2 is 5′-CAAGCAGAAGACGGCATACGA-3′. PCR conditions for TEA and E_δ were: 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 final extension for 10 min at 72 °C. PCR conditions for INT2 were identical except that annealing was at 58 °C. For the INT1 viewpoint, one PCR of 30 cycles was performed using primers INT1-F (Adaptor 2- 5′-AGAAGGGGAGGAATCTGTTG-3′) and INT1-R (Adaptor 1- 5′-ACTGACAAGCAGCAAGAAGC-3′) with annealing at 58 °C. For both rounds of PCR, ten individual PCR reactions were run and 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 using PicoGreen (Life Technologies), were multiplexed by pooling in equimolar ratios, and were supplemented by addition of either a 15% or 30% spike of PhiX control library (Illumina). Prior to sequencing, pooled libraries were quality-tested using the Bioanalyzer platform

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(Agilent). Multiplexed libraries were then subjected to 50bp single-end sequencing using the Illumina HiSeq 2000 platform.

Sequencing data was analyzed using a workflow modified from that described[43]. FASTQ files containing raw multiplexed data were split using viewpoint-specific primer sequences TEA-F, INT1-R, INT2-R and E_{δ}-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 using Bowtie, allowing 0 mismatches and sequences repeated up to 10 times to be aligned (-v 0 -m 10 -all -best -strata). A map of genomic *Hind*III restriction fragments was generated, and reads per *Hind*III restriction fragment were counted using Python scripts as described[43] and visualized using the UCSC Genome Browser. Data were expressed as reads per million mapped sequence reads.

***Tcrd* and *Tcra* recombination—**Genomic DNA was isolated from sorted DN3 or DP thymocytes to analyze *Tcrd* or *Tcra* recombination, respectively. *Tcrd* rearrangements were quantified by Taqman-qPCR using primers and probes described in Supplementary Table 1 and PCR conditions identical to those for 3C Taqman-qPCR[14]. *Tcra* rearrangements were quantified by SYBR Green qPCR (Qiagen) as described[14], using primers shown in Supplementary Table 1. In both instances *Cd14* PCR was used for normalization.

Retention of chromosomal DNA—Genomic DNA isolated from total thymocytes of WT and INT1-2KO littermates was quantified by SYBR Green qPCR (Qiagen) as discribed[14]. Amplicon abundance in thymocyte DNA was compared to that in kidney DNA using a kidney DNA standard curve. Samples were initially normalized to each other based on the abundance of E_{α}; retention of the TEA amplicon in WT was then set to 100% and amplicons in both genotypes were expressed relative to this value. PCR amplicons were located upstream of the identified gene segments. Although V_{α}-J_{α} rearrangement will excise amplicons onto extrachromosomal circles, this material should be retained in DP thymocyte genomic DNA preparations assuming that no thymocyte proliferation occurs after V_{α}-J_{α} rearrangement. In practice, and consistent with previous work[44], we observed that retention of TEA was 50% that of E_{α}, indicating that some proliferation occurs after V_{α}-J_{α} rearrangement. Loss of signal due to V_{α}-J_{α} rearrangement was controlled for by setting TEA amplicon abundance to 100% in WT.

Chromatin immunoprecipitation (ChIP)—6-8 × 10^6 thymocytes were cross-linked in 1ml RPMI containing 10% fetal bovine serum and 1% paraformaldehyde for 10 min at 25 °C. Cross-linked thymocytes were washed in PBS, pelleted and incubated in 1 ml of 5 mM PIPES, pH 8.0, 85 mM KCl, 0.5% NP-40 for 10 min on ice, after which they were disrupted by Dounce homogenization using 15 strokes with pestle “A”. Nuclei were precipitated, washed, and lysed in 500 μl of 50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS. Chromatin was sheared using a Sonicator 3000 (Qsonica) for 4.5 min (six cycles of 15 s on, 30 s off at power=2). For one ChIP experiment, 200 μl sonicated chromatin was diluted 10-fold and precipitated with 5 μl anti-CTCF (07-729; Millipore) or anti-H3Ac (06-599; Millipore) or 5 μg control rabbit IgG (ab-105-c; R&D Systems). Immune complexes were isolated with Protein A agarose/salmon sperm DNA (Millipore), washed, eluted and incubated at 65 °C for 4 h to reverse cross-links. DNA was purified by phenol:chloroform

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

extraction and isopropanol precipitation. Enrichment of chromatin was measured by qPCR as previously described[14] with primers listed in Supplementary Table 1. Data from CTCF-ChIP and H3ac-ChIP were expressed as bound/input and then normalized to values for *Myc* and *B2m*, respectively.

CBE orientation analysis—FASTA sequences corresponding to called peaks from CTCF ChIP-seq data in DN thymocytes[14] were obtained using 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) using default parameters to scan both strands for one or zero occurrences of a particular 6–30 bp motif per input sequence. The top-scoring motif matched the previously defined CTCF binding motif from nucleotides 5–20 (ref. 45). Individual sequences were then manually curated to eliminate those that corresponded to a very minor CTCF ChIP peak, did not align to the center of a CTCF ChIP peak, or were ambiguous with respect to orientation.

Statistical methods—Data were analyzed by 2-way ANOVA or by unpaired two-tailed Student's *t*-test with corrections for multiple comparisons, as appropriate, using 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank C. Bock of the Duke University Knockout and Transgenic Mouse Resource for the production of gene-targeted mice, Y.-W. He for the pGKneoF2L2DTA targeting vector, O. Fedrigo of the Duke University Genome Sequencing & Analysis Core Resource for sequencing of 4C samples, D. Corcoran of the Duke Integrative Genomic Analysis Shared Resource for advice on the processing of 4C sequencing data, L. Martinek, N. Martin and M. Cook of the Duke Cancer Institute Flow Cytometry Facility for help with cell sorting and analysis, J. Liang for technical advice, and Z. Huang for technical support. We thank E. Oltz (Washington University) and Y. Zhuang (Duke University) for their thoughtful comments on the manuscript. Supported by the National Institutes of Health (R37 GM41052 to M.S.K.).

