



# TCF-1 promotes chromatin interactions across topologically associating domains in T cell progenitors

Wenliang Wang<sup>1,2,3,4,6</sup>, Aditi Chandra<sup>1,2,3,4,6</sup>, Naomi Goldman<sup>1,2,3,4,6</sup>, Sora Yoon<sup>1,2,3,4,6</sup>, Emily K. Ferrari<sup>1,2,3,4,6</sup>, Son. C. Nguyen<sup>1,3</sup>, Eric F. Joyce<sup>1,3</sup> and Golnaz Vahedi<sup>1,2,3,4,5</sup> ✉

**The high mobility group (HMG) transcription factor TCF-1 is essential for early T cell development. Although in vitro biochemical assays suggest that HMG proteins can serve as architectural elements in the assembly of higher-order nuclear organization, the contribution of TCF-1 on the control of three-dimensional (3D) genome structures during T cell development remains unknown. Here, we investigated the role of TCF-1 in 3D genome reconfiguration. Using gain- and loss-of-function experiments, we discovered that the co-occupancy of TCF-1 and the architectural protein CTCF altered the structure of topologically associating domains in T cell progenitors, leading to interactions between previously insulated regulatory elements and target genes at late stages of T cell development. The TCF-1-dependent gain in long-range interactions was linked to deposition of active enhancer mark H3K27ac and recruitment of the cohesin-loading factor NIPBL at active enhancers. These data indicate that TCF-1 has a role in controlling global genome organization during T cell development.**

The mammalian genome is folded into higher-order 3D structures with globular interaction domains<sup>1</sup>. These structures, variously called topologically associating domains (TADs) or insulated neighborhoods<sup>2–6</sup>, sequester regulatory elements, insulating them from genomic regions outside such domains. The ubiquitously expressed protein CTCF has a prominent role in creating insulated neighborhoods through its convergent binding events, which can block cohesin-mediated loop extrusion<sup>7,8</sup>. A small number of lineage-determining transcription factors (LDTFs), which can endow competence to new developmental programs have been shown to control 3D genome structure<sup>8–12</sup> but the underlying mechanisms of this control are not fully understood.

The T cell developmental pathway represents a leading exemplar to understand global genome organization<sup>13</sup>. T cell development starts when bone marrow-derived precursors seed the thymus and give rise to early thymic progenitors (ETPs)<sup>14</sup>. TCF-1, encoded by *Tcf7*, is upregulated at ETP stage by Notch1 and remains highly expressed until maturation. This protein can control the expression of *Gata3* and *Bcl11b*, which are necessary for T lineage commitment<sup>15,16</sup>. TCF-1 can also bind to nucleosomal DNA and create new accessible regulatory elements<sup>17,18</sup>. Biochemical studies have indicated that the HMG proteins such as TCF-1 and LEF1 can bend DNA<sup>19,20</sup>. Moreover, TCF-1 is linked to 3D genome organization of peripheral CD8<sup>+</sup> T cells<sup>21</sup> and T leukemic cells<sup>22,23</sup>. Yet, the role of TCF-1 on genome folding during T cell development is unknown.

Here we assessed whether the early and continuous expression of TCF-1 in the thymus had a role in chromatin folding. Employing gain- and loss-of-function experiments, we observed that TCF-1 targeted boundaries of insulated neighborhoods in T cell progenitors, weakened the insulation between adjacent neighborhoods and

enhanced long-range interactions between regulatory elements and target genes located on previously insulated domains. TCF-1-dependent long-range interactions were linked to the recruitment of the cohesin-loading factor NIPBL to active enhancers. Our finding of TCF-1-dependent chromatin interactions across insulated neighborhoods demonstrates an LDTF that can dismantle insulation during a developmental trajectory.

## Results

**TCF-1 occupancy correlates with intra-TAD interactions.** To study the relationship between TCF-1 occupancy and 3D genome folding during T cell development, we employed publicly available chromatin immunoprecipitation sequencing (ChIP-seq) measurements of TCF-1 binding events in thymocytes<sup>24</sup> and Hi-C measurements of 3D genome interactions at different T cell developmental stages, including pre-commitment (common lymphoid progenitor (CLP), ETP and double negative (DN)2) and post-commitment (DN3, DN4 and double positive (DP)) stages<sup>25</sup>. We defined TADs in DPs using the insulation score strategy<sup>26</sup>, grouped TADs based on TCF-1 binding density and evaluated the extent of intra-TAD interactions. The increase in intra-TAD interactions in DPs compared to ETPs-DN4s was significantly correlated with the density of TCF-1 binding events within TADs (Fig. 1a), suggesting that genomic domains with the strongest TCF-1 occupancy acquired the largest extent of de novo 3D interactions in DPs.

To explore the association between TCF-1 binding and finer scale 3D structures, we identified cohesin-associated loops using SMC1 HiChIP in DPs<sup>27</sup>. Of the 149,315 loops detected in DPs, more than 40% were occupied by TCF-1 at least on one anchor, which was comparable to CTCF binding pattern (Fig. 1b) while 32% of

<sup>1</sup>Department of Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. <sup>2</sup>Institute for Immunology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. <sup>3</sup>Epigenetics Institute, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. <sup>4</sup>Institute for Diabetes, Obesity and Metabolism, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. <sup>5</sup>Abramson Family Cancer Research Institute, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. <sup>6</sup>These authors contributed equally: Wenliang Wang, Aditi Chandra, Naomi Goldman, Sora Yoon, Emily K. Ferrari. ✉e-mail: [vahedi@penmedicine.upenn.edu](mailto:vahedi@penmedicine.upenn.edu)

loops were co-bound by TCF-1 and CTCF (Fig. 1b). To evaluate the genome-wide co-occupancy of these two proteins, we grouped CTCF and TCF-1 binding events in DPs into CTCF alone, TCF-1 alone and regions co-occupied by both proteins (Fig. 1c and Supplementary Table 1). CTCF and TCF-1 co-occupancy accounted for ~20% of TCF-1 bound regions (Fig. 1c). The HMG motif was enriched at CTCF and TCF-1 co-bound sites (Fig. 1d), implying the direct binding of TCF-1 protein at these loci. The occupancy of CTCF and TCF-1 was associated with dense cohesin loops at the *Cd8a* (Fig. 1e), *Ets1* (Extended Data Fig. 1a) and *Tcf7* (Extended Data Fig. 1b) loci.

We next assessed the extent of long-range interactions between the regulatory elements in each TCF-1 binding group and genes that are selectively regulated during T cell development using single-cell RNA-seq profiling of thymocytes<sup>28</sup> (Supplementary Table 2) and found that TCF-1 and CTCF co-bound sites had the strongest interactions with stage-specific genes (Fig. 1f). Thus, TCF-1 and CTCF co-occupancy might be a structurally salient feature of T cell development.

**TCF-1 and CTCF co-binding weakens insulation at boundaries.** We next measured the average 3D interactions anchored at the CTCF, TCF-1 or TCF-1+CTCF peaks and a set of randomly sampled regions that lacked either protein as control. In CLPs, ETPs and DN2s, TCF-1+CTCF sites had limited interactions between their upstream and downstream regions (Fig. 2a). The insulation at TCF-1+CTCF sites was gradually lost starting in DN3s (Fig. 2a), suggesting interactions between upstream and downstream insulated domains centered at TCF-1+CTCF sites. The interactions between the two sides of TCF-1-only regions also started to increase in DN3s (Fig. 2a), while the CTCF-occupied sites remained insulated throughout T cell development (Fig. 2a). To quantitatively assess changes in insulation at TCF-1+CTCF sites, we calculated genome-wide insulation scores<sup>26</sup>. We found that insulation at TCF-1+CTCF sites, but not CTCF-only sites, gradually declined during T cell development starting from DN3s (Fig. 2b and Extended Data Fig. 2a). The increase in insulation score in DPs compared to CLPs was significantly larger at TCF-1+CTCF sites compared to TCF-1 or CTCF regions (Extended Data Fig. 2b), indicating that co-binding of TCF-1 and CTCF could increase interactions across insulated neighborhoods, leading to a loss of insulation as T cells become more mature. Co-bound sites were preferentially enriched closer to genes expressed at DN3 and DN4 stages<sup>28</sup> compared to genes associated with other stages (Fig. 2c and Supplementary Table 3). Thus, insulation at TCF-1+CTCF sites declined post-T cell commitment due to de novo long-range interactions at T cell commitment genes.

Based on the quantitative definition of TAD boundaries, the number of boundaries steadily decreased from ETP to DP (Extended Data Fig. 2c), suggesting the reorganization of TADs during T cell

development. We further classified the boundaries into different groups according to their quantitative detection at each stage. We detected boundaries with mixed patterns of gain or loss (group I), boundaries that remained invariant during development (group H), boundaries that formed after T cell commitment (groups A and B) and boundaries that were lost at different stages (groups C–G) (Extended Data Fig. 2d and Supplementary Table 4). Examination of the long-range contacts across the nine groups indicated increased cross-boundary interactions at DN3s to DPs compared to CLPs to DN3s at lost boundaries and fewer cross-boundary interactions in DN3s to DPs at gained boundaries (Extended Data Fig. 2e,f). The interactions across domains separated by dynamic boundaries were confirmed by insulation score and directionality index analyses (Extended Data Fig. 3a,b). The percentage of group C boundaries that were in compartment B ( $PC1 < 0$ ) increased markedly between CLPs and DPs (Extended Data Fig. 3c). Lost boundaries in groups E and F were largely in compartment A in both CLPs and DPs (Extended Data Fig. 3c,d), suggesting no compartmental switching as thymocytes matured.

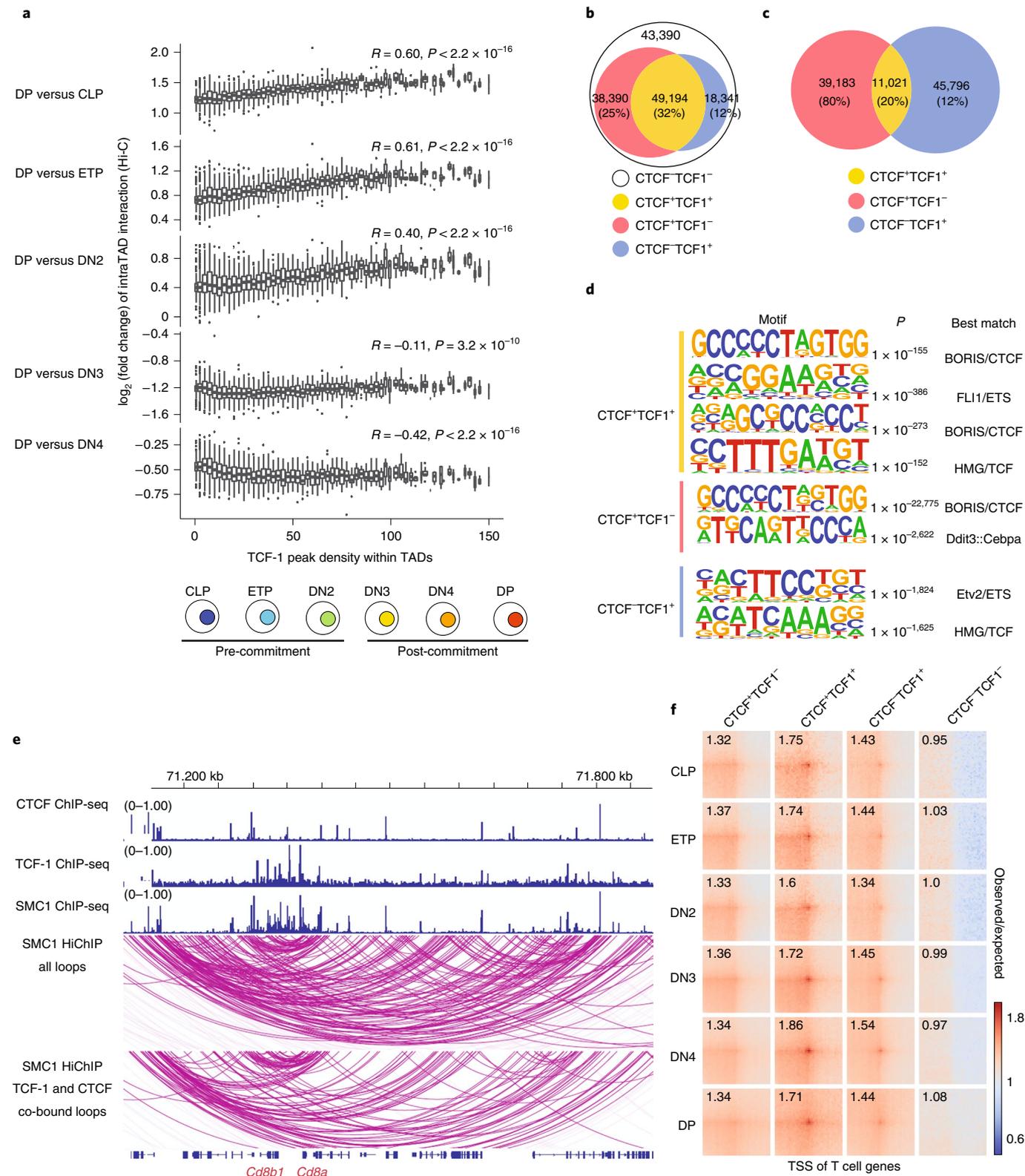
Next, we calculated the average occurrence of peaks for each binding group within 500 kb of dynamic boundaries. While CTCF-only sites were enriched at most boundaries (Extended Data Fig. 3e–h), TCF-1+CTCF as well as TCF-1-only sites were enriched at lost boundaries in groups E and F (Extended Data Fig. 3f–i). Analysis of histone acetylation and chromatin accessibility<sup>29</sup> indicated that boundaries lost at DN3s and DN4s (group E and F) were highly associated with active enhancer marks, as well as transcriptionally active and accessible chromatin states across T cell development (Extended Data Fig. 3h,i), suggesting that TCF-1 binding events, especially TCF-1+CTCF regions, were abundant in transcriptionally active and open chromatin domains.

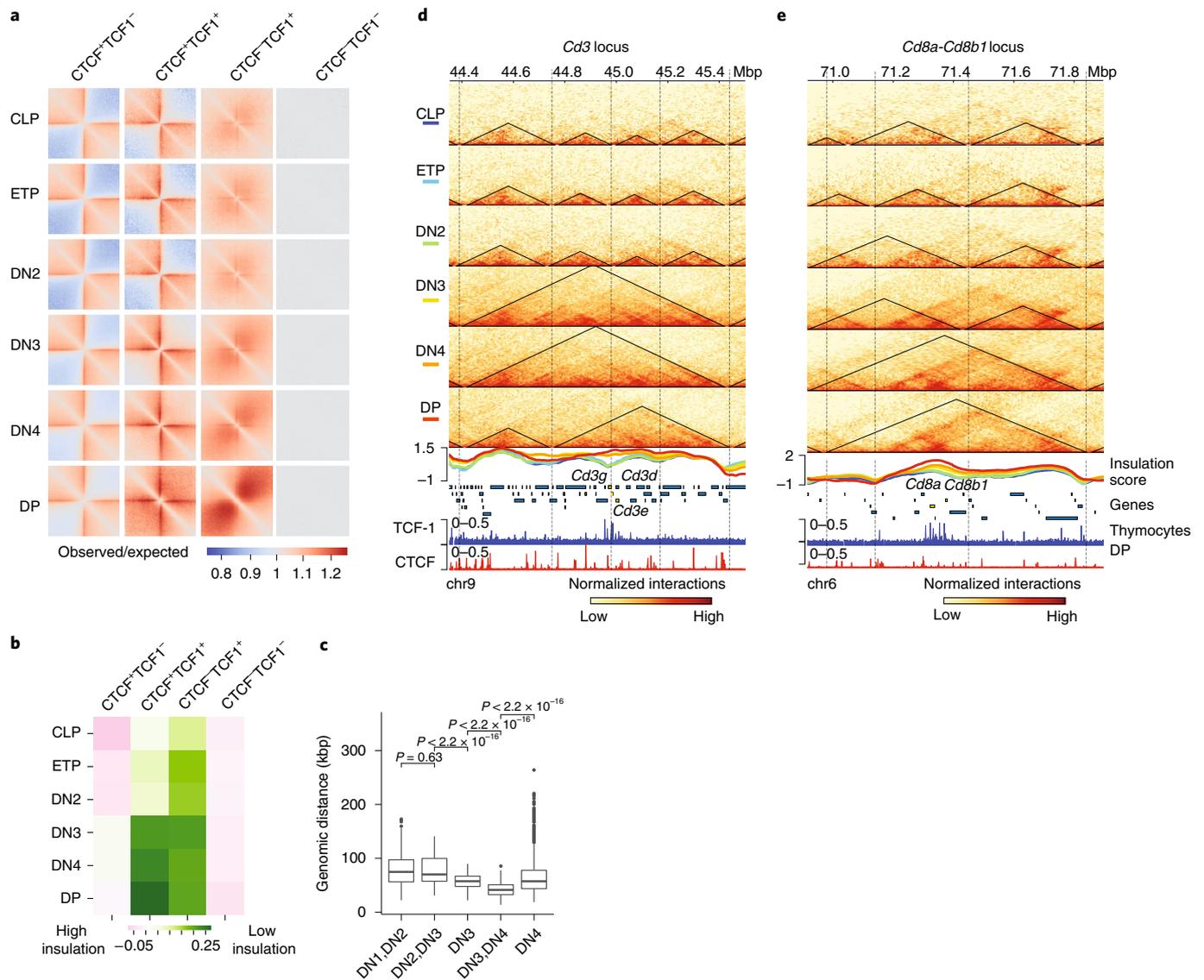
We found multiple boundaries in the *Cd3d-g* and *Cd8a-b* loci in CLPs and ETPs that weakened in DPs (Fig. 2d,e), suggesting that boundary weakening between two adjacent domains at loci encompassing T cell development genes corresponded to new interactions between previously insulated domains. Dense TCF-1 occupancy was accompanied by increased expression of the *Cd3d-g* and *Cd8a-b* genes in DN3s to DPs (Fig. 2d,e). A cluster of TCF-1 binding events at a boundary encompassing the *Cd3* locus corresponded to loss of insulation and gain in long-range interactions across the boundary at DN3s to DPs (Fig. 2d). Moreover, multiple TCF-1 binding events that occurred near the boundary of *Cd8a-b* locus, were also associated with loss of insulation and extensive long-range interactions (Fig. 2e). New boundaries were also detected at select loci such as a cluster of TCF-1 binding events at the *Rag1-Rag2* locus (Extended Data Fig. 3j), corresponding to increased expression of these genes in mature T cells. Thus, TCF-1+CTCF co-binding was associated with dynamic domain boundaries encompassing T cell identity genes.