References

1. Schatz DG, Swanson PC. V(D)J recombination: mechanisms of initiation. Annu Rev Genet. 2011; 45:167–202. [PubMed: 21854230]
2. Schatz DG, Ji Y. Recombination centres and the orchestration of V(D)J recombination. Nat Rev Immunol. 2011; 11:251–263. [PubMed: 21394103]
3. Roldán E, et al. Locus 'decontraction' and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. Nat Immunol. 2005; 6:31–41. [PubMed: 15580273]
4. Skok JA, et al. Reversible contraction by looping of the *Tcra* and *Tcrb* loci in rearranging thymocytes. Nat Immunol. 2007; 8:378–387. [PubMed: 17334367]
5. Jhunjhunwala S, et al. The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. Cell. 2008; 133:265–279. [PubMed: 18423198]
6. Jhunjhunwala S, van Zelm MC, Peak MM, Murre C. Chromatin architecture and the generation of antigen receptor diversity. Cell. 2009; 138:435–448. [PubMed: 19665968]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

7. Shih HY, Krangel MS. Distinct contracted conformations of the *Tcra/Tcrd* locus during *Tcra* and *Tcrd* recombination. J Exp Med. 2010; 207:1835–1841. [PubMed: 20696701]
8. Degner SC, et al. CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the *Igh* locus and antisense transcription in pro-B cells. Proc Natl Acad Sci USA. 2011; 108:9566–9571. [PubMed: 21606361]
9. Guo C, et al. CTCF-binding elements mediate control of V(D)J recombination. Nature. 2011; 477:424–430. [PubMed: 21909113]
10. Guo C, et al. Two forms of loops generate the chromatin conformation of the immunoglobulin heavy-chain gene locus. Cell. 2011; 147:332–343. [PubMed: 21982154]
11. Ribeiro de Almeida C, et al. The DNA-Binding Protein CTCF Limits Proximal V_κ Recombination and Restricts κ Enhancer Interactions to the Immunoglobulin κ Light Chain Locus. Immunity. 2011; 35:501–513. [PubMed: 22035845]
12. Seitan VC, et al. A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. Nature. 2011; 476:467–471. [PubMed: 21832993]
13. Lin YC, et al. Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. Nat Immunol. 2012; 13:1196–1204. [PubMed: 23064439]
14. Shih HY, et al. *Tcra* gene recombination is supported by a *Tcra* enhancer- and CTCF-dependent chromatin hub. Proc Natl Acad Sci USA. 2012; 109:E3493–3502. [PubMed: 23169622]
15. Verma-Gaur J, et al. Noncoding transcription within the *Igh* distal V_H region at PAIR elements affects the 3D structure of the *Igh* locus in pro-B cells. Proc Natl Acad Sci USA. 2012; 109:17004–17009. [PubMed: 23027941]
16. Medvedovic J, et al. Flexible long-range loops in the V_H gene region of the *Igh* locus facilitate the generation of a diverse antibody repertoire. Immunity. 2013; 39:229–244. [PubMed: 23973221]
17. Majumder 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; 212:107–120. [PubMed: 25512470]
18. Ong CT, Corces VG. CTCF: an architectural protein bridging genome topology and function. Nat Rev Genet. 2014; 15:234–246. [PubMed: 24614316]
19. Xiang Y, Park SK, Garrard WT. A major deletion in the V_κ -J_κ intervening region results in hyperelevated transcription of proximal V_κ genes and a severely restricted repertoire. J Immunol. 2014; 193:3746–3754. [PubMed: 25187654]
20. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature. 2012; 489:109–113. [PubMed: 22955621]
21. Shen Y, et al. A map of the cis-regulatory sequences in the mouse genome. Nature. 2012; 488:116–120. [PubMed: 22763441]
22. Phillips-Cremins JE, et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell. 2013; 153:1281–1295. [PubMed: 23706625]
23. Symmons O, et al. Functional and topological characteristics of mammalian regulatory domains. Genome Res. 2014; 24:390–400. [PubMed: 24398455]
24. Krangel MS, Carabana J, Abbarategui I, Schlimgen R, Hawwari A. Enforcing order within a complex locus: current perspectives on the control of V(D)J recombination at the murine T-cell receptor α/δ locus. Immunol Rev. 2004; 200:224–232. [PubMed: 15242408]
25. Bosc N, Lefranc MP. The mouse (Mus musculus) T cell receptor α (TRA) and δ (TRD) variable genes. Dev Comp Immunol. 2003; 27:465–497. [PubMed: 12697305]
26. Zhao Z, et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet. 2006; 38:1341–1347. [PubMed: 17033624]
27. Hao B, Krangel MS. Long-distance regulation of fetal V_δ gene segment TRDV4 by the *Tcrd* enhancer. J Immunol. 2011; 187:2484–2491. [PubMed: 21784972]
28. Rudan MV, et al. Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. Cell Rep. 2015; 10:1297–1309. [PubMed: 25732821]
29. Rao SS, et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell. 2014; 159:1665–1680. [PubMed: 25497547]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

30. Weber-Arden J, Wilbert OM, Kabelitz D, Arden B. V_{δ} repertoire during thymic ontogeny suggests three novel waves of γδ TCR expression. J Immunol. 2000; 164:1002–1012. [PubMed: 10623850]
31. Monroe RJ, et al. Developmental regulation of TCRδ locus accessibility and expression by the TCRδ enhancer. Immunity. 1999; 10:503–513. [PubMed: 10367896]
32. Abarrategui I, Krangel MS. Noncoding transcription controls downstream promoters to regulate T-cell receptor α recombination. EMBO J. 2007; 26:4380–4390. [PubMed: 17882258]
33. Ji Y, et al. Promoters, enhancers, and transcription target RAG1 binding during V(D)J recombination. J Exp Med. 2010; 207:2809–2816. [PubMed: 21115692]
34. Hawwari A, Krangel MS. Role for rearranged variable gene segments in directing secondary T cell receptor α recombination. Proc Natl Acad Sci USA. 2007; 104:903–907. [PubMed: 17210914]
35. Pasqual N, et al. Quantitative and qualitative changes in V-J_{α} and qualitative changes in V-thymocytes differentiation: implication for a limited T cell receptor α chain repertoire. J Exp Med. 2002; 196:1163–1173. [PubMed: 12417627]
36. Jouvin-Marche E, Fuschiotti P, Marche PN. Dynamic aspects of TCRα gene recombination: qualitative and quantitative assessments of the TCRα chain repertoire in man and mouse. Adv Exp Med Biol. 2009; 650:82–92. [PubMed: 19731803]
37. Genolet R, Stevenson BJ, Farinelli L, Osteras M, Luescher IF. Highly diverse TCRα chain repertoire of pre-immune $CD8^{+}$ T cells reveals new insights in gene recombination. EMBO J. 2012; 31:4247–4248. [PubMed: 23128857]
38. Bassing CH, et al. T cell receptor (TCR) α/δ locus enhancer identity and position are critical for the assembly of TCRδ and α variable region genes. Proc Natl Acad Sci USA. 2003; 100:2598–2603. [PubMed: 12604775]
39. Shrimali S, et al. An ectopic CTCF-dependent transcriptional insulator influences the choice of V_{β} gene segments for VDJ recombination at *Tcrb* locus. Nucleic Acids Res. 2012; 40:7753–7765. [PubMed: 22718969]
40. Giorgetti L, et al. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. Cell. 2014; 157:950–963. [PubMed: 24813616]
41. Passoni L, et al. Intrathymic δ selection events in γδ cell development. Immunity. 1997; 7:83–95. [PubMed: 9252122]
42. Hagege H, et al. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat Protoc. 2007; 2:1722–1733. [PubMed: 17641637]
43. Stadhouders R, et al. Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. Nat Protoc. 2013; 8:509–524. [PubMed: 23411633]
44. Livak F, Schatz DG. T-cell receptor α locus V(D)J recombination byproducts are abundant in thymocytes and mature T cells. Mol Cell Biol. 1996; 16:609–618. [PubMed: 8552089]
45. Nakahashi H, et al. A genome-wide map of CTCF multivalency redefines the CTCF code. Cell Rep. 2013; 3:1678–1689. [PubMed: 23707059]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

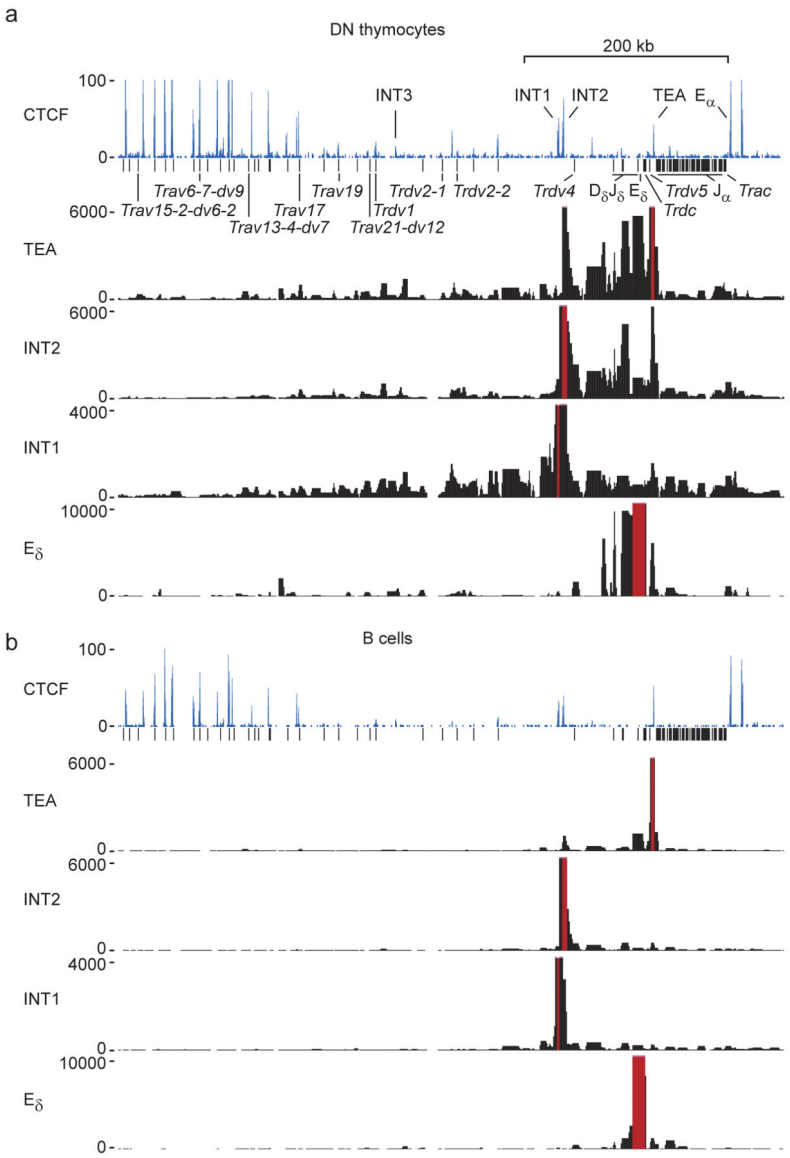

Figure 1.