**Fig. 1 | TCF-1 is associated with intra-TAD interactions and cohesin loops.** **a**, Box plots showing the association between TCF-1 binding density and  $\log_2$ (fold change) difference in intra-TAD interactions between DP and CLP-to-DN4 (top) and schematic of T cell developmental stages (bottom). Long-range interactions at various T cell developmental stages were measured using Hi-C<sup>26</sup>. Intra-TAD interactions were calculated based on TADs detected in DP T cells ( $n=3,023$ ) and peak density were calculated with number of TCF-1 peaks per 1-Mbp region. TCF-1 peaks were defined from TCF-1 ChIP-seq in thymocytes<sup>25</sup>. Data are shown as box plots (center, median; box limits, upper (75th) and lower (25th) percentiles; whiskers, 1.5x interquartile range; points, outliers). *P* value was calculated with two-sided Spearman correlation. **b**, Venn diagram showing the total number of cohesin loops measured by SMC1 HiChIP and the proportion of them co-bound by CTCF and/or TCF-1 at least on one anchor. The four classes of peaks were defined using CTCF ChIP-seq in DPs<sup>29</sup> and publicly available TCF-1 ChIP-seq in thymocytes<sup>25</sup>. **c**, Venn diagram showing the number of TCF-1 and CTCF peaks and the overlap between them in DPs. **d**, Seq-logos demonstrating motif enrichment using random background based on HOMER analysis at TCF-1+CTCF, CTCF-only and TCF-1-only sites in DPs. *P* values were calculated using hypergeometric test. **e**, Genome browser view showing the association of CTCF and TCF-1 peaks with cohesin loops at the *Cd8a* and *Cd8b1* locus. Loops co-bound by CTCF and TCF-1 at least on one anchor are shown in the bottom panel. **f**, Pileups of interactions between pairs of T cell developmental genes and CTCF-only, TCF-1+CTCF, TCF-1-only peaks and random regions within 100-kbp windows. T cell genes were defined from single-cell RNA-seq analysis of the thymus<sup>30</sup> ('T cell subset markers'; Supplementary Table 2 in ref. <sup>30</sup>). The number in each box indicates the strength of interaction at the center pairing transcription start site (TSS) of T cell genes and different sets of peaks.

**TCF-1+CTCF sites are evolutionarily conserved.** We next examined the conservation scores of DNA sequences centered on different classes of CTCF and TCF-1 binding events. We found a sharp increase in sequence conservation at genomic regions with CTCF-only or TCF-1-only peaks and an even broader and stronger degree of sequence conservation at TCF-1+CTCF sites

(Extended Data Fig. 4a). To further explore the insulation status of conserved binding sites in human T cells, we employed Hi-C data from T cell acute lymphoblastic leukemia (T-ALL) and ETP leukemia (ETP-ALL)<sup>30</sup>, which arise from malignant transformation of T cell precursors. Moreover, we generated Hi-C maps from mature lymph node CD8<sup>+</sup> T cells from one healthy organ donor (Extended





**Fig. 2 | TCF-1 + CTCF sites earmark weakening of insulated neighborhoods.** **a**, Heat map showing local pileup interactions using Hi-C data at CTCF-only, CTCF + TCF-1, TCF-1-only peaks and random regions as control. The horizontal line and the vertical line of each plot are the 500-kb region centered on different peak groups. **b**, Heat map showing average insulation score during T cell development at the CTCF-only, CTCF + TCF-1 co-bound, TCF-1-only peaks and random regions as control. **c**, Box plots showing distance of marker genes of T cell developmental stages to TCF-1 + CTCF peaks. Data are shown as box plots (center, median; box limits, upper (75th) and lower (25th) percentiles; whiskers, 1.5x interquartile range; points, outliers). The comparison between different stages was conducted with a two-sided Wilcoxon signed-rank test. The distance between these marker genes and TCF-1 + CTCF peaks significantly ( $P < 2.22 \times 10^{-16}$ ) changed except between DN1 and DN2. Selection of marker genes as described in Fig. 1g. **d, e**, Genome browser view showing the interactions at *Cd3d-e* (**d**) and *Cd8a, b1* (**e**) loci during T cell development, as well as TCF-1 and CTCF binding events in DPs. Triangles in the heat map represent TADs called with cooltools. Boundaries are marked as vertical lines. Insulation score in each T cell developmental stage is shown as a line plot and colored according to their developmental stage as indicated in the heat map legend. Heat maps were normalized with sequencing depth and ChIP-seq tracks were normalized with count per million.

Data Fig. 4b). We evaluated the extent of insulation and long-range 3D interactions at homologous CTCF and TCF-1 binding sites in the human genome and found loss of insulation at TCF-1 + CTCF sites in mature CD8<sup>+</sup> T cells compared to T-ALLs and ETP-ALLs (Extended Data Fig. 4b). We further analyzed the average local interactions at TCF-1 + CTCF sites in multiple mouse cell types (Extended Data Fig. 4c). While the CTCF sites seemed insulated in all examined cell types, TCF-1 sites showed increased interactions only in naive CD4<sup>+</sup> T cells and DPs (Extended Data Fig. 4c). Based on analysis of public Hi-C data from the 4D Nucleome project<sup>31</sup>, the TCF-1 + CTCF sites exhibited the same level of insulation as

CTCF sites in all non-T cell lines examined (Extended Data Fig. 4d). Thus, the highly conserved TCF-1 + CTCF sites were insulated in non-T cells and early T cell progenitors in both humans and mice.

**Overexpression of TCF-1 in fibroblasts creates de novo 3D interactions.** To ask whether TCF-1 expression was required for long-range interactions across co-bound TCF-1 + CTCF boundaries, we used fibroblasts because T cell lineage collaborators of TCF-1 are not expressed in this cell type. We overexpressed TCF-1 in the fibroblast cell line NIH 3T3 using lentiviral transduction with a doxycycline-inducible gene expression system (hereafter 3T3<sup>TCF-1</sup>).

We measured TCF-1 and CTCF binding events and transcriptional outputs before and 72 h after TCF-1 induction by doxycycline in biological replicates (Supplementary Fig. 1), using ChIP-seq and RNA-seq, respectively. Transcriptomic measurements indicated the increased expression of 273 genes and decreased expression of 194 genes in 3T3<sup>TCF-1</sup> compared to non-induced fibroblasts (Extended Data Fig. 5a and Supplementary Table 5). Overall, 1,872 CTCF binding events were predominantly stronger and 431 CTCF binding events were weaker in 3T3<sup>TCF-1</sup> fibroblasts (Extended Data Fig. 5b). Thus, TCF-1 expression can interfere with the transcriptional control of fibroblasts<sup>17</sup>.

Next, we measured cohesin-mediated 3D genome interactions using SMC1 HiChIP before and after TCF-1 induction. Only a small number of regions switched compartments between 3T3<sup>TCF-1</sup> and non-induced fibroblasts (Extended Data Fig. 5c). The overall compartmentalization, measured by the compartment strength (Extended Data Fig. 5d) or the extent of BB, AA and AB interactions (Extended Data Fig. 5e), was similar in 3T3<sup>TCF-1</sup> and non-induced fibroblasts, suggesting that TCF-1 expression did not alter the compartmentalization of the fibroblast genome 72 h after induction. Most TAD boundaries were conserved (Extended Data Fig. 5f,g); however, we observed a significant increase in intra-TAD interactions within invariant TADs in 3T3<sup>TCF-1</sup> compared to non-induced fibroblasts (Fig. 3a). Measurements of the connectivity within a TAD, referred to as 'domain score'<sup>32</sup>, indicated significantly larger domain scores in 3T3<sup>TCF-1</sup> compared to non-induced fibroblasts (Fig. 3b), suggesting that TCF-1 can promote intra-TAD interactions. The density of TCF-1 binding correlated significantly with changes in domain score after TCF-1 induction (Extended Data Fig. 5h), suggesting that dense TCF-1 binding corresponded to a substantial increase in intra-TAD interactions. TCF-1 + CTCF sites had the largest increase in 3D interactions in 3T3<sup>TCF-1</sup> (Fig. 3c), suggesting their cooperativity in mediating 3D interactions. The increase in interactions originating from each genomic region was positively correlated with the number of TCF-1 peaks in the neighborhood (Extended Data Fig. 5i,j). The greatest increase in 3D interactions occurred when TCF-1 occupied both anchors and had high density at neighboring regions (Extended Data Fig. 6k,l). Thus, as exemplified at the *Irf2bp2* locus (Fig. 3d), TCF-1 could increase intra-TAD interactions in fibroblasts at many loci.

### TCF-1 controls the 3D chromatin conformation of T cells.

T cells undergo a remarkable genome reorganization during the DN2-to-DN3 transition<sup>25</sup>. To explore whether TCF-1 was required for the 3D genome organization during T cell development, we genetically ablated TCF-1 expression using CRISPR/Cas9 (Extended Data Fig. 6a) in a DN3-like T cell line, also called *Scid.adh*<sup>33</sup>. In addition, we generated Vav-Cre*Tcf7*<sup>fl/fl</sup> mice, in which TCF-1 is conditionally ablated in all hematopoietic cells (Extended Data Fig. 6b),

as indicated by immunoblot experiments (Extended Data Fig. 6c) and flow cytometry in the thymus (Extended Data Fig. 6d). We next mapped the global chromatin conformation using ultra-deep Hi-C measurements in DN3 cell lines and primary DPs in wild-type and TCF-1 deficient T cells (hereafter *Tcf7*<sup>-/-</sup> DN3s and DPs). We detected overall modest changes of PC1 values between wild-type and *Tcf7*<sup>-/-</sup> cells (Extended Data Fig. 6e,f); however, we observed weakening of compartment strengths in *Tcf7*<sup>-/-</sup> DPs (Fig. 3e). There was also a significant decrease of interactions in the A compartment in both *Tcf7*<sup>-/-</sup> DN3s and DPs, a significant decrease in compartment strength in *Tcf7*<sup>-/-</sup> DPs and a significant increase of interactions between A and B compartments in *Tcf7*<sup>-/-</sup> DPs (Fig. 3f and Extended Data Fig. 6h) compared to their wild-type counterparts. These data suggest TCF-1 was required for the establishment and maintenance of long-range interactions at the compartment level in DN3s and DPs. Most TAD boundaries were conserved between wild-type and *Tcf7*<sup>-/-</sup> cells (Extended Data Fig. 6i,j), but there was a significant reduction in intra-TAD interactions and hence, reduced domain scores in *Tcf7*<sup>-/-</sup> cells compared to wild-type DN3s and DPs (Fig. 3g), positing that TCF-1 was required for the increase in de novo 3D interactions at these stages.

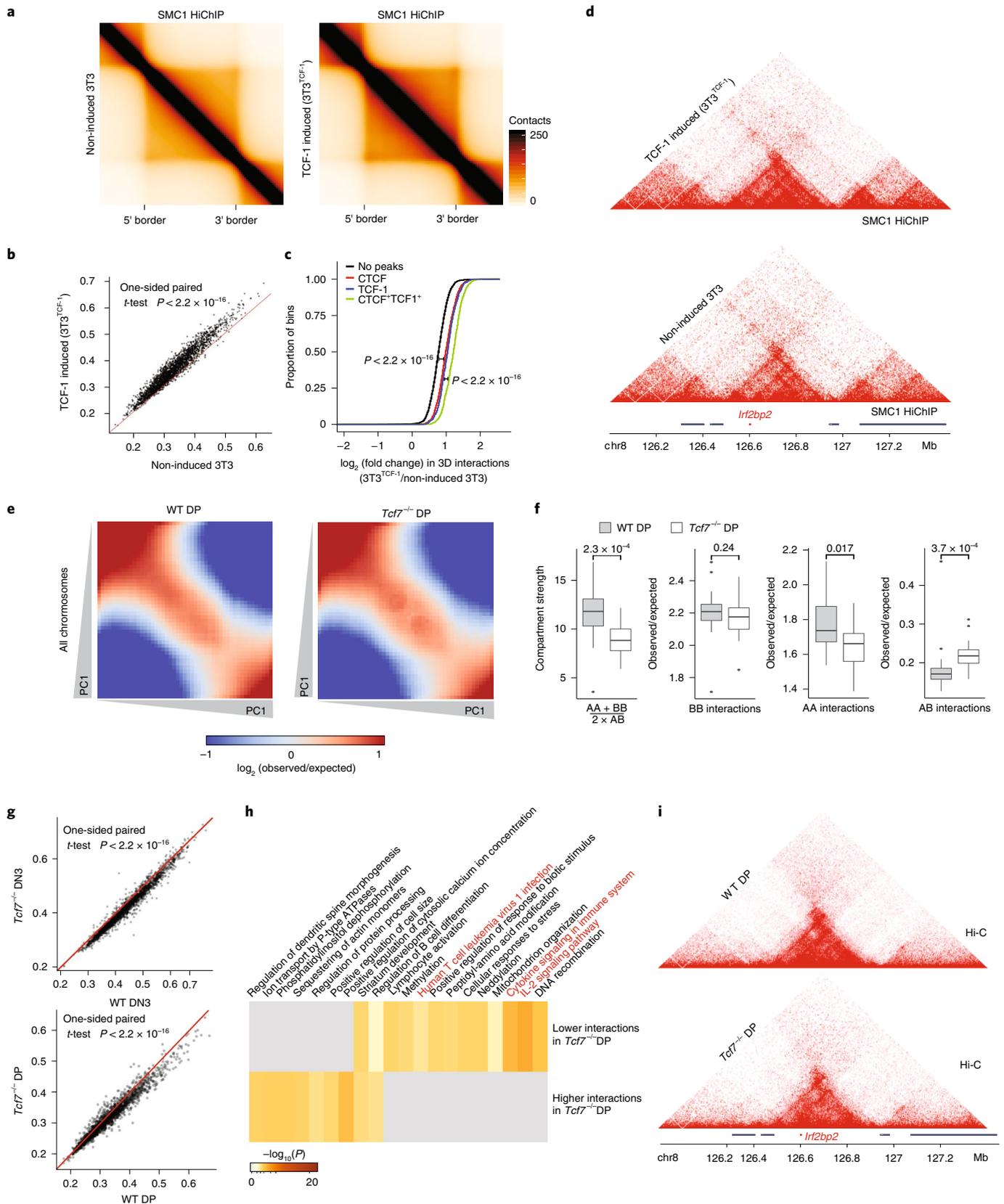
In *Tcf7*<sup>-/-</sup> DN3s (or *Tcf7*<sup>-/-</sup> DPs), 196 (or 525) genes were downregulated, whereas 374 (or 664) genes were upregulated (Supplementary Tables 6–9). To ascertain whether TCF-1-dependent changes in intra-TAD interactions related to changes in transcriptional outputs, we investigated the expression of genes located in the top 50 TADs, which had the greatest change in their domain scores in *Tcf7*<sup>-/-</sup> cells. Genes positioned in TADs whose domain scores increased in *Tcf7*<sup>-/-</sup> DN3s and DPs were upregulated in *Tcf7*<sup>-/-</sup> DN3s and DPs (Extended Data Fig. 6k,l and Supplementary Table 10), whereas genes whose domain scores decreased after TCF-1 deletion were downregulated in *Tcf7*<sup>-/-</sup> DN3s and DPs (Extended Data Fig. 6k,l and Supplementary Table 10). Gene Ontology analysis indicated that genes positioned in TADs whose domain scores decreased in *Tcf7*<sup>-/-</sup> DPs were enriched in ontologies including 'lymphocyte activation', 'cytokine signaling in immune system' and 'interleukin (IL)-2 signaling' (Fig. 3h,i and Supplementary Table 11), whereas genes whose domain scores decreased in *Tcf7*<sup>-/-</sup> DN3 were enriched in 'cellular response to interferon (IFN)- $\beta$ ' (Extended Data Fig. 6m and Supplementary Table 11), indicating a potential regulatory role of TCF-1 on IFN-induced genes by directly reorganizing 3D genome folding. These findings indicated the requirement of TCF-1 for intra-TAD interactions, which are necessary for the transcriptional control of genes with developmental roles in T cells.