Long-range interaction network within the *Tcra-Tcrd* locus. (**a**) (Top) CTCF binding to the 3' portion of the C57BL/6 *Tcra-Tcrd* locus in DN thymocytes (GEO accession number GSE41743)[14]. Several CTCF binding elements (CBEs) are labeled above the CTCF track. Black vertical lines below the CTCF track mark gene segments, a subset of which are labeled. (Bottom) Interactomes of TEA, INT2, INT1 and E_δ viewpoints determined by 4C-seq analysis of *Rag2*$^{-/-}$ thymocytes (C57BL/6 background). (**b**) (Top) CTCF binding and (Bottom) 4C-seq analyses of splenic B cells. Sequence reads were averaged from two independent experiments for each cell type and were mapped to *Hind*III fragments. The viewpoint *Hind*III fragment is marked in red.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

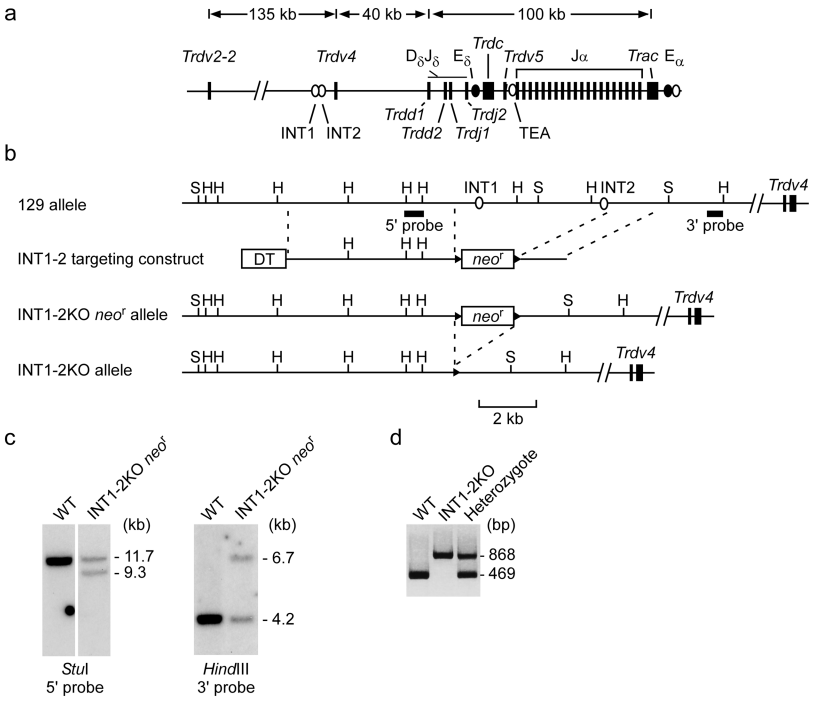

Figure 2.

Generation of INT1-2KO mice. (**a**) The relative positions of gene segments (black rectangles), enhancers (black ovals), and CBEs (white ovals) within the 3' 300 kb of the *Tcra-Tcrd* locus. (**b**) WT 129/SvJ allele (129), targeting construct INT1-2, neomycin-resistant allele INT1-2KO *neo*^{r}, and *neo*-deleted allele INT1-2KO are shown. DT, diphtheria toxin cassette; H, *Hind*III; S, *Stu*I; Southern blot probes are indicated. (**c**) Southern blot analyses of genomic DNA from WT and INT1-2KO neo^{r}-targeted ES cells. Results are representative of 3 and 2 experiments, respectively, with the 5' and 3' probes. (**d**) Genotyping PCR of WT, homozygous INT1-2KO or heterozygous INT1-2KO littermates. Results are representative of >3 experiments.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

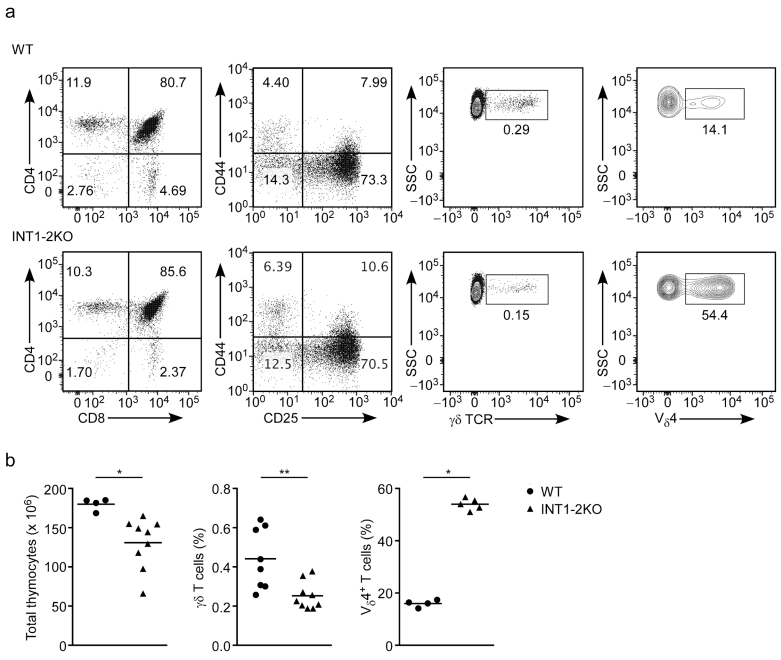

Figure 3.

Thymocyte development in INT1-2KO mice. (**a**) Flow cytometry analysis of thymocytes from WT and INT1-2KO littermates. CD4-CD8 and γδ TCR staining are shown for total thymocytes; CD44-CD25 staining is shown for DN thymocytes depleted of $CD4^+$ and $CD8^+$ cells and pre-gated as follows: $7AAD^-CD4^-CD8^-CD11b^-Ter119^-B220^-Gr\text{-}1^-CD3\varepsilon^-$; $V_\delta4$ staining is shown for pre-gated γδ TCR^+ thymocytes. Data are representative of at least 3 independent experiments. (**b**) Number of total thymocytes (left), abundance of γδ TCR^+ thymocytes as a percentage of total thymocytes (middle), and percentage of $V_\delta4^+$ thymocytes among pregated γδ TCR^+ thymocytes (right) in WT and INT1-2KO mice. Each data point represents an individual mouse and the horizontal line indicates the mean. Statistical significance was evaluated by unpaired Student's t-test (left) or Mann-Whitney U-test (middle, right). *, $P<0.05$; **, $P<0.01$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

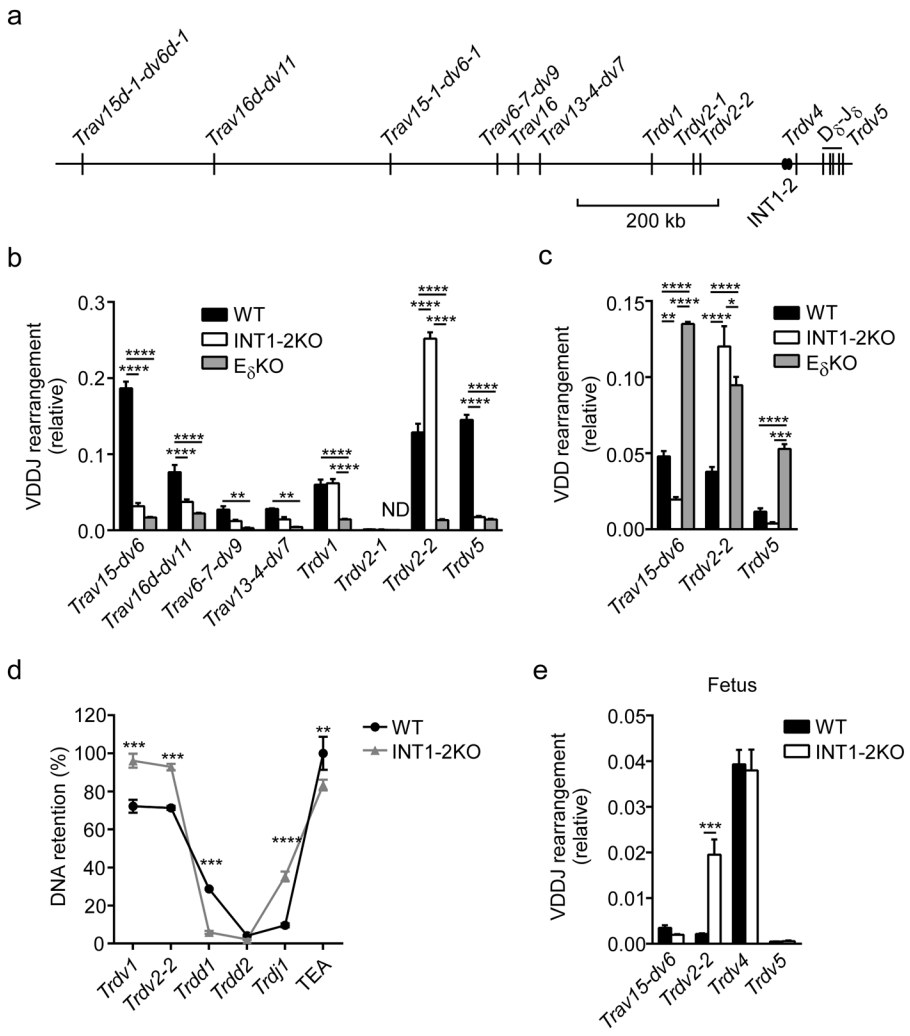

Figure 4.

Restricted TCRδ repertoire in INT1-2KO mice. (**a**) Locus map identifying V_δ gene segments analyzed. Genomic DNA extracted from DN3 thymocytes from WT, INT1-2KO and E_δKO mice was analyzed for (**b**) VDD-*Trdj1* rearrangements or (**c**) VD-*Trdd2* rearrangements by Taqman-qPCR with normalization to *Cd14*. The *Trav15-dv6* PCR detects *Trav15-1-dv6-1* and *Trav15d-1-dv6d-1*. The *Trav16d-dv11* PCR detects *Trav16d-dv11* and *Trav16*. Data represent the mean ± SEM of 3 WT, 3 INT1-2KO and 2–3 E_δKO samples, with each sample representing a pool of 2–3 mice. Statistical significance was evaluated by 2-way ANOVA with Tukey's multiple comparison test. (**d**) Rearrangement was quantified by measuring retention of chromosomal DNA in total thymocytes as compared to kidney using SYBR Green qPCR. Data represent the mean ± SEM of 3 WT and 3 INT1-2KO samples, with each sample representing an individual mouse. Samples were initially normalized to each other based on the abundance of E_α; retention of the TEA amplicon in WT was then set to 100% and amplicons in both genotypes were expressed relative to this value. PCR amplicons were located upstream of the identified gene segments. Statistical significance was evaluated by 2-way ANOVA with Sidak's multiple comparison test. (**e**) Genomic DNA extracted from