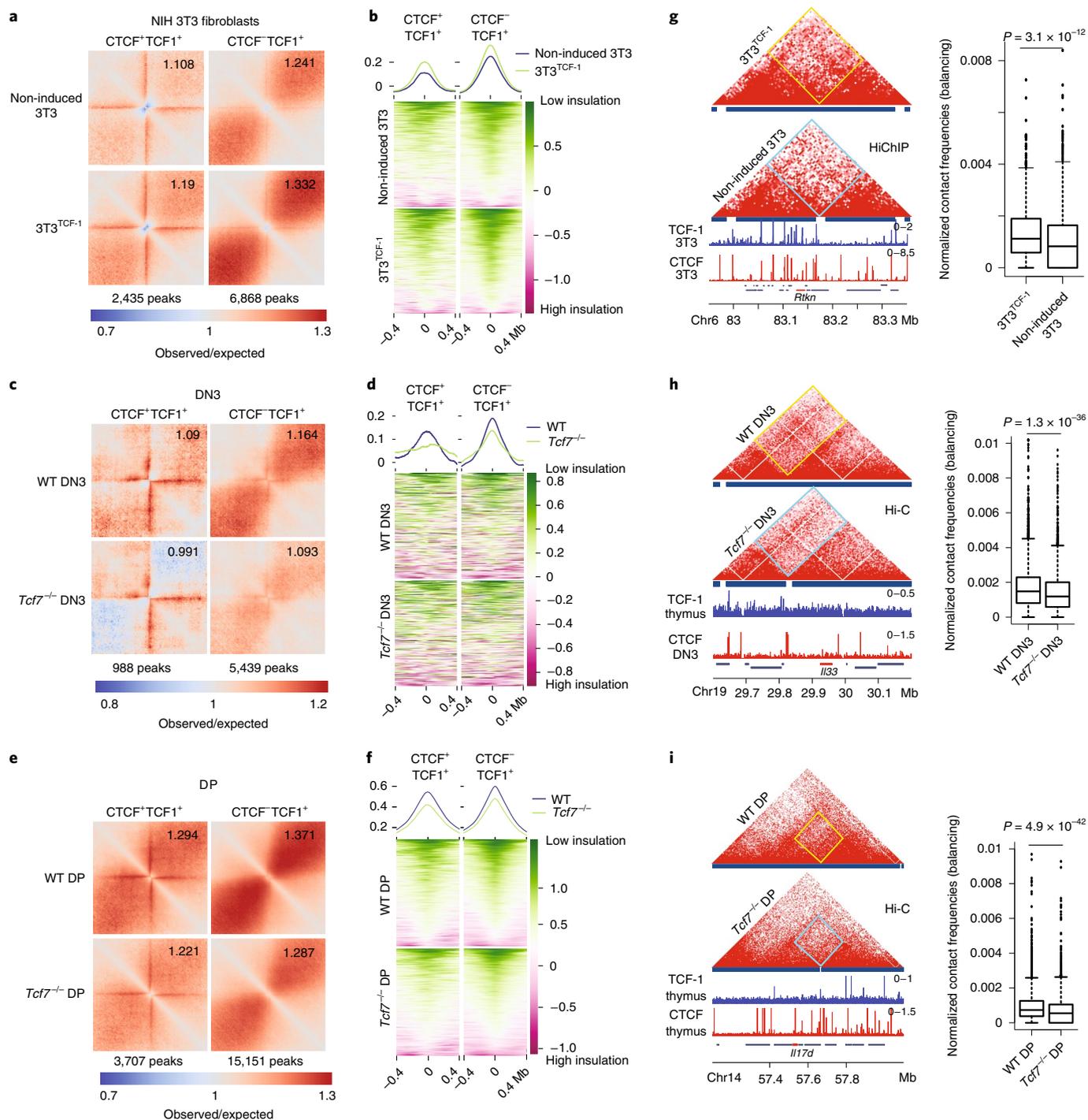
A dense cluster of TCF-1 binding events on IFN-induced genes corresponded to TCF-1 dependent SMC1 occupancy and 3D interactions (Extended Data Fig. 7a), suggesting a potential role of TCF-1 protein in recruiting the cohesin subunit. To further investigate

**Fig. 3 | Gain and loss of TCF-1 reshapes long-range interactions.** **a**, Heat maps showing the aggregate of interactions across invariant TADs using SMC1 HiChIP in non-induced 3T3 and 3T3<sup>TCF-1</sup>. TADs are rescaled and normalized using GENOVA. One HiChIP experiment was performed per experimental condition. **b**, Scatter-plot showing domain scores in non-induced 3T3 and 3T3<sup>TCF-1</sup>. One-sided paired Student's *t*-test assesses that domain scores were significantly ( $P < 2.21 \times 10^{-16}$ ) higher in 3T3<sup>TCF-1</sup> compared to non-induced 3T3. **c**, Cumulative distribution plot showing the increase of interactions in genomic bins with TCF-1 + CTCF or other groups. Comparisons between different groups using two-sided Kolmogorov-Smirnov test evaluates the significance of the difference. TCF-1 + CTCF sites gained significantly ( $P < 2.2 \times 10^{-16}$ ) more interactions than other groups. **d**, Contact matrix and genome browser view of the TAD encompassing the *Irf2bp2* locus in 3T3. **e**, Saddle plot showing the compartmentalization across all chromosomes in wild-type and *Tcf7*<sup>-/-</sup> DPs based on Hi-C measurements. One Hi-C experiment was performed per genotype. **f**, Data are shown as box plots (center, median; box limits, upper (75th) and lower (25th) percentiles; whiskers, 1.5x interquartile range; points, outliers) showing the compartment strength, BB interactions, AA interactions and AB interactions in wild-type and *Tcf7*<sup>-/-</sup> DPs ( $n = 20$  mouse chromosomes). The statistical test was performed using a two-sided Student's *t*-test. (NS, not significant; \* $P < 0.05$ ; \*\*\* $P < 0.001$ ). **g**, Scatter-plot showing domains scores in DN3s and DPs. The red line represents equal domain scores in two conditions. One-sided paired Student's *t*-test shows that domain score is significantly different in wild-type DN3s and DPs compared to *Tcf7*<sup>-/-</sup> cells ( $P < 2.21 \times 10^{-16}$ ). One Hi-C experiment was performed in the wild-type condition and two Hi-C experiments were performed in two distinct *Tcf7*<sup>-/-</sup> DN3 clones. **h**, Gene Ontology analysis of genes in the top 50 TADs that gain domain score and the top 50 TADs that lose domain score after TCF-1 deletion in DP T cells using metaScape. **i**, Contact matrix and genome browser view of the TAD encompassing *Irf2bp2* that lost intra-TAD interactions in *Tcf7*<sup>-/-</sup> DPs.

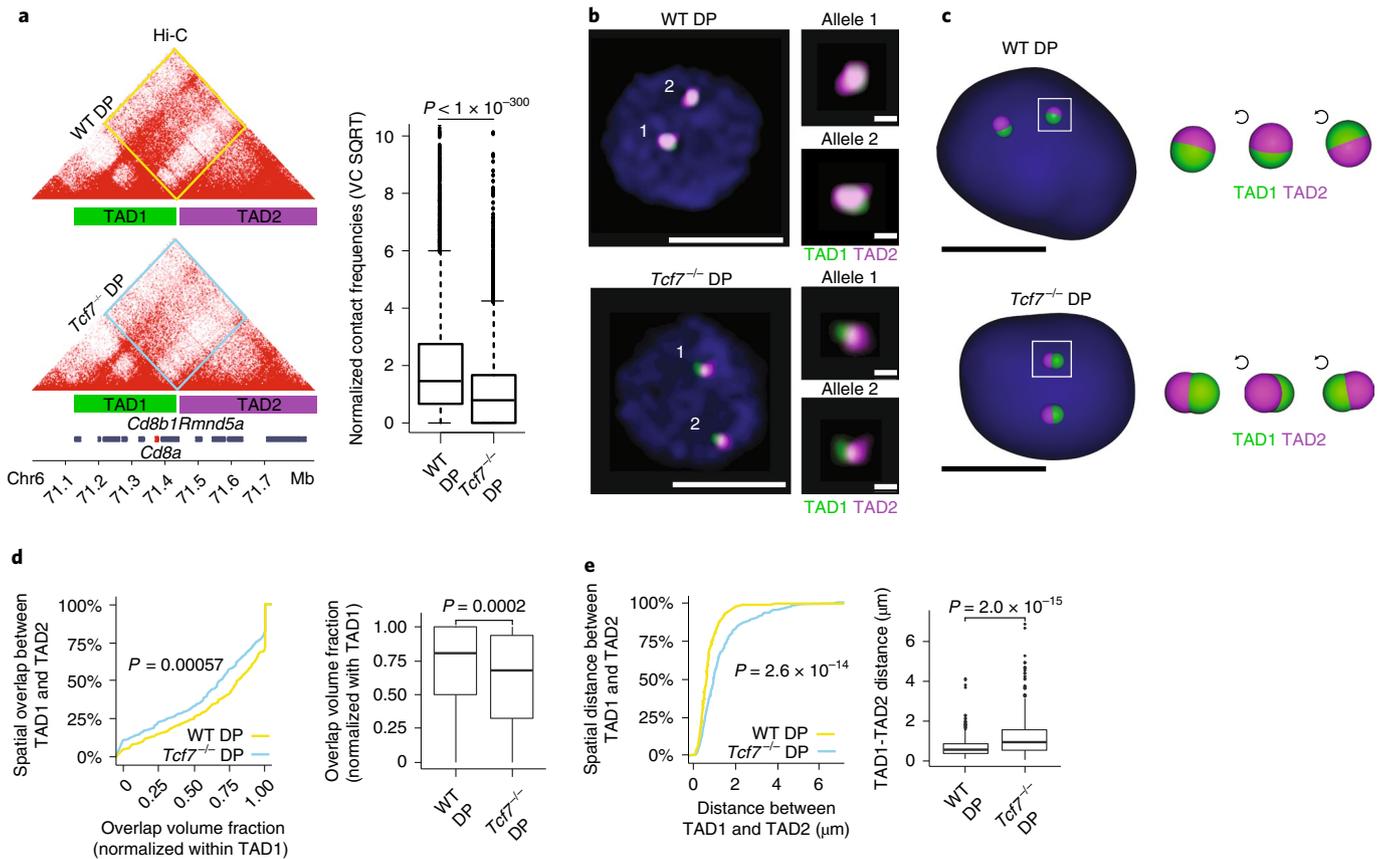
whether sequence variation at this TCF-1 binding cluster can alter 3D genome interactions and thus the expression of IFN-induced genes within this domain, we utilized the SMC1 HiChIP data in DPs in C57BL/6 and NOD<sup>27</sup>, which showed differences in the intra-TAD interactions (Extended Data Fig. 7b). A cluster of TCF-1

binding events coincided with stronger interactions in NOD compared to C57BL/6 DPs (Extended Data Fig. 7b). Consequently, most genes had higher expression in NOD compared to C57BL/6 DPs (Extended Data Fig. 7c). Long-range interactions at this domain were also TCF-1 dependent in fibroblasts and DPs (Extended Data





**Fig. 4 | TCF-1 is required to dismantle boundaries.** **a**, Local pileup plot of long-range interactions using SMC1 HiChIP in non-induced 3T3 and 3T3<sup>TCF-1</sup> at a subset of CTCF + TCF-1 and TCF-1 only peaks which lost insulation in 3T3<sup>TCF-1</sup>. The two classes of peaks were defined using TCF-1 and CTCF ChIP-seq in 3T3<sup>TCF-1</sup>. **b**, Summary plot and heat map of insulation score in non-induced 3T3 and 3T3<sup>TCF-1</sup>. **c**, Local pileup plot of long-range interactions using ultra-deep Hi-C in wild-type and *Tcf7*<sup>-/-</sup> DN3 at a subset of CTCF + TCF-1 and TCF-1 only peaks that gained insulation in *Tcf7*<sup>-/-</sup> DN3. The two classes of peaks were defined using TCF-1 and CTCF ChIP-seq in thymocytes and DPs, respectively. **d**, Summary plot and heat map showing insulation score in wild-type and *Tcf7*<sup>-/-</sup> DN3. **e**, Local pileup plot of long-range interactions using ultra-deep Hi-C in wild-type and *Tcf7*<sup>-/-</sup> DPs at a subset of CTCF + TCF-1 and TCF-1 only peaks that gained insulation in *Tcf7*<sup>-/-</sup> DPs. **f**, Summary plot and heat map showing insulation score in wild-type and *Tcf7*<sup>-/-</sup> DPs at a subset of CTCF + TCF-1 and TCF-1 only peaks that gained insulation after TCF-1 deletion in DPs. **g-i**, Contact matrix of HiChIP and Hi-C data at the *Rtkn*, *Il22* and *Il17d* loci in NIH 3T3 (**g**) and DN3 (**h**) and DPs (**i**) respectively. Inter-domain interactions are quantified by box plots ( $n=1,080$  (**g**), 3,404 (**h**) and 3,300 genomic interactions (**i**) for 3T3<sup>TCF-1</sup>, DN3 and DPs, respectively; center, median; box limits, upper (75th) and lower (25th) percentiles; whiskers, 1.5x interquartile range; points, outliers). Two-sided paired Wilcoxon rank-sum test was performed.



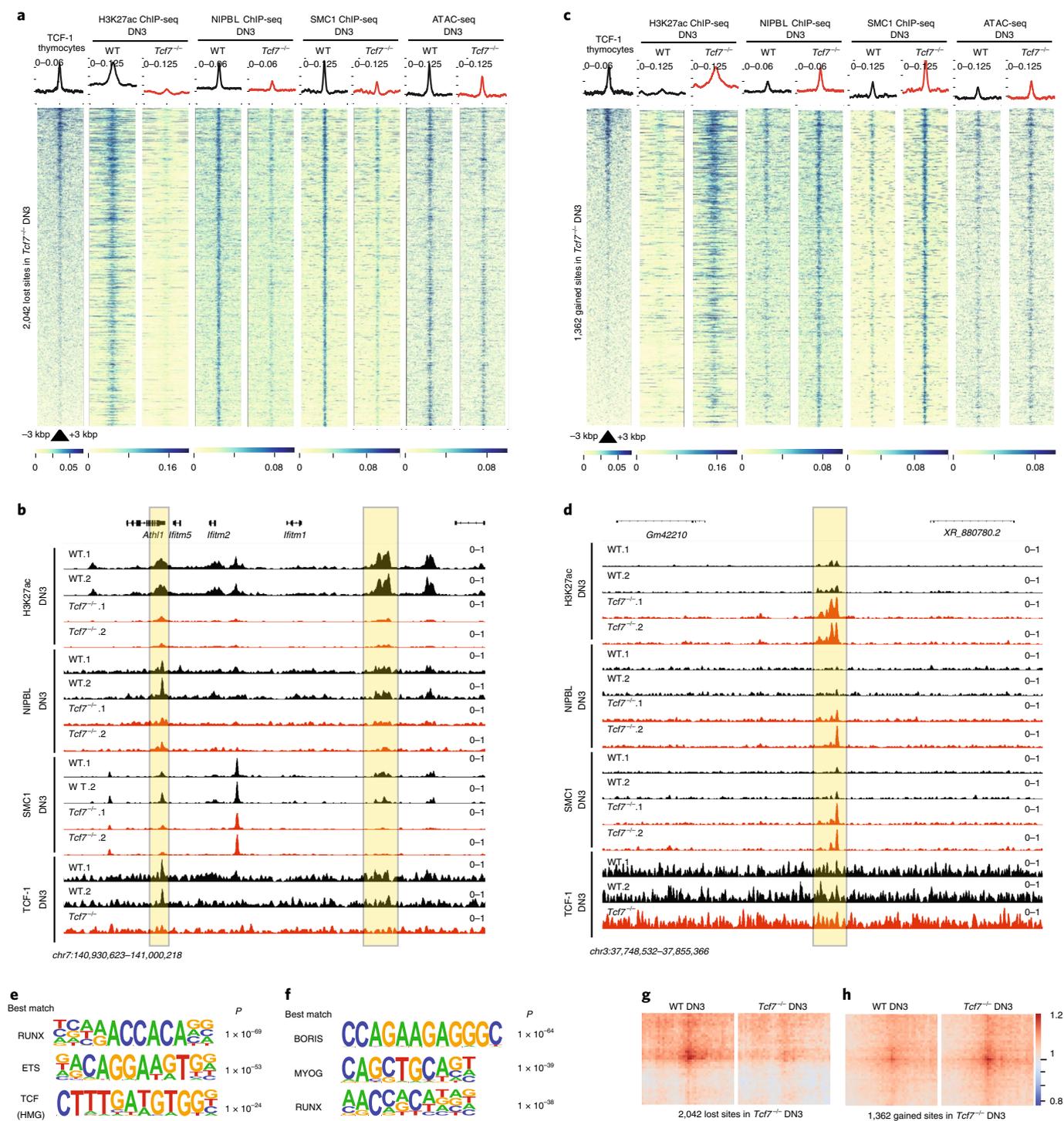
**Fig. 5 | 3D FISH reveals TCF-1 dependent TAD intermingling at the *CD8a* locus in DPs.** **a**, Contact matrix of Hi-C data at the *Cd8a-Cd8b1* locus which includes the location of Oligopaint probes for TAD1 and TAD2 depicted as green and magenta bars. Inter-domain interactions were quantified by box plot in wild-type and *Tcf7*<sup>-/-</sup> DPs ( $n = 7,221$  genomic interactions; box plot center, median; box limits, upper (75th) and lower (25th) percentiles; whiskers, 1.5 $\times$  interquartile range; points, outliers). A two-sided paired Wilcoxon rank-sum test was performed. TAD1 spans chr6:71163477-71432963, whereas TAD2 spans chr6:71440321-71825773 in mm10. **b**, A representative image of Oligopaint FISH probes in wild-type and *Tcf7*<sup>-/-</sup> DPs with magnification of each allele. Scale bars, 5  $\mu\text{m}$  (full cell), 1  $\mu\text{m}$  (magnified allele). Nuclei are stained with 4,6-diamidino-2-phenylindole (DAPI) (blue). **c**, 3D rendering of TAD1 and TAD2 in wild-type and *Tcf7*<sup>-/-</sup> single DP T cells. Zoomed view of one allele per cell is shown in three rotations of 90°. Scale bar, 5  $\mu\text{m}$ . **d,e**, Cumulative distribution plot (left) and corresponding box plot (right) of overlap volume (d) and distance (e) between TAD1 and TAD2 across 312 individual wild-type DPs and 367 individual *Tcf7*<sup>-/-</sup> DPs. Kolmogorov-Smirnov test  $P$  values and two-sided Wilcoxon rank-sum test  $P$  values are shown for cumulative distribution plots and corresponding box plots, respectively. Cells were pooled from two to three mice per genotype. Overlap volume was defined using a 3D segmentation strategy<sup>48</sup> across a minimum of 300 alleles per condition. The overlap volume of TAD1 and TAD2 per allele was normalized to the volume of TAD1 (d). Distance was measured between the centroids of each domain across individual alleles (e). Data are shown as box plots (center, median; box limits, upper (75th) and lower (25th) percentiles; whiskers, 1.5 $\times$  interquartile range; points, outliers).

Fig. 7d,e). Thus, TCF-1 was required to maintain long-range interactions at various length scales.

**TCF-1 is required to diminish insulation.** To investigate the extent to which TCF-1 acted on insulated neighborhoods, we examined interactions at TCF-1 + CTCF sites in 3T3<sup>TCF-1</sup> fibroblasts in addition to *Tcf7*<sup>-/-</sup> DN3s and DPs. We first quantified the interactions at a 500-kbp region centered around different classes of TCF-1 and CTCF binding events in 3T3<sup>TCF-1</sup> fibroblasts. Among 12,143 TCF-1 + CTCF sites, interactions between upstream and downstream domains centered at 2,435 co-bound sites increased in 3T3<sup>TCF-1</sup> fibroblasts compared to non-induced fibroblasts (Fig. 4a and Supplementary Table 12). Among 27,353 TCF-1 peaks, the interaction and insulation changed at 6,868 sites (Fig. 4a,b).