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

E15.5 WT and INT1-2KO fetal thymi was analyzed for VDD-*Trdj1* rearrangements by SYBR Green qPCR, with normalization to *Cd14*. Data represent the mean ± SEM of 3 WT and 4 INT1-2KO samples, with each sample representing a pool of 2–3 mice. Statistical significance was evaluated by 2-way ANOVA with Sidak's multiple comparison test. *, $P<0.05$; **, $P<0.01$; ***, $P<0.001$; ****, $P<0.0001$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

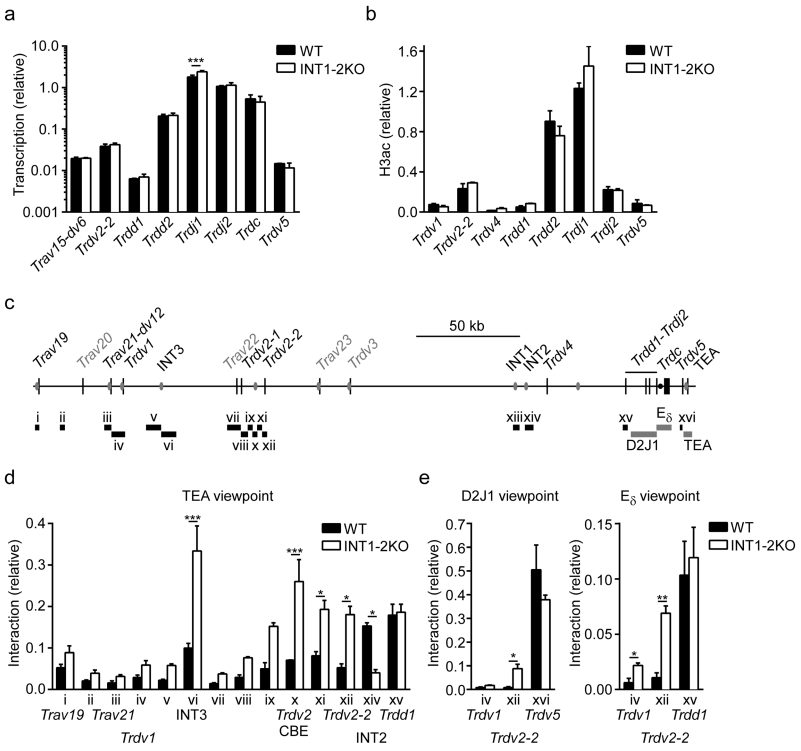

Figure 5.

INT1-2-deletion alters chromatin loop organization but not chromatin accessibility. (**a**) Germline transcription was analyzed in WT and INT1-2KO DN thymocytes (both on a *Rag2*$^{-/-}$ background). Data represent the mean ± SEM of 2-4 WT and 2-3 INT1-2KO cDNA preparations, each representing a pool of 2-3 mice, with all values normalized to *Hprt*. Statistical analysis was by 2-way ANOVA with Sidak's multiple comparison test. (**b**) Histone H3 acetylation (H3ac) was analyzed in WT and INT1-2KO DN thymocytes (both on a *Rag2*$^{-/-}$ background). Data represent the mean ± SEM of 2-3 WT and 3 INT1-2KO chromatin preparations, each representing a pool of 8-10 mice, with values of bound/input normalized to values for *B2m*. Statistical analysis was as in (**a**). (**c**) Long-distance interactions analyzed by 3C. CBEs (gray ovals) are indicated on the map. Viewpoint (gray rectangles) and target (numbered black rectangles) *Hind*III fragments are shown below the map. V gene segments shaded gray are pseudogenes. (**d**) WT and INT1-2KO DN thymocytes (both on a *Rag2*$^{-/-}$ background) were analyzed by 3C from the TEA viewpoint. Data represent the mean ± SEM of 3–5 WT and 3–6 INT1-2KO preparations, with normalization to a TEA nearest neighbor fragment. Statistical analysis was as in (**a**). (**e**) Similar 3C analyses from D2J1 and E_δ viewpoints. Data for D2J1 represent the mean ± SEM of 3 WT and 3 INT1-2KO preparations, with normalization to interaction between TEA and its neighbor fragment. Data for E_δ represent the mean ± SEM of 4 WT and 3–4 INT1-2KO preparations, with interactions normalized to an E_δ nearest neighbor fragment. Statistical analysis was by unpaired Student's *t*-test with Holm-Sidak correction for multiple comparisons. All 3C preparations represent pools of 8–10 mice. *, $P<0.05$; **, $P<0.001$; ***, $P<0.0001$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

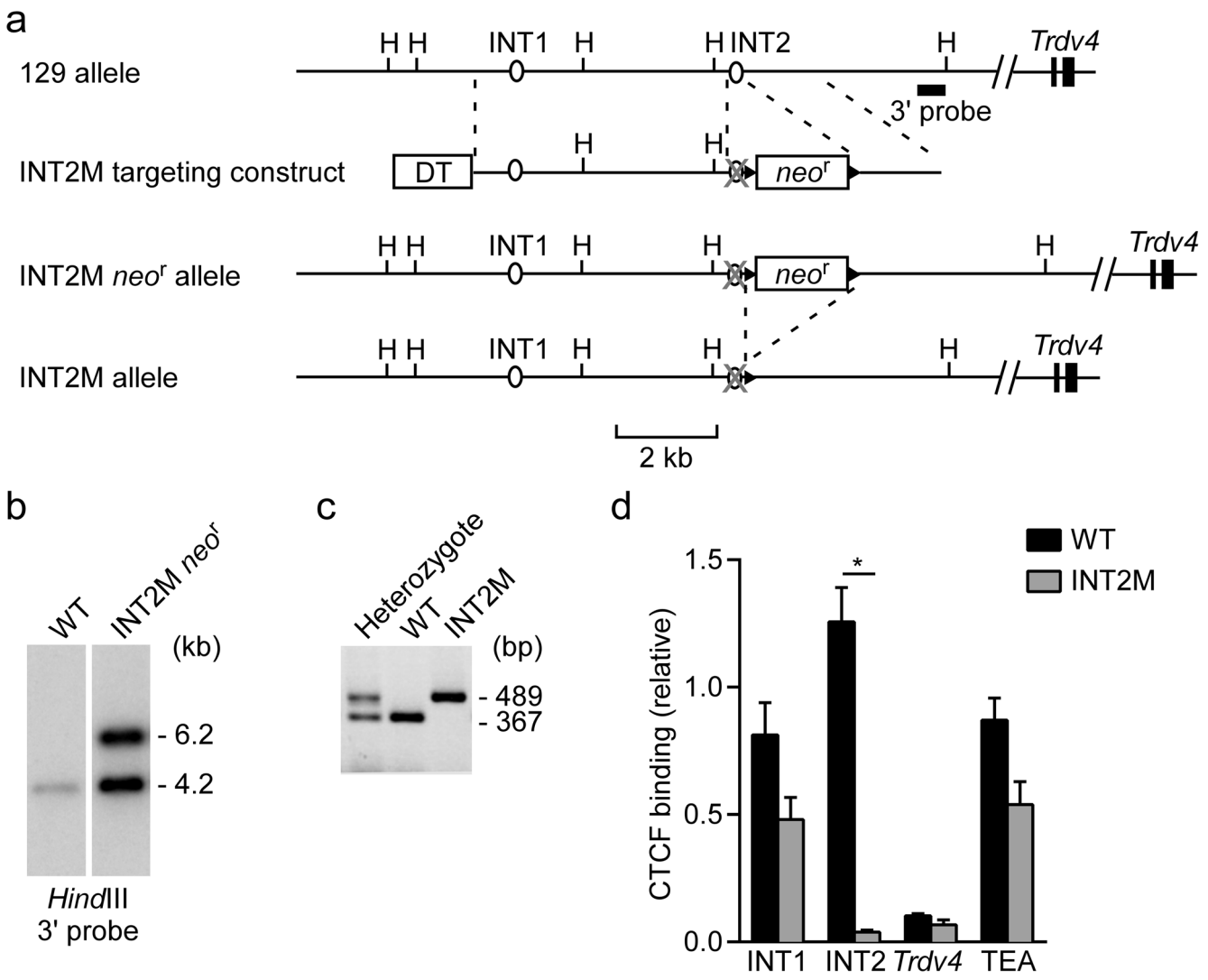

Figure 6.

Generation of INT2 mutant (INT2M) mice. (**a**) WT 129/SvJ allele (129), targeting construct INT2M, neomycin-resistant allele INT2M *neo*^{r}, and *neo*-deleted allele INT2M are shown. DT, diphtheria toxin cassette; H, *Hind*III; S, *Stu*I; Southern blot probe is indicated. (**b**) Southern blot analysis of genomic DNA from WT and INT2M *neo*^{r}-targeted ES cells. Results are representative of 2 experiments. (**c**) Genotyping PCR of INT2M heterozygous, WT and INT2M homozygous littermates. Results are representative of >3 experiments. (**d**) ChIP analysis of CTCF binding to WT and INT2M alleles in *Rag2*$^{-/-}$ and *Rag2*$^{-/-}$INT2M thymocytes, respectively. *Trdv4* served as a negative control. Data represent the mean ± SEM of 3 WT and 2 INT2M samples, with each sample representing a pool of 2-3 mice. Statistical significance was determined by unpaired Student's *t*-test with Holm-Sidak correction for multiple comparisons. *, $P<0.01$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

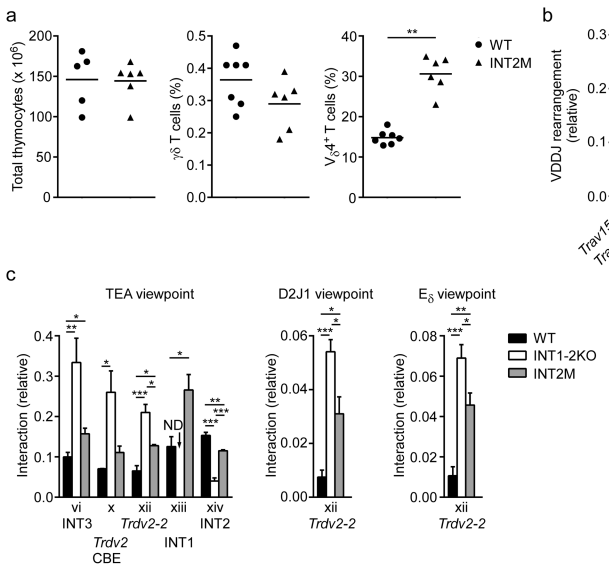

Figure 7.