To test whether TCF-1 was required to reshape the insulated neighborhoods, we quantified 3D interactions in wild-type, *Tcf7*<sup>-/-</sup> DN3s and DPs using our ultra-deep Hi-C measurements. We mapped TCF-1 binding events in DN3s using TCF-1 ChIP-seq and compared TCF-1 occupancy in DN3s with the genome-wide profile

based on the public TCF-1 ChIP-seq data in thymocytes<sup>24</sup> (Extended Data Fig. 8a). Based on the number of peaks, the public TCF-1 ChIP-seq data in thymocytes was of higher quality (Extended Data Fig. 8a), so we employed them in analyses of DN3s and DPs. Loss of TCF-1 in DN3s led to reduced interactions at 988 TCF-1 + CTCF sites (Fig. 4c,d and Supplementary Table 13) and 5,439 TCF-1-only sites (Fig. 4c,d), with a corresponding increase in insulation. These effects were reproduced in another *Tcf7*<sup>-/-</sup> DN3 clone (Extended Data Fig. 8b). The effect of TCF-1 deletion in DPs on long-range interactions centered at TCF-1 + CTCF sites was even more prominent than that in DN3s and led to reduced interactions at 3,707 TCF-1 and CTCF co-bound peaks, whereas 15,151 TCF-1-only sites showed loss of 3D interactions (Fig. 4e,f and Supplementary Table 14). Hi-C data from wild-type and *Bcl11b*-deficient naive CD4<sup>+</sup> T cells<sup>25</sup> indicated no effect of *Bcl11b* on insulation (Extended Data Fig. 8c). As representative examples, the *Rtkn* (Fig. 4g), *Il33* (Fig. 4h) and *Il17d* (Fig. 4i) loci showed statistically significant changes in chromatin interactions between 3T3<sup>TCF-1</sup> and non-induced 3T3 in addition to wild-type and *Tcf7*<sup>-/-</sup> DN3s and



**Fig. 6 | TCF-1 promotes NIPBL recruitment to active enhancers to reconstruct genome organization. a,c.** Heat map depicting genomic regions where TCF-1 led to a loss (**a**) or a gain (**c**) on the binding events of SMC1 and NIPBL in addition to deposition of H3K27ac in DN3s. A binding atlas of cohesin occupancy was created by combining SMC1 and NIPBL peaks across various conditions. DESeq2 was used to determine a TCF-1-dependent set of peaks between wild-type and *Tcf7*<sup>-/-</sup> cells. Heat maps show  $\pm 3$  kb around the peak center. Publicly available TCF-1 ChIP-seq from thymocytes was used<sup>25</sup>. H3K27Ac, NIPBL and SMC1 ChIP-seq and ATAC-seq experiments were generated in DN3s. **b.** Genome browser view at a cluster of IFN-induced transmembrane genes. Yellow highlight shows binding of TCF-1 at genomic loci where H3K27ac, SMC1 and NIPBL binding events were concordantly reduced in *Tcf7*<sup>-/-</sup> DN3s. **d.** Genome browser view of a representative example locus, yellow highlight shows genomic loci where H3K27ac, SMC1 and NIPBL binding events were concordantly increased in *Tcf7*<sup>-/-</sup> DN3s. **e,f.** Seq-logos demonstrating top motif enrichment based on HOMER analysis of genomic regions in which occupancy of cohesin was lost (**e**) or gained (**f**) in *Tcf7*<sup>-/-</sup> DN3s compared to random background regions. *P* values are calculated using hypergeometric test. **g,h.** Local pileup of long-range interactions anchored at genomic regions described in (**a**) and (**c**) with lost (**g**) and gained (**h**) occupancy. Hi-C data in DN3s were used.

DPs. Thus, TCF-1 binding could change local interactions between insulated domains.

**Oligopaint 3D FISH corroborates genomic measurements.** A robust identification of TADs is still lacking due to the relatively low concordance of TAD definitions<sup>34,35</sup>. We used alternative methodologies to quantify interaction frequencies across TCF-1 dependent boundaries. We tested how chromatin interactions measured by Hi-C related to spatial localization of adjacent TADs using high-resolution Oligopaint DNA fluorescence in situ hybridization (FISH) with confocal imaging in 3D<sup>36,37</sup>. Because we detected TCF-1-dependent chromatin interactions across domains at the *Cd8a-Cd8b1* locus in fibroblasts, DN3s and DPs (Fig. 5a and Extended Data Fig. 9a,b) we tiled Oligopaint probes along two domains of this locus, referred to as TAD1 (~284 kbp) and TAD2 (~437 kbp) (Fig. 5a), which were insulated in ETP-DN2 and combined into one domain after DN3. Because the largest extent of interactions across boundaries was detected at DPs, we performed Oligopaint FISH experiments in wild-type and *Tcf7*<sup>-/-</sup> DPs. To assess the spatial proximity of TADs, we measured the distance between centroids of each domain across individual alleles and used a 3D segmentation strategy<sup>38</sup> to trace the edges of each domain's signal, generating a distribution of domain volumes across ~300 alleles per condition (Fig. 5b). To report the fraction of overlap between two domains across single alleles, the overlap volume of TAD1 and TAD2 per allele was normalized to the volume of TAD1. If chromatin interactions across boundaries at the *Cd8a-Cd8b1* locus in DPs were TCF-1-dependent, we expect a reduction in overlap between TAD1 and TAD2 volumes and an increase in distance between TAD1 and TAD2 centroids in *Tcf7*<sup>-/-</sup> compared to wild-type DPs. Representative examples of alleles at the *Cd8* locus in wild-type and *Tcf7*<sup>-/-</sup> DPs and 3D rendering of TAD1 and TAD2 at the single-cell level corroborated the TCF-1-dependent spatial localization of two adjacent TADs (Fig. 5b,c). This finding was also confirmed for hundreds of DPs across biological replicates (Fig. 5d,e and Extended Data Fig. 9c,d). As such, high-resolution imaging quantified the TCF-1-dependent chromatin interactions across boundaries, independent of sequencing techniques at the single-allele level.

**TCF-1 recruits NIPBL to reconstruct 3D genome.** TADs are formed by cohesin-mediated extrusion of DNA<sup>1</sup>. Subunits of the cohesin complex are loaded onto chromatin by NIPBL and MAU2 (refs. 39–41) and once loaded, they extrude DNA until convergent CTCF sites are recognized<sup>42–45</sup>. Although TCF-1 overexpression in fibroblasts or TCF-1 deletion in DN3s changed CTCF occupancy (Extended Data Figs. 5b and 10a), CTCF binding at TCF-1-dependent boundaries remained unchanged (Extended Data Fig. 7). We mapped the genome-wide binding patterns of NIPBL and SMC1 and measured the chromatin signature of active enhancers using H3K27ac ChIP-seq in wild-type and *Tcf7*<sup>-/-</sup> DN3s in biological replicates (Supplementary Figs. 2–3). We created a binding atlas of cohesin occupancy in DN3s by combining SMC1 and NIPBL peaks across various conditions and evaluated genomic regions with gain or loss of SMC1 and NIPBL, in addition to deposition of H3K27ac after TCF-1 deletion in DN3s. H3K27ac was lost in *Tcf7*<sup>-/-</sup> DN3s at more than 2,000 genomic regions, concomitant with an evident diminution of NIPBL and SMC1 binding at these sites (Fig. 6a and Supplementary Table 15). Notably, the overt chromatin accessibility at these regions, as measured by ATAC-seq, did not change commensurately (Fig. 6a). As a representative example, H3K27ac, SMC1 and NIPBL binding events were concordantly reduced in *Tcf7*<sup>-/-</sup> DN3s at a cluster of IFN-induced transmembrane genes (Fig. 6b). Gain of H3K27ac and co-occupancy of NIPBL and SMC1 occurred at more than 1,300 genomic regions in *Tcf7*<sup>-/-</sup> DN3s (Fig. 6c,d and Supplementary Table 16). ETS, RUNX and TCF recognition motifs were enriched at sites where cohesin complex

occupancy was lost in *Tcf7*<sup>-/-</sup> DN3s (Fig. 6e), whereas CTCF-like recognition motifs were enriched at genomic regions that gained cohesin occupancy in *Tcf7*<sup>-/-</sup> DN3s (Fig. 6f) compared to wild-type DN3s. SMC1 ChIP-seq indicated a similar motif enrichment lost in *Tcf7*<sup>-/-</sup> DN3s (Extended Data Fig. 10c and Supplementary Table 8). The long-range interactions anchored at TCF-1-dependent SMC1 + NIPBL co-occupied regions that carried the H3K27ac signature were also TCF-1-dependent (Fig. 6g,h). Moreover, the TCF-1-dependent SMC1 + NIPBL co-bound enhancers were enriched near weakened boundaries bound by TCF-1 + CTCF (Extended Data Fig. 10d). Together, TCF-1 established the active enhancer repertoire and sequestered the cohesin machinery to these enhancers, a process which can create de novo long-range interactions and cause weakening of the previously insulated neighborhoods.

## Discussion

Here, we showed that co-binding of TCF-1 and CTCF earmarked TAD boundaries that were reorganized during T cell development. The disruption of TAD boundaries, such as domains encompassing the *Cd8* and *Cd3* clusters, was accompanied by interactions of regulatory elements previously located in different insulated neighborhoods. Key regulatory elements and their target genes were spatially distanced on different TADs during early developmental stages when these target genes were transcriptionally silent. Our data indicated that TCF-1 endowed competence to the T cell developmental program by minimizing the spatial distance between regulatory elements required for the proper expression of T cell genes, even when located within insulated neighborhoods in progenitors.

The ability of TCF-1 to recruit NIPBL and the cohesin complex at active enhancers implied a possible mechanism through which TCF-1 could diminish insulation at TAD boundaries. Degrading the histone acetyltransferases P300 leads to loss of NIPBL loading, suggesting that histone acetylation may represent a signal for cohesin loading<sup>46</sup>. We speculate that TCF-1-mediated histone acetylation can lead to accumulation of NIPBL and recruitment of the cohesin complex, which in turn would lead to chromatin interactions across TCF-1-bound loci. Considering that TCF-1 expression is induced at ETP, it is puzzling why it takes so long for the boundary dynamics to occur after DN3s. We speculate that the cohesin complex-mediated genome reorganization during T cell development might be rate limiting. This hypothesis is supported by the increased expression of NIPBL at DPs (ImmGen data), but further experiments are required. Whether the HMG domain of TCF-1 bends DNA to assist the cohesin complex during genome reorganization remains to be shown.

Although our work established an effect of TCF-1 on genome folding, it remains unclear to which extent the TCF-1-dependent global genome organization relates to transcriptional control. The emerging view is that transcription factors can alter genome organization before mediating transcriptional control. Pre-existing loops and the requirement of stage-specific histone acetylation for gene expression has been reported<sup>47,48</sup>. It is enticing to speculate that TCF-1 sets up the thymocytes' 3D genome landscape for transcriptional control once T cells specialize and expand in the periphery. Additional work is needed to elucidate whether CTCF and TCF-1 co-binding events enable spatial contacts between regulatory elements of genes that are essential in stem-like or memory CD8<sup>+</sup> T cells. Despite these new questions, our study describes an LDTF with the ability to eliminate TAD-enforced spatial distance between regulatory elements and their target genes in progenitors. We speculate that other LDTFs might be proficient to enable interactions between regulatory elements and target genes in other developmental pathways.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of

author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41590-022-01232-z>.

Received: 29 July 2021; Accepted: 5 May 2022;

Published online: 20 June 2022

## References

- Beagan, J. A. & Phillips-Cremins, J. E. On the existence and functionality of topologically associating domains. *Nat. Genet.* **52**, 8–16 (2020).
- Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
- Downen, J. M. et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387 (2014).
- Hnisz, D., Shrivivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A phase separation model for transcriptional control. *Cell* **169**, 13–23 (2017).
- Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
- Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
- Hansen, A. S. CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for a multi-step mechanism. *Nucleus* **11**, 132–148 (2020).
- Stadhouders, R., Filion, G. J. & Graf, T. Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* **569**, 345–354 (2019).
- Stadhouders, R. et al. Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nat. Genet.* **50**, 238–249 (2018).
- Weintraub, A. S. et al. YY1 is a structural regulator of enhancer-promoter loops. *Cell* **171**, 1573–1588 (2017).
- Di Giammartino, D. C. et al. KLF4 is involved in the organization and regulation of pluripotency-associated three-dimensional enhancer networks. *Nat. Cell Biol.* **21**, 1179–1190 (2019).
- Johanson, T. M. et al. Transcription-factor-mediated supervision of global genome architecture maintains B cell identity. *Nat. Immunol.* **19**, 1257–1264 (2018).
- Singh, H., Khan, A. A. & Dinner, A. R. Gene regulatory networks in the immune system. *Trends Immunol.* **35**, 211–218 (2014).
- Allman, D. et al. Thymopoiesis independent of common lymphoid progenitors. *Nat. Immunol.* **4**, 168–174 (2003).
- Li, L., Leid, M. & Rothenberg, E. V. An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. *Science* **329**, 89–93 (2010).
- Kueh, H. Y. et al. Asynchronous combinatorial action of four regulatory factors activates Bcl11b for T cell commitment. *Nat. Immunol.* **17**, 956–965 (2016).
- Johnson, J. L. et al. Lineage-determining transcription factor TCF-1 initiates the epigenetic identity of T cells. *Immunity* **48**, 243–257 (2018).
- Emmanuel, A. O. et al. TCF-1 and HEB cooperate to establish the epigenetic and transcription profiles of CD4<sup>+</sup>CD8<sup>+</sup> thymocytes. *Nat. Immunol.* **19**, 1366–1378 (2018).
- Giese, K., Cox, J. & Grosschedl, R. The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell* **69**, 185–195 (1992).
- Love, J. J. et al. Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* **376**, 791–795 (1995).
- Shan, Q. et al. Tcf1 and Lef1 provide constant supervision to mature CD8<sup>+</sup> T cell identity and function by organizing genomic architecture. *Nat. Commun.* **12**, 5863 (2021).
- Zhou, Y. et al. EBF1 nuclear repositioning instructs chromatin refolding to promote therapy resistance in T leukemic cells. *Mol. Cell* **82**, 1003–1020 e1015 (2022).
- Antoszewski, M. et al. Tcf1 is essential for initiation of oncogenic Notch1-driven chromatin topology in T-ALL. *Blood* **139**, 2483–2498 (2022).
- Dose, M. et al.  $\beta$ -Catenin induces T-cell transformation by promoting genomic instability. *Proc. Natl Acad. Sci. USA* **111**, 391–396 (2014).
- Hu, G. et al. Transformation of accessible chromatin and 3D nucleome underlies lineage commitment of early T cells. *Immunity* **48**, 227–242 e228 (2018).
- Crane, E. et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).
- Fasolino, M. et al. Genetic variation in type 1 diabetes reconfigures the 3D chromatin organization of T cells and alters gene expression. *Immunity* **52**, 257–274 (2020).
- Kernfeld, E. M. et al. A single-cell transcriptomic atlas of thymus organogenesis resolves cell types and developmental maturation. *Immunity* **48**, 1258–1270 (2018).
- Zhang, J. A., Mortazavi, A., Williams, B. A., Wold, B. J. & Rothenberg, E. V. Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell* **149**, 467–482 (2012).
- Kloetgen, A. et al. Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. *Nat. Genet.* **52**, 388–400 (2020).
- Dekker, J. et al. The 4D nucleome project. *Nature* **549**, 219–226 (2017).
- Krijger, P. H. et al. Cell-of-origin-specific 3D genome structure acquired during somatic cell reprogramming. *Cell Stem Cell* **18**, 597–610 (2016).
- Dionne, C. J. et al. Subversion of T lineage commitment by PU.1 in a clonal cell line system. *Dev. Biol.* **280**, 448–466 (2005).
- Forcato, M. et al. Comparison of computational methods for Hi-C data analysis. *Nat. Methods* **14**, 679–685 (2017).
- Eres, I. E. & Gilad, Y. A TAD skeptic: is 3D genome topology conserved? *Trends Genet.* **37**, 216–223 (2021).
- Beliveau, B. J. et al. Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nat. Commun.* **6**, 7147 (2015).
- Beliveau, B. J. et al. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl Acad. Sci. USA* **109**, 21301–21306 (2012).
- Ollion, J., Cochenec, J., Loll, F., Escude, C. & Boudier, T. TANGO: a generic tool for high-throughput 3D image analysis for studying nuclear organization. *Bioinformatics* **29**, 1840–1841 (2013).
- Ciosk, R. et al. Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. *Mol. Cell* **5**, 243–254 (2000).
- Murayama, Y. & Uhlmann, F. Biochemical reconstitution of topological DNA binding by the cohesin ring. *Nature* **505**, 367–371 (2014).
- Petela, N. J. et al. Scc2 is a potent activator of cohesin's ATPase that promotes loading by binding Scc1 without Pds5. *Mol. Cell* **70**, 1134–1148 e1137 (2018).
- Vian, L. et al. The energetics and physiological impact of cohesin extrusion. *Cell* **173**, 1165–1178 (2018).
- Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
- Kueng, S. et al. Wapl controls the dynamic association of cohesin with chromatin. *Cell* **127**, 955–967 (2006).
- Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
- Zhu, Y., Denholtz, M., Lu, H. & Murre, C. Calcium signaling instructs NIPBL recruitment at active enhancers and promoters via distinct mechanisms to reconstruct genome compartmentalization. *Genes Dev.* **35**, 65–81 (2021).
- Beagan, J. A. et al. Three-dimensional genome restructuring across timescales of activity-induced neuronal gene expression. *Nat. Neurosci.* **23**, 707–717 (2020).
- Petrovic, J. et al. Oncogenic notch promotes long-range regulatory interactions within hyperconnected 3D Cliques. *Mol. Cell* **73**, 1174–1190 (2019).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2022

## Methods

This table shows the source and identifiers for reagents and resources used in the current study.

Reagent or resource	Source	Identifier
<b>Antibodies</b>		
Anti-CTCF	Millipore	Cat. no. 07-729; RRID:AB_441965
Anti-SMC1	Bethyl	Cat. no. A300-055A; RRID:AB_2192467
Normal rabbit IgG	CST	Cat. no. 2729; RRID:AB_1031062
Anti-TCF-1	CST	Cat. no. 2206S; RRID:AB_2199300
Biotin anti-mouse CD8a (53-6.7)	BioLegend	Cat. no. 100703, RRID:AB_312742
Anti-NIPBL	Bethyl	A301-779A; RRID:AB_1211232
Histone H3 (acetyl K27) antibody	Abcam	ab4729; RRID:AB_2118291
Anti-Histone H3 Antibody, Unconjugated	CST	Cat. no. 9715, RRID:AB_331563
PE anti-mouse CD4 (RM4-4)	BioLegend	Cat. no. 116005; RRID:AB_313690
APC anti-mouse CD8 (53-6.7)	BioLegend	Cat. no. 100711; RRID:AB_312750
PeCy7 anti-mouse CD25 (PC61)	BioLegend	Cat. no. 102015; RRID:AB_312864
Brilliant Violet 785 anti-mouse/human CD44 (IM7)	BioLegend	Cat. no. 103041; RRID:AB_11218802
APC streptavidin	BioLegend	Cat. no. 405207
Biotin anti-mouse Ly-6G/Ly-6C (Gr-1) (RB6-8C5)	BioLegend	Cat. no. 108403; RRID:AB_313368
Biotin anti-mouse NK1.1 (PK136)	BioLegend	Cat. no. 108703; RRID:AB_313390
Biotin anti-mouse CD11b (M1/70)	BioLegend	Cat. no. 101203, RRID:AB_312786
Biotin anti-mouse Ter-119 (TER-119)	BioLegend	Cat. no. 116203, RRID:AB_313704
Biotin anti-mouse B220 (RA3-6B2)	BioLegend	Cat. no. 103203, RRID:AB_312988
Anti-rabbit IgG, HRP-linked antibody	CST	Cat. no. 7074, RRID:AB_2099233
<b>Bacterial and virus strains</b>		
One-shot Stbl3 chemically competent <i>E. coli</i>	Thermo Fisher	Cat. no. C737303
<b>Chemicals, peptides and recombinant proteins</b>		
Dulbecco's modified eagle medium	Thermo Fisher	Cat. no. 11965084
FBS (HyClone)	Cytiva	Cat. no. SH30910.03
Bovine serum, heat inactivated	Thermo Fisher	Cat. no. 26170035
RPMI 1640 medium	Invitrogen	Cat. no. 11875085
L-Glutamine 200 mM	Lonza	Cat. no. 17-605E
MEM Non-Essential Amino Acids Solution (100x)	Invitrogen	Cat. no. 11140050
Sodium pyruvate (100 mM)	Invitrogen	Cat. no. 11360070
Penicillin-streptomycin	Gibco	Cat. no. 15140122
Gateway Clonase II LR	Thermo Fisher	Cat. no. 11791020
Gateway Clonase II BP	Thermo Fisher	Cat. no. 11789020
Polybrene	Sigma	Cat. no. TR-1003-G
Lipofectamine 3000	Thermo Fisher	Cat. no. L3000008
Fugene HD	Promega	Cat. no. E2311
Chloroquine	Sigma	Cat. no. C6628-25G

Continued

Reagent or resource	Source	Identifier
G418	Gibco	Cat. no. 11811-023
Doxycycline	Sigma	Cat. no. D9891
Dimethylformamide	Sigma-Aldrich	Cat. no. D4551
Puromycin	Takara	Cat. no. 631305
Ethylene glycol-bis (succinic acid N-hydroxysuccinimideester)	Thermo Fisher	Cat. no. 21565
Dimethylsulfoxide	Thermo Fisher	Cat. no. BP231-1
Formaldehyde solution 16%	Thermo Fisher	Cat. no. P128908
Glycine	Thermo Fisher	Cat. no. 15527-013
cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail	Roche	Cat. no. 11836170001
Mbol	NEB	Cat. no. R0147
Biotin-14-dATP	Invitrogen	Biotin-14-dATP
dCTP	Invitrogen	Cat. no. 18253-013
dTTP	Invitrogen	Cat. no. 18255-018
dGTP	Invitrogen	Cat. no. 18254-011
Tris-HCl, pH 7.5, 1M	Thermo Fisher	Cat. no. 15567027
NaCl, 5 M	Thermo Fisher	Cat. no. AM9759
MgCl <sub>2</sub> 1M	Thermo Fisher	Cat. no. AM9530G
IGEPAL CA-630	Sigma	Cat. no. 18896-50ML
EDTA	Invitrogen	Cat. no. 15575-038
Protein-A magnetic beads	Pierce	Cat. no. 88846
Lithium chloride, 8M	Sigma-Aldrich	Cat. no. L7026
Sodium bicarbonate	Sigma-Aldrich	Cat. no.144-55-8
T4 DNA ligase	NEB	Cat. no. M0202L
BSA, 50 mg ml <sup>-1</sup>	Invitrogen	Cat. no. AM2616
Tween-20	BIO-RAD	Cat. no. 170-6531
DNA polymerase I, large (Klenow) fragment	NEB	Cat. no. M0210S
Qubit dsDNA HS Assay kit	Invitrogen	Uniprot ID <a href="#">Q32851</a>
Dynabeads MyOne Streptavidin C-1	Invitrogen	Cat. no. 65001
Phusion PCR Master Mix	NEB	Cat. no. M0531
Nextera XT Index kit	Illumina	Cat. no. FC-131-1001
SPRIselect	Beckman Coulter	ID <a href="#">B23318</a>
MinElute Reaction Cleanup kit	QIAGEN	Cat. no. 28204
QIAQuick PCR Purification kit	QIAGEN	Cat. no. 28104
PBS, 1x	Thermo Fisher	Cat. no. 10010031
RNase	Thermo Fisher	Cat. no. EN0531
Proteinase K	Thermo Fisher	Cat. no. AM2546
Protein G Dynabeads	Thermo Fisher	Cat. no. 10003D
2-Mercaptoethanol	Thermo Fisher	Cat. no. 21985023
Buffer RLT Plus	QIAGEN	Cat. no. 1053393
RNeasy Plus Micro kit	QIAGEN	Cat. no. 74034
Trypsin-EDTA (0.05%)	Gibco	Cat. no. 25300054
RIPA buffer (10x)	CST	Cat. no. 9806
NuPAGE LDS Sample buffer (4x)	Thermo Fisher	Cat. no. NP0007
NuPAGE 4 to 12%, Bis-Tris, 1.0 mm, Mini Protein Gel, 10-well	Thermo Fisher	Cat. no. NP0321BOX
iBlot 2 Transfer Stacks, nitrocellulose, mini	Thermo Fisher	Cat. no. IB23002
Tris Buffered Saline with Tween 20	CST	Cat. no. 99975
SuperSignal West Femto Maximum Sensitivity Substrate	Thermo Fisher	Cat. no. 34095
Polysine adhesion glass slides	Electron Microscopy Sciences	Cat. no. 63412-01

Continued

Reagent or resource	Source	Identifier	Reagent or resource	Source	Identifier
Silicone Isolators	Electron Microscopy Sciences	Cat. no. 70339-05	EasySep Mouse CD4 Positive Selection kit II	Stemcell	Cat. no. 18952
Ethanol	Decon Laboratories	2716	<b>Recombinant DNA</b>		
Dimethylformamide	Sigma-Aldrich	Cat. no. D4551	Plasmid: pInducer-20	Addgene	Cat. no. 44012; RRID: <a href="#">Addgene_44012</a>
Triton X-100	Sigma-Aldrich	Cat. no. T8787-250ML	Plasmid: pSL21-vex	Addgene	Cat. no. 158230
Polyvinylsulfonic acid	Sigma-Aldrich	Cat. no. 278424	Plasmid: MSCV-Cas9_Puro	Addgene	Cat. no. 65655
Fisherbrand Premium cover glasses	Fisher Scientific	Cat. no. 12-548-5M	Plasmid: pCL-Eco	Naviaux et al., 1996	
No-wrinkle rubber cement	Elmer's	Cat. no. 34633	<b>Deposited data</b>		
Slowfade Gold Antifade Reagent	Invitrogen by Thermo Fisher Scientific	ID <a href="#">S36936</a>	Hi-C, HiChIP, ChIP-seq, RNA-seq	This study	<a href="#">GSE178348</a>
Secondary Oligopaint probe NDB_1279_Atto565: /5ATTO565N/ AGC GAA TCC GAC GCA CCG CTA /3ATTO565N/			<b>Publicly accessible data</b>		
Secondary Oligopaint probe NDB_1281_Alexa647: /5Alex647N/ AGG ACC CTG TTC GGC TAA CCA /3Alex647N/			TCF-1 ChIP-seq in thymus	<a href="#">24</a>	<a href="#">GSE46662</a>
Dextran sulfate	Sigma-Aldrich	Cat. no. D8906-50G	CTCF ChIP-seq in DP T cells	<a href="#">27</a>	<a href="#">GSE141853</a>
20x SSC buffer	Corning	Cat. no. 46-020-CM	SMC1 HiChIP in DP T cells	<a href="#">27</a>	<a href="#">GSE141853</a>
PBS, 10x	Stemcell Technologies	Cat. no. 37354	3e Hi-C during T cell development	<a href="#">25</a>	<a href="#">GSE79422</a>
Formamide, molecular grade	Promega	Cat. no. H5052	Histone modifications during T cell development	<a href="#">29</a>	<a href="#">GSE31235</a>
Formamide, >99.5% (4x4l)	Sigma-Aldrich	Cat. no. 221198-4X4L	ATAC-seq during T cell development	<a href="#">49</a>	<a href="#">GSE100738</a>
dNTP mixture, 100 mM (25 mM of each)	Thomas Scientific	Cat. no. C788T68	RNA-seq during T cell development	<a href="#">25</a>	<a href="#">GSE79422</a>
RNase A, DNase and protease-free (10 mg ml <sup>-1</sup> )	Thermo Fisher	Cat. no. EN0531	Mouse B cell hic	<a href="#">42</a>	<a href="#">4DNFISA93XFU</a>
Defy & Inspire clear nail polish	Target Corporation	Cat. no. 052-06-3573	Mouse cardiomyocyte hic	Rosa-Garrido et al., 2017	<a href="#">GSE96693</a>
<b>Critical commercial assays</b>			Mouse mESC hic	Bonev et al., 2017	<a href="#">4DNFIC21MG3U</a>
Illumina Tagment DNA Enzyme and Buffer Large kit	Illumina	Cat. no. 20034198	Human hESC hic	Oksuz et al., 2020	<a href="#">4DNFI7JNCNFB</a>
SMARTer Stranded Total RNA-Seq kit v2 - Pico Input Mammalian	Takara	Cat. no. 634413	Human IMR90 hic	<a href="#">45</a>	<a href="#">4DNFIJTOIGOI</a>
NEBNext Ultra II DNA Library Prep kit for Illumina	NEB	Cat. no. E76455	Human olfactory sensory neurons hic	Zazhytska et al., 2021	<a href="#">4DNFI7V61PXA</a>
Arima-Hi-C+ kit	Arima Genomics		Human HCT116 hic	<a href="#">42</a>	<a href="#">4DNFIV3PIEQJ</a>
Accel-NGS 2S Plus DNA Library kit	Swift Biosciences	Cat. no. 21024	Human K562 hic	Ray et al., 2019	<a href="#">4DNFI244AS29</a>
2S Set A Indexing kit	Swift Biosciences	Cat. no. 26148	Human HFF hic	Oksuz et al., 2020	<a href="#">4DNFIWMLVHWX</a>
KAPA Library Amplification kit with Primer Mix	Roche	KK2620	<b>Experimental models: cell lines</b>		
Kapa library Quantification kit	Roche	7960140001	NIH 3T3	ATCC	Cat. no. CRL-1658 RRID: <a href="#">CVCL_0594</a>
D1000 ScreenTape	Agilent	Cat. no. 5067-5582	HEK293T	ATCC	Cat. no. CRL-3216 RRID: <a href="#">CVCL_0063</a>
D1000 Reagents	Agilent	Cat. no. 5067-5583	Scid.adh	W. Pear laboratory at UPENN	
High Sensitivity D1000 ScreenTape	Agilent	Cat. no. 5067-5584	<b>Experimental models: organisms and strains</b>		
High Sensitivity D1000 Reagents	Agilent	Cat. no. 5067-5585	Mouse: C57BL/6J	Jackson Laboratory	RRID:IMSR_JAX:000664
Genomic DNA ScreenTape	Agilent	Cat. no. 5067-5365	Mouse: B6(Cg)-Tcf7tm1Hhx/J	Jackson Laboratory	RRID:IMSR_JAX:030909
Genomic DNA Reagents	Agilent	Cat. no. 5067-5366	Mouse: B6.Cg-CommD10Tg(Vav1-icre)A2Kio/J	Jackson Laboratory	RRID:IMSR_JAX:008610
RNA ScreenTape	Agilent	Cat. no. 5067-5576	<b>Software and algorithms</b>		
RNA ScreenTape Ladder	Agilent	Cat. no. 5067-5578	FastQC		<a href="https://www.bioinformatics.babraham.ac.uk/projects/fastqc/">https://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a> ; RRID:SCR_014583
RNA ScreenTape Sample Buffer	Agilent	Cat. no. 5067-5577	Trim Galore		<a href="https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/">https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/</a> ; RRID:SCR_011847
EasySep Release Mouse Biotin Positive Selection kit	Stemcell	Cat. no. 17655			

Continued

Continued

Reagent or resource	Source	Identifier
Bowtie2	50	<a href="http://bowtie-bio.sourceforge.net/bowtie2/index.shtml">http://bowtie-bio.sourceforge.net/bowtie2/index.shtml</a>
STAR	51	<a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>
Picard		<a href="https://broadinstitute.github.io/picard/">https://broadinstitute.github.io/picard/</a>
HTSeq	52	<a href="https://htseq.readthedocs.io/en/master/">https://htseq.readthedocs.io/en/master/</a> ; RRID:SCR_005514
DESeq2	53	<a href="https://github.com/Bioconductor-mirror/DESeq2">https://github.com/Bioconductor-mirror/DESeq2</a> ; RRID:SCR_015687
R		<a href="https://cran.r-project.org/">https://cran.r-project.org/</a> ; RRID:SCR_001905
MACS2	54	<a href="https://github.com/taoliu/MACS">https://github.com/taoliu/MACS</a>
HOMER	55	<a href="http://homer.ucsd.edu/homer/">http://homer.ucsd.edu/homer/</a> ; RRID:SCR_010881
bedtools		<a href="https://bedtools.readthedocs.io/en/latest/#">https://bedtools.readthedocs.io/en/latest/#</a> ; RRID:SCR_006646
wigToBigWig		<a href="http://hgdownload.cse.ucsc.edu/admin/execute/">http://hgdownload.cse.ucsc.edu/admin/execute/</a>
Hi-C-Pro	56	<a href="https://github.com/nservant/HiC-Pro">https://github.com/nservant/HiC-Pro</a> ; RRID:SCR_017643
ggplot2		<a href="https://ggplot2.tidyverse.org/">https://ggplot2.tidyverse.org/</a>
Metascape	57	<a href="https://metascape.org/gp/index.html#/main/step1">https://metascape.org/gp/index.html#/main/step1</a> ; RRID:SCR_016620
Juicer	58	<a href="https://github.com/aidenlab/juicer">https://github.com/aidenlab/juicer</a>
Cooler		<a href="https://github.com/open2c/cooler">https://github.com/open2c/cooler</a>
cooltools		<a href="https://github.com/open2c/cooltools">https://github.com/open2c/cooltools</a>
HiCExplorer	59	<a href="https://hicexplorer.readthedocs.io/en/latest/">https://hicexplorer.readthedocs.io/en/latest/</a>
UpSetR	60	<a href="https://cran.r-project.org/web/packages/UpSetR/index.html">https://cran.r-project.org/web/packages/UpSetR/index.html</a>
deeptools	61	<a href="https://deeptools.readthedocs.io/en/develop/">https://deeptools.readthedocs.io/en/develop/</a>
dekker-cworld		<a href="https://github.com/dekkerlab/cworld-dekker">https://github.com/dekkerlab/cworld-dekker</a>
hichipper	62	<a href="https://github.com/aryeelab/hichipper">https://github.com/aryeelab/hichipper</a>
Coolpup.py	63	<a href="https://github.com/open2c/coolpuppy">https://github.com/open2c/coolpuppy</a>
TANGO		
Fiji		
IMARIS		
Sushi		
OligoMiner		
FlowJo v.10.6.1		

**Resource availability.** Further information and requests for resources and reagents should be directed to the corresponding author, G.V.

**Materials availability.** This study did not generate new unique reagents.

**Cell culture.** NIH 3T3 cells were purchased from the American Type Culture Collection (ATCC) as described<sup>17</sup> and maintained at a low passage number (<12) using high glucose 1× DMEM medium with L-glutamine (Gibco) with 100 U ml<sup>-1</sup> penicillin and 100 mg ml<sup>-1</sup> streptomycin (Gibco) and 10% bovine serum (Gibco). The 293T (ATCC) cells were maintained in high glucose DMEM 1× medium with L-glutamine (Gibco) and 100 U ml<sup>-1</sup> penicillin and 100 mg ml<sup>-1</sup> streptomycin (Gibco) with 10% fetal bovine serum (Fisher Scientific). The Scid.adh cell line, a pro-T cell line derived from spontaneous thymic lymphomas<sup>64</sup>, was a kind gift from W. Pearl's laboratory at the University of Pennsylvania. These cells were grown in RPMI 1640 medium (Invitrogen), supplemented with 10% fetal bovine serum (Fisher Scientific), 1 mM sodium pyruvate (Gibco), 1% non-essential amino acids (Gibco), 2 mM L-glutamine (Lonza), 1% penicillin-streptomycin and 0.1% 2-mercaptoethanol (Gibco). All cells were grown at 37 °C and 5% CO<sub>2</sub>.