Partial redundancy between INT1 and INT2. (**a**) Number of total thymocytes (left), abundance of γδ TCR$^+$ thymocytes as a percentage of total thymocytes (middle), and percentage of $V_\delta4^+$ thymocytes among pre-gated γδ TCR$^+$ thymocytes (right) in WT and INT2M mice. Each data point represents an individual mouse and the horizontal line indicates the mean. Statistical significance was evaluated by unpaired Student's t-test (left) or Mann-Whitney U-test (middle, right). (**b**) Genomic DNA extracted from DN3 thymocytes from 3–4 week old WT and INT2M mice was analyzed for VDD-*Trdj1* rearrangement by Taqman-qPCR, with normalization to *Cd14*. Data represent the mean ± SEM of 3 WT and 3 INT2M preparations, with each preparation representing a pool of 2–3 mice. Statistical significance was evaluated by 2-way ANOVA with Sidak's multiple comparison test. (**c**) Long-distance interactions analyzed by 3C. WT, INT1-2KO and INT2M DN thymocytes (all on a *Rag2*$^{-/-}$ background) were analyzed by 3C from the TEA, D2J1 and E_δ viewpoints, with normalization as in Fig. 5d,e. Data represent the mean ± SEM of 3–5 WT, 3–6 INT1-2KO and 3–4 INT2M preparations for the TEA viewpoint, 4 WT, 4 INT1-2KO and 3 INT2M preparations for the D2J1 viewpoint, and 4 preparations of each genotype for the E_δ viewpoint. All 3C preparations represent pools of 8–10 mice. TEA viewpoint data for WT and INT1-2KO sites vi, x and xiv are identical to Fig. 5d. Statistical significance was evaluated by unpaired Student's *t*-test with Holm-Sidak correction for multiple comparisons (TEA viewpoint) or by unpaired Student's *t*-test (D2J1 and E_δ viewpoints). *, $P<0.05$; **, $P<0.01$; ***, $P<0.001$; ****, $P<0.0001$. ND, not determined.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

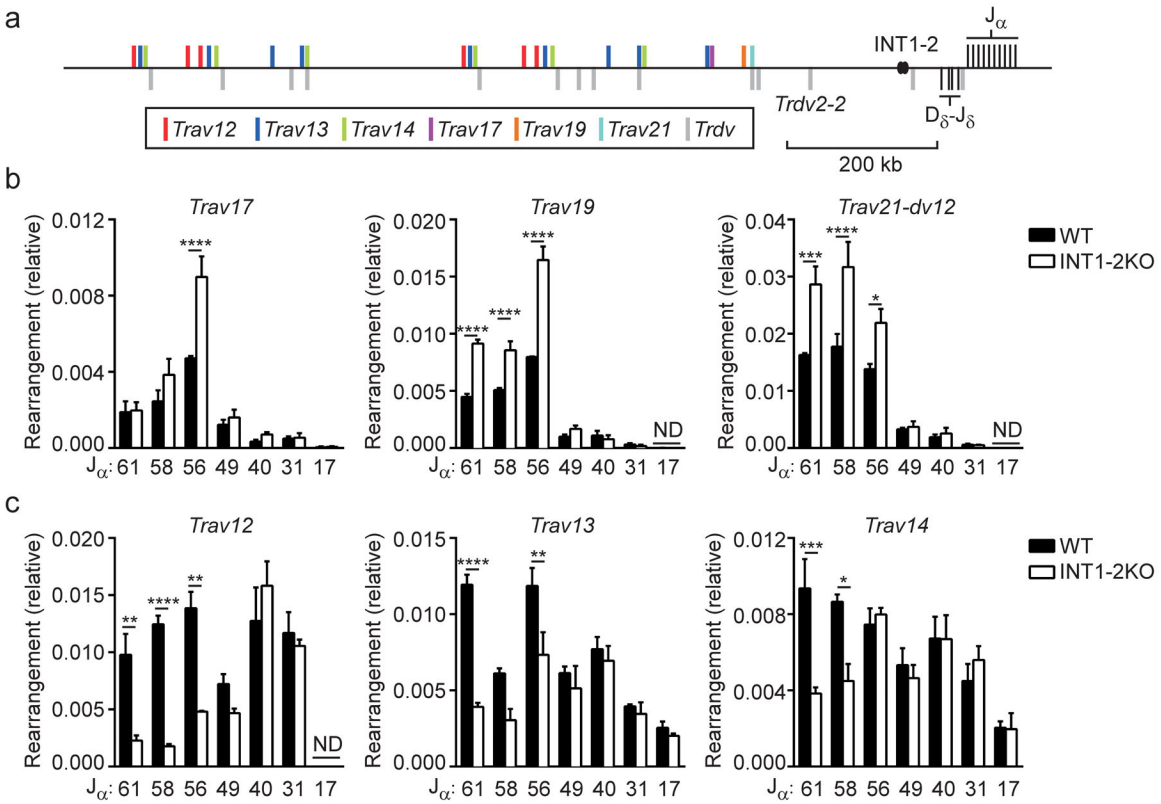

Figure 8.

Restricted TCRα repertoire in INT1-2KO mice. (**a**) Partial locus map, with *Tcra* and *Tcrd* gene segments denoted above and below the horizontal line, respectively. D and J segments are in black, selected *Trav* families are color coded, and *Trdv* segments are in gray. Aligned *Trav* and *Trdv* designations indicate V segments designated as *Trav-Trdv*. Genomic DNA extracted from DP thymocytes from WT and INT1-2KO mice was analyzed for rearrangement of (**b**) J_α-proximal V_α segments and (**c**) J_α-distal V_α-segments to different J_α segments by SYBR Green-qPCR with normalization to *Cd14*. Data represent the mean ± SEM of 3 preparations for each genotype, each preparation representing a different mouse. Statistical significance was evaluated by 2-way ANOVA with Sidak's multiple comparison test. *, $P<0.05$; **, $P<0.01$; ***, $P<0.001$; ****, $P<0.0001$. ND, not detected.