**Mice.** Female and male breeder Vav-iCre transgenic mice (strain 008610)<sup>65-67</sup> and Tcf7<sup>Cre</sup> mice (strain 030909)<sup>68</sup> were purchased from the Jackson Laboratory. Tcf7<sup>-/-</sup> mice were generated by breeding Tcf7<sup>Cre</sup> mice, in which two loxP sites are inserted on either side of exon 2 of the Tcf7 gene, with Vav-iCre mice. The F1 generation was backcrossed to Tcf7<sup>Cre</sup> mice to reach homozygous floxed Cre<sup>+</sup> experimental mice. Vav-iCre<sup>+</sup> control mice were generated by crossing Vav-iCre mice with C57BL/6J (strain 000664) and used as controls in all experiments. Experimental and control mice were 6–10 weeks old of either sex. All mice were bred and housed in an American Association for the Accreditation of Laboratory Animal Care accredited vivarium at the University of Pennsylvania. All husbandry and experimental procedures were performed according to the protocol reviewed and approved by the Institutional Animal Care and Use Committee.

**Cell preparation.** Single-cell suspensions were prepared from thymi of mice by dissociation of tissue through 70-µm mesh filters (Falcon) in RPMI 1640 (Corning) + 1% FBS (Gemini). DP T cells were isolated utilizing serial magnetic bead separation. First EasySep Release Mouse Biotin Positive Selection kit was utilized with biotin anti-mouse CD8a (53-6.7), followed by isolation with EasySep Mouse CD4 Positive Selection kit II. Purity of DP cells after isolation was ≥90% as assessed by flow cytometry.

**Lentiviral packaging and transduction for NIH 3T3 cells.** For TCF-1 overexpression in 3T3 cells, the complementary DNA encoding the short isoform of TCF-1 (p33)<sup>17</sup> was cloned into a Tet-inducible lentiviral vector pINDUCER-20 (Addgene) using a Gateway cloning strategy (Gateway Clonase II, Invitrogen). Lentiviral vectors were packaged in HEK293T cells. Briefly, 4 × 10<sup>6</sup> HEK293T cells were plated in 5 ml medium in 10-cm dishes on the day before transfection. During transfection, 3.9 µg of pINDUCER-20-TCF-1 plasmid was co-transfected with packaging plasmids, 2.6 µg of pCMVδ and 1.3 µg of VSV-G, using 26 µl FuGene HD (Promega). The cells were returned to the incubator for 6 h. Subsequently, the medium was changed to fresh medium. Virions were collected 24 and 48 h after transfection, snap-frozen and stored at -80 °C for future use. NIH 3T3 cells were transduced by addition of virions to culture medium supplemented with polybrene (Sigma-Aldrich, cat. no. H9268) at 8 mg ml<sup>-1</sup> followed by centrifugation at room temperature for 20 min at 2,000 r.p.m. Transduced cells were selected using 1 mg ml<sup>-1</sup> G418 (Gibco) for 7 d. TCF-1 expression was induced using 500 mg ml<sup>-1</sup> of doxycycline for 72 h. Cells which have been transduced with the same vector but have not been treated with doxycycline are referred to as 'untreated NIH 3T3' while doxycycline-treated cells are referred to as 'TCF-1 induced NIH 3T3'.

**Retroviral packaging and transduction for Tcf7<sup>-/-</sup> in Scid.adh cell line.** CRISPR/Cas9 system was used to delete TCF-1 in Scid.adh cells. Single-guide RNAs (sgRNAs) targeting the DNA-binding domain (HMG-box) were designed using CRISPR Targets tracks of the University of California, Santa Cruz's Genome Browser. ACCGCAACCAGATCTGGGTCGCA and AAATGCGACCCAGGATCTGGTGTG were used as sgRNAs. sgRNA was cloned into pSL21-vec (Addgene, 158230)<sup>69</sup> and MSCV-Cas9-puro plasmid (Addgene, 65655) was used for retroviral introduction of Cas9 into the cells. For retroviral packaging, 4 × 10<sup>6</sup> 293T cells were plated in 4 ml DMEM in 10-cm dishes on the day before transfection. Immediately before transfection, chloroquine was added to the medium to a final concentration of 25 mM. The retroviral construct/empty vector and the pCL-Eco plasmid were transiently co-transfected using Lipofectamine 3000 (Invitrogen). The cells were returned to the incubator for 6 h. Subsequently, the medium was changed to fresh medium. Virions were collected 24 and 48 h after transfection and were snap-frozen and stored at -80 °C for future use. Scid.adh cells were transduced by addition of Cas9 and sgRNA virions to culture medium supplemented with polybrene (Sigma-Aldrich, cat. no. H9268) at 8 mg ml<sup>-1</sup> and spininfected at 1,750 r.p.m. for 25 min. Cas9-expressing cells were selected 4 d after spin infection with 1 µg ml<sup>-1</sup> puromycin for 7 d, followed by checking Cas9 expression on the immunoblot. To obtain single-cell clones of the TCF-1-knockout cells, we diluted bulk transduced cells to obtain a dilution of 1 cell

in 100  $\mu$ l medium and plated them in 96-well plates. Knockout clones G9 and E10 were selected after performing an immunoblot to check TCF-1 expression. Hi-C data in both clones were used to evaluate the effect of TCF-1 deletion in DN3 cell lines.

**Immunoblot.** Immunoblotting was performed on whole-cell lysates from Vav-Cre<sup>+</sup>, heterozygous and homozygous *Tcf7*<sup>-/-</sup> thymocytes as well as cells derived from Scid.adh *Tcf7* knockout single-cell clones. Cells were lysed with 1 $\times$  RIPA buffer supplemented with proteinase inhibitor cocktail. Equal numbers of cells per condition were utilized and equal volumes of lysate were loaded on a NuPAGE 4–12% Bis-Tris gel and transferred using the iBlot 2 Gel Transfer Device. Membranes were blocked with 5% nonfat dry milk in 1 $\times$  TBST buffer followed by incubation with primary rabbit-anti-mouse and HRP conjugated secondary antibodies. Blots were visualized with SuperSignal West Femto Maximum Sensitivity Substrate on the ChemiDoc Imaging system.

**Flow cytometry.** Single-cell suspensions were stained following standard protocols. The fluorochrome-conjugated, anti-mouse antibodies were as follows: PE CD4 (RM4-4), APC CD8a (53-6.7), BV785 CD44 (IM7), PECy7 CD25 (PC61) and APC streptavidin. Antibodies used in the lineage cocktail (Lin) include biotinylated antibodies against GR-1 (RB6-8C5), NK1.1 (PK136), CD11b (M1/70), Ter-119 (TER-119) and B220 (RA3-6B2). Cells were stained with LIVE/DEAD Fixable Aqua Dead Cell Stain kit for discrimination of live cells. Data were collected on an LSRII running DIVA software (BD Biosciences) and were analyzed with FlowJo software v.10.6.1.

**ChIP-seq.** ChIP-seq was performed as previously described<sup>27</sup>. Cells were fixed for 30 min in RPMI 1640 (Gibco) with 1% FBS (Sigma-Aldrich) with 1.5 mM ethylene glycol-bis (succinic acid *N*-hydroxysuccinimide ester) (EGS) (Thermo Fisher, 21565) in dimethylsulfoxide followed by 15 min with 1% formaldehyde (formaldehyde solution 16%, Thermo Fisher, 28906) and quenched for 10 min with 0.125 M glycine. Cells were washed 2 $\times$  with PBS and frozen at –80 °C. Crosslinked cells were lysed and then sonicated for ten cycles at 10 s each with 50 s between cycles. Triton X-100 was added to a final concentration of 1% to centrifuge cleared lysates. Lysates were incubated overnight with Protein G Dynabeads (Thermo Fisher, 10003D) conjugated to 10  $\mu$ g of anti-SMC1 antibody (Bethyl, A300-055A), 10  $\mu$ g of anti-CTCF antibody (Millipore 07-729), 10  $\mu$ g of anti-NIPBL antibody (Bethyl, A301-779A), 10  $\mu$ g of anti-H3K27ac antibody (Abcam, ab4729) or 10  $\mu$ g of anti-TCF-1 antibody (CST, 2206S). Beads were washed and complexes were eluted for 30 min at 65 °C with shaking. After reversal of cross-linking, RNase and proteinase K treatment were performed and DNA was purified and quantified. Two biological replicates were generated for each experiment. Two separate aliquot of fixed cells per condition were used as replicates for each experiment. Library preparation was carried out using NEBNext Ultra II DNA Library Prep kit for Illumina (NEB) and were paired-end sequenced (38 bp + 38 bp) on a NextSeq 550 (Illumina).

**RNA-seq.** Cells were washed once with 1 $\times$  PBS before resuspending pellet in 350  $\mu$ l Buffer RLT Plus (QIAGEN) with 1% 2-mercaptoethanol (Sigma), vortexed briefly and stored at –80 °C. Subsequently, total RNA was isolated using the RNeasy Plus Micro kit (QIAGEN). RNA integrity numbers were determined using a TapeStation 2200 (Agilent) and all samples used for RNA-seq library preparation had RNA integrity numbers >9. Libraries were prepared using the SMARTer Stranded Total RNA-seq kit v2- Pico Input Mammalian kit (Takara). Two biological replicates were generated for each experiment. Two separate aliquot of fixed cells per condition were used as replicates for each experiment. Libraries were validated for quality and size distribution using a TapeStation 2200 (Agilent). Libraries were paired-end sequenced (38 bp + 38 bp) on a NextSeq 550 (Illumina).

**ATAC-seq.** ATAC-seq was performed as previously described with minor modifications<sup>27,70</sup>. Fifty thousand cells were pelleted at 550g and washed with 50  $\mu$ l ice-cold 1 $\times$  PBS, followed by treatment with 50  $\mu$ l lysis buffer (10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl<sub>2</sub> and 0.1% IGEPAL CA-630). Nuclei pellets were resuspended in 50  $\mu$ l transposition reaction containing 2.5  $\mu$ l Tn5 transposase (FC-121-1030; Illumina). The reaction was incubated in a 37 °C heat block for 45 min. Tagmented DNA was purified using a MinElute Reaction Cleanup kit (QIAGEN) and amplified with varying cycles, depending on the side reaction results. Libraries were purified using a QIAquick PCR Purification kit (QIAGEN). Libraries were validated for quality and size distribution using a TapeStation 2200 (Agilent). Libraries were paired-end sequenced (38 bp + 38 bp) on a NextSeq 550 (Illumina).

**HiChIP.** HiChIP was performed as described<sup>71</sup> using an antibody to SMC1 (Bethyl A300-055A). Briefly, 10 $\times$ 10<sup>6</sup> cells were crosslinked with 1% formaldehyde (Thermo Fisher Scientific, cat. no. 28908) for 10 min and subsequently quenched with 0.125 M glycine (Invitrogen, cat. no. 15527-013). Chromatin was digested using MboI restriction enzyme (NEB, cat. no. R0147), followed by biotin incorporation with Biotin-14-dATP (Invitrogen, cat. no. 19524-016) during repair, ligation and sonication. Sheared chromatin was fourfold diluted with ChIP dilution buffer (16.7 mM Tris, pH 7.5, 167 mM NaCl, 1.2 mM EDTA, 0.01% SDS and

1.1% Triton X-100), pre-cleared and then incubated with SMC1 antibody at 4 °C overnight. Chromatin–antibody complexes were captured by Protein-A magnetic beads (Pierce, cat. no. 88846) and subsequently washed with Low-Salt Wash Buffer, High-Salt Wash Buffer, LiCl Wash Buffer and eluted. DNA was purified with a MinElute PCR Purification kit (QIAGEN, cat. no. 28004) and quantified using a Qubit dsDNA HS Assay kit (Invitrogen, cat. no. Q32851). 50–150 ng was used for capture with Dynabeads MyOne Streptavidin C-1 (Invitrogen, cat. no. 65001) and an appropriate amount of Tn5 enzyme (Illumina, cat. no. FC-121-1030) was added to captured DNA to generate the sequencing library. Libraries were validated for quality and size distribution using a TapeStation 2200 (Agilent). Paired-end sequencing (38 bp + 38 bp) was performed on a NextSeq 550.

**Hi-C.** Hi-C libraries were generated on 10<sup>6</sup> cells using with Arima-Hi-C+ kit (Arima Genomics) and Accel-NGS @S Plus DNA Library kit (21024 Swift Biosciences), according to the manufacturer's recommendations. Libraries were validated for quality and size distribution using Qubit dsDNA HS Assay kit (Invitrogen, cat. no. Q32851) and TapeStation 2200 (Agilent). Libraries were paired-end sequenced (66 bp + 66 bp) on NovaSeq 6000 (Illumina).

**Oligopaint probe designing.** The TADs and boundaries at CD8a locus were identified using insulation score<sup>26</sup> (as described in more details subsequently). The OligoMiner design pipeline<sup>72</sup> was then applied to design DNA FISH probes to the respective coordinates, chr6: 71163477–71432963 (TAD1) and chr6: 71440321–71825773 (TAD2). Oligopaint probes were designed to have 80 bases of homology with an average of five probes per kb and were purchased from Twist Bioscience. Oligopaint probes were synthesized as described previously<sup>73,74</sup>.

**Oligopaint DNA FISH.** Cells were incubated on 75 $\times$ 25 mm Polysine adhesion glass slides (Electron Microscopy Sciences, cat. no. 63412-01) with silicone isolators (Electron Microscopy Sciences, cat. no. 70339-05) for 1 h at room temperature in a humidified chamber. Cells were then fixed to the slide for 10 min with 4% formaldehyde in PBS and rinsed briefly in PBS. Slides were stored in PBS at 4 °C until use.

For DNA FISH primary probe hybridization, slides were warmed to room temperature in PBS for 10 min. Cells were permeabilized in 0.5% Triton-PBS for 15 min. Cells were then dehydrated in an ethanol row, consisting of 2-min consecutive incubations in 70%, 90% and 100% ethanol. The slides were then allowed to air dry for about 2 min at room temperature. Cells were then washed in 4 $\times$  saline sodium citrate-Tween (SSCT)/50% formamide (0.6 M NaCl, 0.06 M sodium citrate and 0.2% Tween-20) for 5 min, pre-denatured in 4 $\times$  SSCT/50% formamide at 92 °C for 2.5 min and then in 4 $\times$  SSCT/50% formamide at 60 °C for 10 min. Then, 50 pmol of primary Oligopaint probes in hybridization buffer (10% dextran sulfate/2 $\times$  SSCT/50% formamide/4% polyvinylsulfonic acid (PVSA)/1.4 mM dNTPs) were then added to the cells, covered with a 24 $\times$ 50 mm glass coverslip (Fisher Scientific, cat. no. 22037170) and sealed with no-wrinkle rubber cement (Elmer's). Cells were denatured by placing slides on a heat block in a water bath set to 92 °C for 2.5 min, after which slides were transferred to a humidified chamber and incubated overnight at 37 °C. Approximately 16–18 h later, coverslips were removed with a razor blade and slides were washed in 2 $\times$  SSCT (0.3 M NaCl, 0.03 M sodium citrate and 0.1% Tween-20) at 60 °C for 15 min, 2 $\times$  SSCT at room temperature for 10 min and 0.2 $\times$  SSC at room temperature for 10 min.

Secondary probes (10 pmol/25  $\mu$ l<sup>-1</sup>) containing fluorophores Atto-565 and Alexa-647 (IDT) were added to the slides, again resuspended in hybridization buffer containing only 10% formamide and covered with a coverslip sealed with rubber cement. Slides were incubated at room temperature for 2 h in a dark, humidified chamber, followed by washes in 2 $\times$  SSCT at 60 °C for 5 min, 2 $\times$  SSCT at room temperature for 5 min. All slides were stained in PBS with DAPI (0.1 mg ml<sup>-1</sup>) and then washed in 2 $\times$  SSC. Slides were mounted with Slowfade Gold Antifade Reagent (Invitrogen, cat. no. S36936) and sealed with clear fingernail polish.

**Image acquisition and processing.** Imaging was performed on a Leica TCS SP8 Multiphoton Confocal microscope using a 1.3 NA  $\times$ 40 oil immersion objective with pixels of 541.6 $\times$ 541.6 nm. Fields of view were selected, such that confluence was balanced to provide maximum data points and to ensure proper cell segmentation during downstream analysis. Z-stacks were determined, such that both homologs were within the imaged space and represented 10  $\mu$ m in total axial thickness. Localizations were then recorded in 0.3- $\mu$ m steps.

Representative images were also obtained on a Leica TCS SP8 Multiphoton Confocal microscope, this time using a 1.4 NA  $\times$ 63 oil immersion objective with pixels of 343.9 $\times$ 343.9 nm. Z-stacks were determined and recorded as with the 1.3 NA  $\times$ 40 oil immersion objective. Each cell, allele and locus for each strain were individually processed using Fiji software to generate orthogonal projections of the Z-stacks and to split channels into individual TIFs.

3D reconstructions of representative cells were rendered using IMARIS v.7.4.2 software (Bitplane AG). DNA FISH dots were generated using the Spots tool with a 0.8  $\mu$ m diameter, created at the intensity mass center of the fluorescent probe signal. Nuclear volume was created using the Surfaces tool with automatic settings based on the fluorescent signal from the DAPI stain.

**ChIP-seq data analysis.** Bowtie2 (ref.<sup>50</sup>) was used for alignment of ChIP-seq data. Reads aligned to the mitochondrial genome or chrY as well as reads mapping to multiple genomic loci were discarded from downstream analyses. Bigwig files were generated by bedtools<sup>55</sup> genomecov and wigToBigWig normalizing tracks to tags per million. For peak calling, macs2 (ref.<sup>54</sup>) with 'macs2 callpeak -c inputfile--nolambda--nomodel--keep-dup all -p 0.00001' was used. Input sample was prepared by the same approach without immunoprecipitation and used as input control for peak calling. ChIP-seq peaks from two conditions and both replicates were merged and the number of fragments in each peak were counted with bedtools. The count data of each peak was then fed to DESeq2 for differential analysis.

**RNA-seq data analysis.** RNA-seq samples were aligned by STAR\_2.5.0a\_alpha<sup>51</sup> with parameters '-readFilesCommand zcat--outFilterMultimapNmax 1--outSAMtype SAM--alignEndsType Local--outReadsUnmapped Fastx--outFilterMismatchNmax 1--alignMatesGapMax 400,000--sjdbGTFfile: HTSeq v.0.6.1' facilitated counting RNA-seq reads on Gencode v.M11 gene models with parameters '-s yes -t exon -m intersection-nonempty'. DESeq2 was subsequently applied on gene counts to identify genes that were differentially expressed.

**ATAC-seq data analysis.** Bowtie2 was used for alignment (bowtie2 -p 20 -X2000 -t). Reads aligned to the mitochondrial genome or chrY as well as reads mapping to multiple genomic loci were discarded from downstream analyses. Additionally, Picard was used to mark and remove duplicates. Furthermore, for each ATAC-seq library the insert size was calculated by Picard. The insert size distribution of sequenced fragments had clear periodicity of approximately 200 bp, suggesting many fragments are protected by integer multiples of nucleosomes. Bigwig files were generated by bedtools genomecov and wigToBigWig normalizing tracks to tags per million. For peak calling, macs2 with 'macs2 callpeak--nomodel -B--keep-dup all--broad--broad-cutoff 0.1 -q 0.1' was used.

**HiChIP and Hi-C data analysis.** HiChIP data processing. Raw reads for HiChIP sample were processed with Hi-C-Pro (v.2.11.1)<sup>56</sup> to obtain putative interactions with default parameters, except 'ligation\_site' = GATCGATC and 'genome\_fragment' was generated for the MboI restriction enzyme. For the purpose of downstream analysis, ValidPairs were converted to cool files and hic files using the 'hicpro2higlass.sh' and 'hicpro2juicebox.sh' in utils of Hi-C-Pro, respectively.

**Hi-C data processing.** For the 3e Hi-C data during T cell development<sup>55</sup>, we downloaded the raw data from GSE79422 and processed it with Hi-C-Pro without setting the 'ligation\_site' and 'genome\_fragment'. For the Hi-C data that we generated in the wild-type, Tcf7<sup>-/-</sup> DPs and DN3s, we processed the data with Hi-C-Pro using parameters 'ligation\_site' = GATCGATC and 'genome\_fragment' was generated for the MboI restriction enzyme, keeping other parameters as default. ValidPairs generated by Hi-C-Pro were further converted to cool and hic files as described above.

**Compartment analysis.** We performed the compartment analysis of HiChIP and Hi-C data following the tutorial on cooltools GitHub page (<https://github.com/open2c/cooltools>) and further visualized the compartment strength utilizing saddle plot implemented in cooltools. A scatter-plot of PC1 values between the two different conditions was produced with ggplot2 in R, of which the correlation was calculated with stat\_cor() function in ggpubr package. The compartment strength was calculated with the top 30% PC1 regions representing A compartment and bottom 30% PC1 value representing B compartment. We calculated the compartment strength, AA interactions, BB interactions and AB interactions of each chromosome and used a box plot to compare the two conditions, implementing a Student's *t*-test for statistical analysis.

**TAD and boundary analysis.** TADs and boundaries were identified with insulation score<sup>56</sup>, which was implemented in cooltools using the 'diamond-insulation' function. We calculated the genome-wide insulation score at 20-kb resolution with a 500-kb window size utilizing cooltools diamond-insulation. The insulation score result was converted to a bedgraph file, which was further converted to bigwig files for visualization purposes. Boundaries were identified with the bins whose boundary strength >0.1 and were further used to define TADs with a custom Python script.

Common and unique boundaries between conditions and among T cell developmental stages were identified with a custom Python script, in which we clustered the boundaries whose minimum distance was less than 100 kb and the maximum distance between boundaries of within 200 kb. We generated a set of consensus boundaries based on the boundary clusters by taking the two ends of each cluster as boundary coordinates. In this way, we generated the common and unique boundaries between conditions and identified the boundaries that are shared between different T cell developmental stages. The boundaries that were shared or unique to the six T cell developmental stages were visualized with UpSetR<sup>60</sup>.

Common and unique TADs between conditions were identified with bedtools intersect (-wao -f 0.8 -r)<sup>75</sup> and followed with filtering the TADs by only keeping TADs whose distances of starts and ends in two conditions were less than 100 kb. Only TADs that were larger than 200 kb were kept for downstream analysis. We conducted domain score analysis based on these common TADs between two

conditions. Cooler was used to dump the contact matrix from cool files and custom script was used to calculate the PETs within the TAD, which was divided by PETs between the TAD and the chromosome to generate the domain score.

**Directional index.** A matrix was generated from cool files using cooltools dump-cworld with parameter '-balancing-type IC\_unity'. Then directional index was calculated with script from dekker-cworld (<https://github.com/dekkerlab/cworld-dekker>). The results were further converted to bedgraph and bigwig files for visualization.

**TCF-1 association with interaction change.** In the TCF-1 association with increased intra-TAD interactions during T cell development, we identified the common TADs between five early stages and DP T cells. Then, we calculated the log<sub>2</sub>(fold change) of paired-end-tag sequence (PETs) within the common TADs from early stages to DP, as well as TCF-1 densities in the TADs. The relationship between TCF-1 density and increase of intra-TAD interactions was visualized with box plot in ggplot2 and the correlation was calculated with stat\_cor() function.

In the association of TCF-1 binding and domain score change in NIH 3T3, we subtracted the domain scores of common TADs between TCF-1-induced and normal NIH 3T3 and calculated the TCF-1 densities in the TADs. Then we plotted the relationship between these two features with box plot in ggplot2. For the bin-based analysis, we calculated the number of TCF-1 peaks within 50 kb of each bin and further classified them according to whether there was TCF-1 binding in that bin. Then we calculated the total PETs between the bin and other regions of the TAD in both TCF-1-induced and normal NIH 3T3, which was used to determine log<sub>2</sub>(fold change) of interactions of the bin. We plotted the relationship between number of TCF-1 peaks within 50 kb and log<sub>2</sub>(fold change) of interactions of the bins using box plot in ggplot2 and further separated the bins into those with or without TCF-1 binding. We used the classified bins and aggregated PETs between bin pairs with the same TCF-1 peaks and binding status. The log<sub>2</sub>(fold change) of the pairwise interaction between the bins was calculated and visualized with Python script.

**Loop analysis.** Valid pairs from Hi-C-Pro and SMC1 ChIP-seq peaks were fed to hichipper<sup>62</sup> to call significant interactions. Only loops with PETs ≥ 2 were kept for downstream analysis. The comparison of increased loops after TCF-1 induction in fibroblast between different peaks was conducted using violin plot and box plot in addition to stat\_compare\_means() for statistical tests in ggplot2.

We used loops from Mango to conduct the association analysis of loops with TCF-1 and CTCF ChIP-seq peaks, utilizing the bedtools pairtobed function. The TCF-1 and CTCF peaks were extended 5 kb at both sides. Then we plotted the loops associated with TCF-1 and CTCF peaks with a Venn diagram in Python (<https://pypi.org/project/matplotlib-venn/>). Genome Browser views of cohesin loops and ChIP-seq signals on TCF-1, CTCF and SMC1 were shown using Integrative Genomics Viewer.

**3D pileup analysis.** Local pileup analysis at different sets of peaks was conducted with coolpup.py<sup>63</sup> using the parameters '-pad 250 -local'. The average interaction of the upright corner of pileup plots is quantified with a custom Python script, by parsing the results from coolpup.py. Interactions between regions in bedfiles were conducted with parameters '-mindist 200,000--maxdist 2,000,000'. Pileup of loops was also plot with coolpup.py. To identify the regions that have insulation change, we quantified the average interactions between upstream and downstream of each peak. The peaks whose difference of average interactions between two conditions greater than 0.05 were considered a change of insulation. To perform multiple pileup analysis parallelly, we used GNU parallel to run the shell script.

**Stripe analysis.** We used Stripenn<sup>76</sup> to quantify stripiness of the stripe originating from the TCF-1 cluster between NOD and C57BL/6.

**One-dimensional pileup analysis.** The pileup analysis of one-dimensional features including insulation score, directional index, ChIP-seq and phastCons score was conducted with deeptools<sup>61</sup>.

**Visualization of one-dimensional and 3D features at selected regions.** We used HiCEXplorer<sup>59</sup> to visualize the interactions, insulation score, genes and ChIP-seq signals of selected regions. The cool files were converted to h5 format with hicConvertFormat using parameter '-load\_raw\_values'. The color scales of the interaction heat maps were normalized with total PETs.

**Triangle heat maps.** Triangle heat maps for 3D chromatin conformation data and corresponding tracks were generated using Sushi R package (v.1.28.0)

**Quantification of inter-domain interactions.** The inter-domain interactions marked as yellow and sky-blue boxes in the Sushi plots were quantified using the matrix.fetch function in the cooler package<sup>77</sup>. The square root of vanilla coverage method<sup>45</sup> or matrix balancing normalization<sup>78</sup> was used.

**Detection of regions which gained or lost insulation at TCF-1 binding sites.** We used custom script following the tutorial on cooltools GitHub page (<https://github.com/open2c/cooltools>) to calculate the interactions in a 500-kb window centered

on each peak, for both untreated and TCF-1-induced 3T3 and for wild-type and TCF-1 knockout DN3 cells. Those peaks that have an increase or decrease of interactions between upstream and downstream of the peaks by greater than 0.05 were selected for TCF-1-induced 3T3 and TCF-1 knockout DN3 cells, respectively. The pileup at these peaks was plotted using coolpup.py.

**Distance to T cell development genes.** We downloaded the signature genes for each T cell development stage from single-cell studies of the thymus in mice<sup>38</sup>. Then, we calculated the distance of these genes to the closest TCF-1 and CTCF co-bound peak with the bedtools closest function.

**Oligopaint FISH imaging analysis.** The TANGO<sup>38</sup> plug-in of Fiji software<sup>79</sup> was utilized for the Oligopaint FISH imaging analysis. An image with a LIF file extension was used as an input. The imaging analysis was composed of two steps: (1) cell/probe segmentations and (2) the measurement of the overlapping volume and minimum distance between two probes. Before cell segmentation, Fast Filter 3D (median filter) and Misc 3D Filters (Gaussian 3D filter) were applied to the image. Then, the Otsu method was used for cell segmentation. After cell segmentation, Morphological Filters 3D (Fill Holes 3D and Fill holes 2D) was applied for cells as post-filtering. For probe channels, Fast Filters 3D (mean filter), Subtract Background 2D (radius of five pixels, rolling ball method) and Misc 3D filters (Gaussian 3D filter) were applied before segmentation. For segmentation, the percentage of bright pixels was set as 3.5%. As post-filtering, Morphological Filters 3D and Size and Edge Filter (minimum volume of 5) were applied. After cell segmentation, those including more than one cell were eliminated from the analysis. To measure the overlapping volume and the distance between two probes, object colocalization and minimal distance functions were utilized.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Genomics data generated in this study are publicly available in the Gene Expression Omnibus at accession code GSE178348. Oligopaint FISH data generated in this study are provided as Source Data. Source data are provided with this paper.

## Code availability

The codes for Stripenn and data analysis are available from the author's GitHub page ([https://github.com/VahediLab/TCF13D\\_code](https://github.com/VahediLab/TCF13D_code)).

## References

- Yoshida, H. et al. The *cis*-regulatory atlas of the mouse immune system. *Cell* **176**, 897–912 (2019).
- Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
- Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
- Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
- Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
- Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).
- Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
- Wolff, J. et al. Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* **46**, W11–W16 (2018).
- Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940 (2017).
- Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
- Lareau, C. A. & Aryee, M. J. hicchipper: a preprocessing pipeline for calling DNA loops from HiChIP data. *Nat. Methods* **15**, 155–156 (2018).

- Flyamer, I. M., Illingworth, R. S. & Bickmore, W. A. Coolpup.py: versatile pile-up analysis of Hi-C data. *Bioinformatics* **36**, 2980–2985 (2020).
- Carleton, M. et al. Signals transduced by CD3ε, but not by surface pre-TCR complexes, are able to induce maturation of an early thymic lymphoma in vitro. *J. Immunol.* **163**, 2576–2585 (1999).
- de Boer, J. et al. Transgenic mice with hematopoietic and lymphoid specific expression of Cre. *Eur. J. Immunol.* **33**, 314–325 (2003).
- Ogilvy, S. et al. Promoter elements of *vav* drive transgene expression in vivo throughout the hematopoietic compartment. *Blood* **94**, 1855–1863 (1999).
- Shimshak, D. R. et al. Codon-improved Cre recombinase (iCre) expression in the mouse. *Genesis* **32**, 19–26 (2002).
- Yang, Q. et al. TCF-1 upregulation identifies early innate lymphoid progenitors in the bone marrow. *Nat. Immunol.* **16**, 1044–1050 (2015).
- Chen, Z. et al. In vivo CD8(+) T cell CRISPR screening reveals control by Fli1 in infection and cancer. *Cell* **184**, 1262–1280 (2021).
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
- Mumbach, M. R. et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* **13**, 919–922 (2016).
- Beliveau, B. J. et al. OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes. *Proc. Natl Acad. Sci. USA* **115**, E2183–E2192 (2018).
- Moffitt, J. R. & Zhuang, X. RNA imaging with multiplexed error-robust fluorescence in situ hybridization (MERFISH). *Methods Enzymol.* **572**, 1–49 (2016).
- Rosin, L. F., Nguyen, S. C. & Joyce, E. F. Condensin II drives large-scale folding and spatial partitioning of interphase chromosomes in *Drosophila* nuclei. *PLoS Genet.* **14**, e1007393 (2018).
- Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
- Yoon, S., Chandra, A. and Vahedi, G. Stripenn detects architectural stripes from chromatin conformation data using computer vision. *Nat. Commun.* <https://doi.org/10.1038/s41467-022-29258-9> (2022).
- Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311–316 (2020).
- Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999 (2012).
- Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

## Acknowledgements

We thank R. B. Faryabi, K. Zaret, A. Karnay, E. J. Wherry and R. Jain for helpful discussions. We thank J. Shi and his laboratory for help and discussion about the CRISPR/Cas9 plasmids. This work was supported by National Institutes of Health grants UC4 DK112217, U01DK112217, R01HL145754, U01DK127768 and U01DA052715, the Burroughs Wellcome Fund, the Chan Zuckerberg Initiative, W. W. Smith Charitable Trust, the Penn Epigenetics Institute pilot and the Sloan Foundation awards to G.V.

## Author contributions

A.C. performed all genomics and sequencing experiments in collaboration with N.G. N.G. generated the *Tcf7<sup>-/-</sup>* mice. A.C. performed computational analysis of ChIP-seq data in DN3. W.W. performed all computational analysis of genomics data in the first submission and wrote the initial draft of the first submission. S.Y. performed all computational analysis for genomics and imaging data for two revisions. E.K.F. performed all FISH experiments and assisted in mouse breeding. S.C.N. and E.F.J. assisted in Oligopaint probe design. G.V. conceptualized and supervised the project, wrote the manuscript, revised it with comments from all authors and acquired funding.

## Competing interests

The authors declare no competing interests.

## Additional information

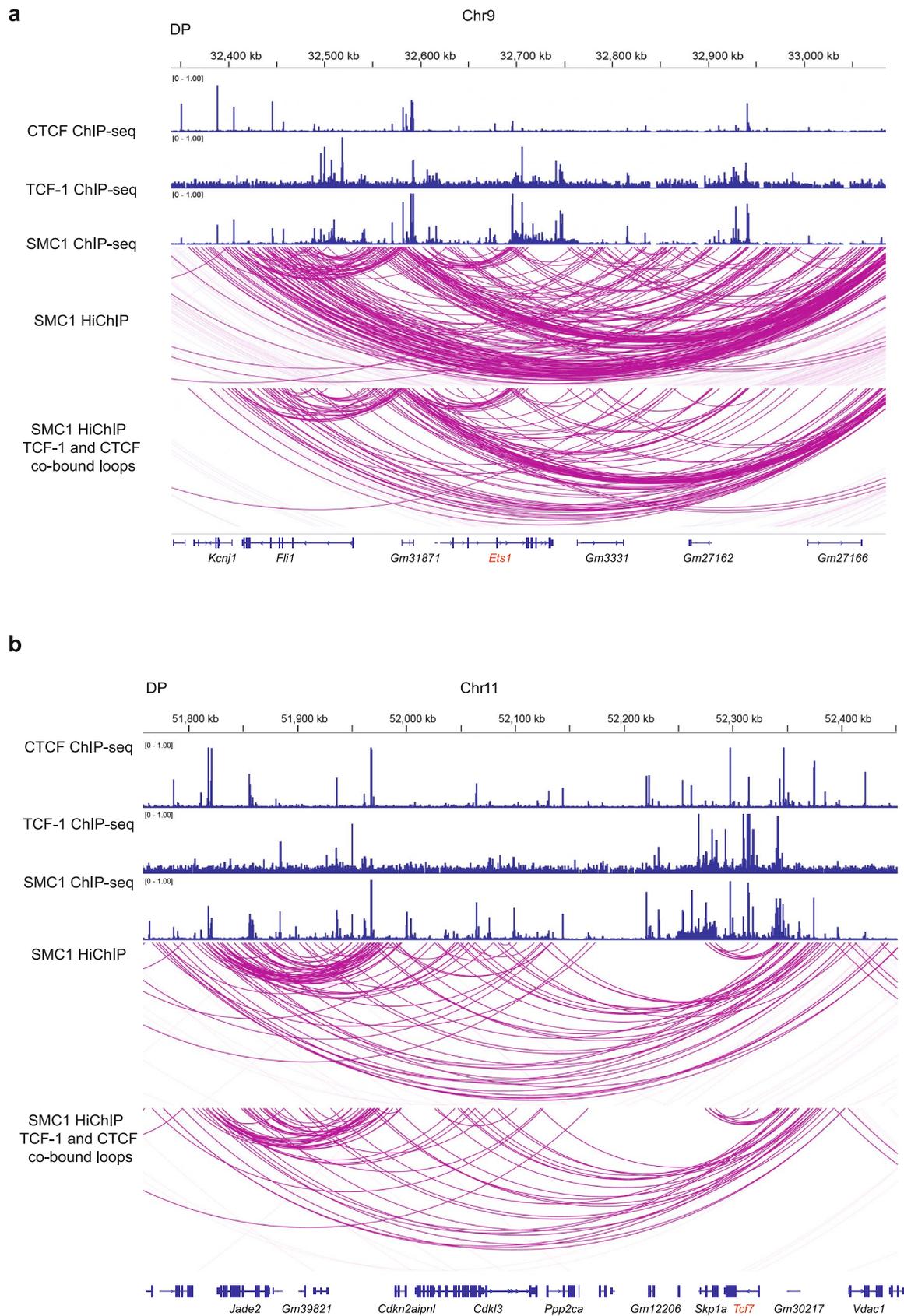
**Extended data** is available for this paper at <https://doi.org/10.1038/s41590-022-01232-z>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41590-022-01232-z>.

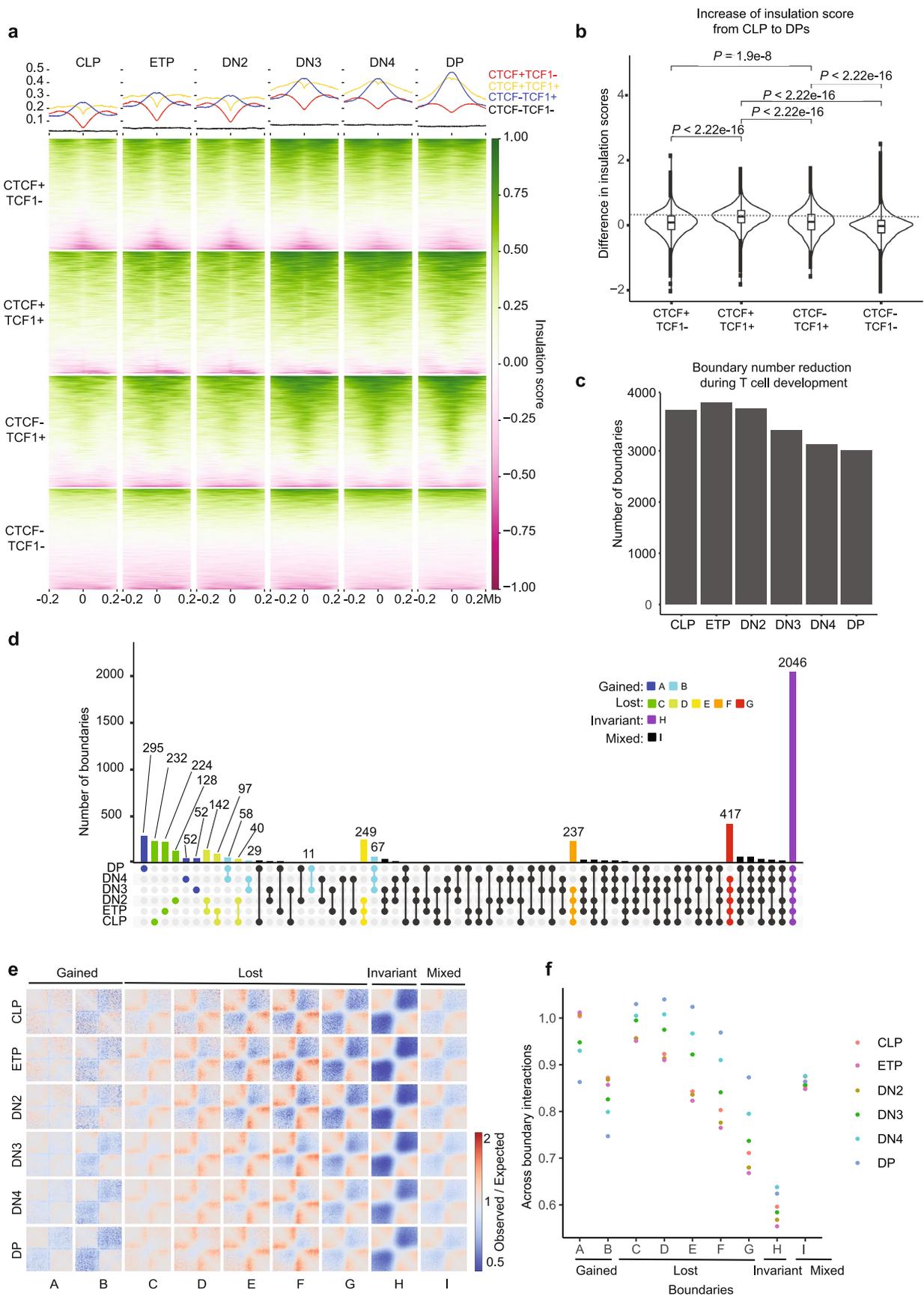
**Correspondence and requests for materials** should be addressed to Goltz Vahedi.

**Peer review information** *Nature Immunology* thanks the anonymous reviewers for their contribution to the peer review of this work. Ioana Visan was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

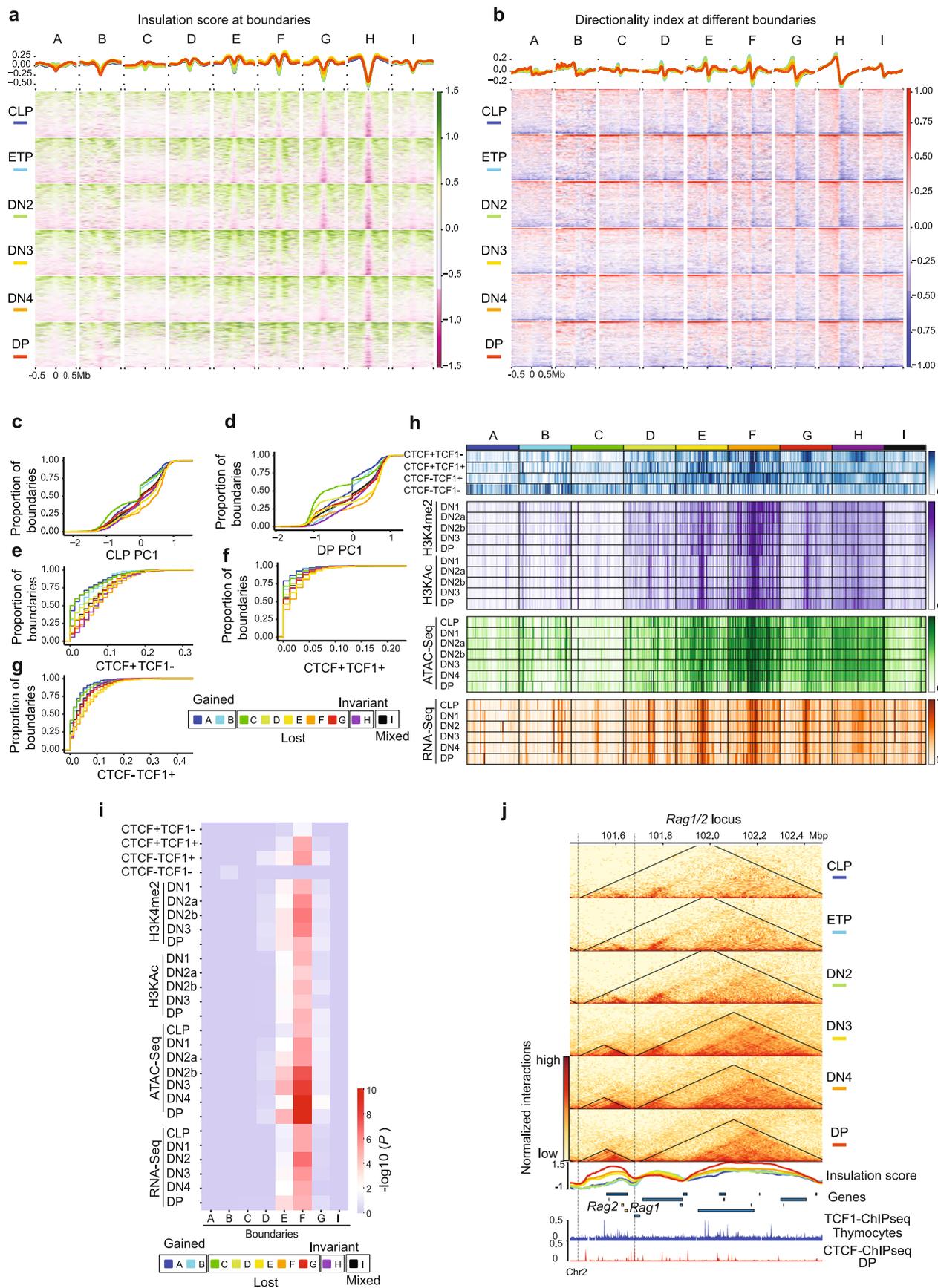


**Extended Data Fig. 1 | Loops associated with CTCF and TCF-1.** Genome browser view showing cohesin loops and CTCF, TCF-1, SMC1 binding at the *Ets1* (**a**) and *Tcf7* (**b**) loci. The loops associated with both TCF-1 and CTCF are shown in the bottom panel.



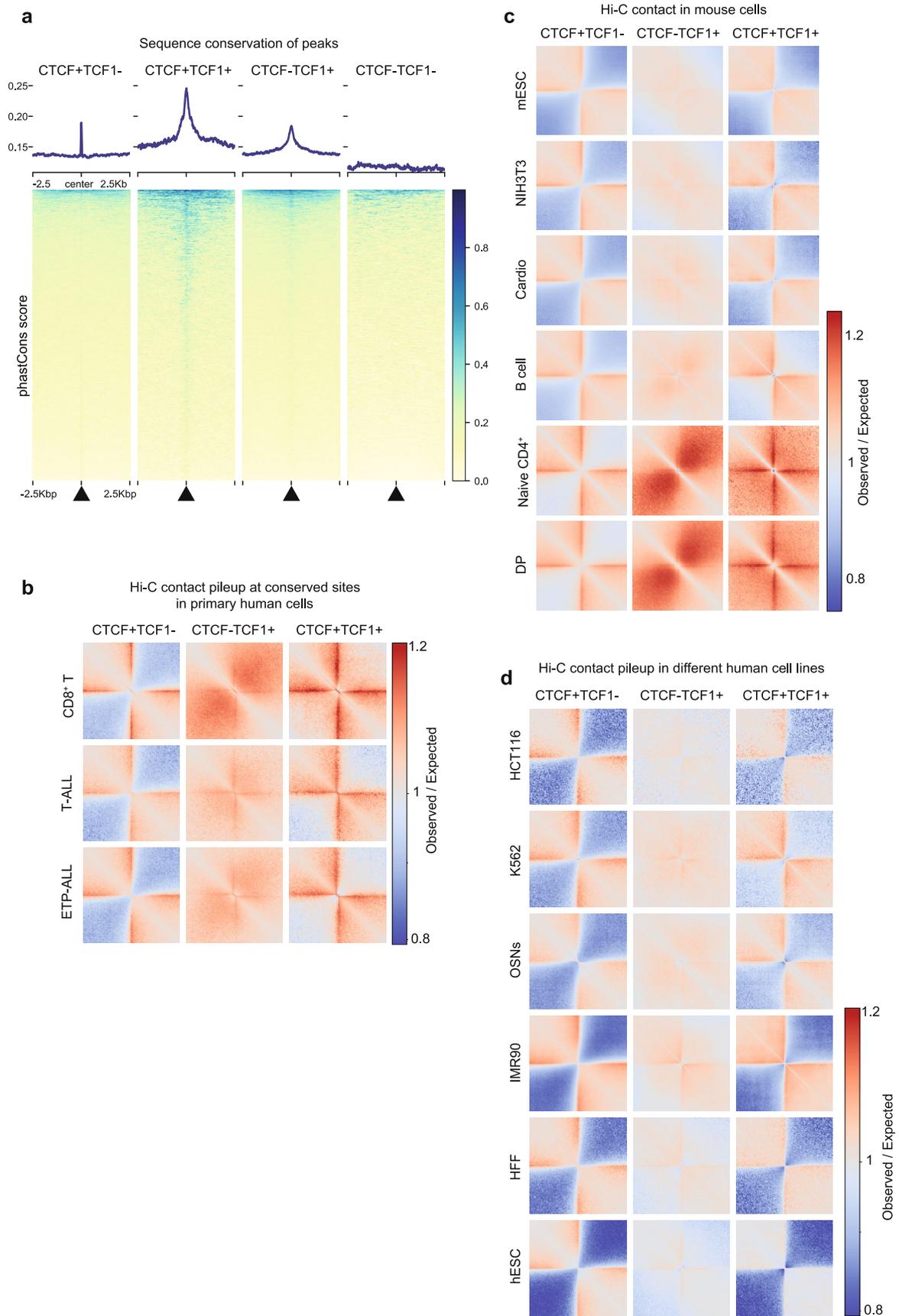
Extended Data Fig. 2 | See next page for caption.

**Extended Data Fig. 2 | Loss of insulation and boundaries during T cell development.** **a.** Summary plot and heat map of insulation score in six T cell developmental stages at CTCF only (red), CTCF+TCF-1 (yellow), TCF-1-only (blue) peaks and random regions as control (black), where lower value indicates higher insulation. Peaks defined in Fig. 1c. **b.** Violin plot and boxplot (center, median; box limits, upper (75th) and lower (25th) percentiles; whiskers, 1.5x interquartile range; points, outliers) showing differences in insulation score from CLP to DP at the four different sets of peaks. The comparison between different groups was done with one-sided Student's t-test. The insulation score change was significantly ( $P < 2.2 \times 10^{-16}$ ) different between different sets of peaks. **c.** Bar plot showing the total number of boundaries in six T cell developmental stages. **d.** Upset plot showing unique and shared boundaries among the six T cell developmental stages. The color shows different groups of boundaries that are lost or gained during T cell development. **e.** Local pileup in six T cell developmental stages at the nine groups of boundaries classified in (d). **f.** Quantification of interactions in six T cell developmental stages between upstream and downstream regions of the nine groups of boundaries.



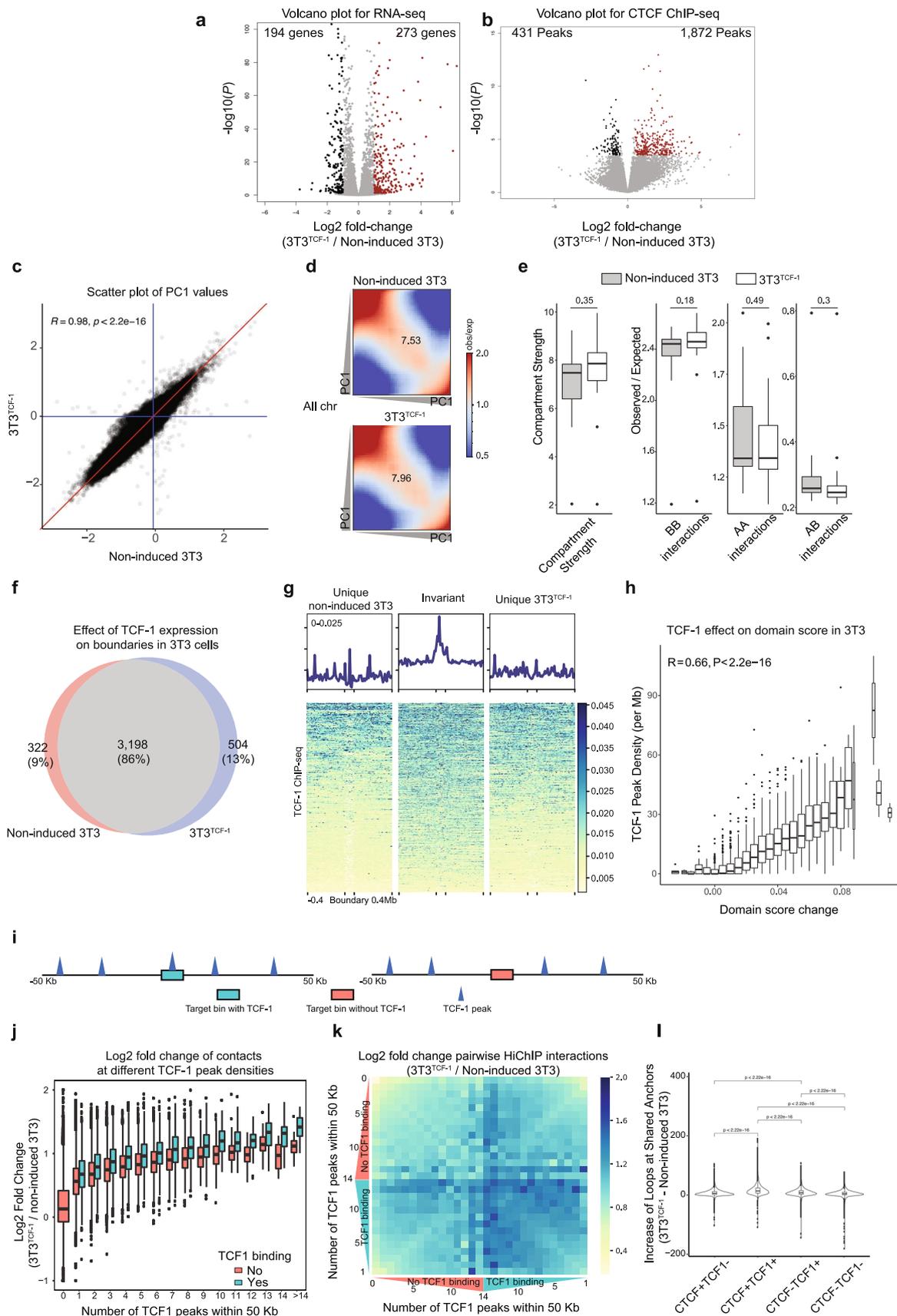
Extended Data Fig. 3 | See next page for caption.

**Extended Data Fig. 3 | Chromatin features at dynamic boundaries.** **a.** Summary plot and heat map of insulation score in 1 Mbp regions centered at the nine groups of boundaries in six T cell developmental stages. **b.** Summary plot and heat map of directionality index at the nine groups of boundaries in six T cell developmental stages. Cumulative distribution of PC1 values in CLP cells (**c**) and DP cells (**d**) at nine boundary groups. Cumulative distribution of CTCF only peak density (**e**), CTCF+TCF-1 peaks density (**f**) and TCF-1-only peak density (**g**) as defined in Fig. 1c at nine boundary groups **h.** Heatmap showing enrichment of CTCF only, CTCF+TCF-1, TCF-1 only peaks and random regions at nine boundary groups, as well as the H3K4me2, H3KAc, ATAC-seq and RNA-seq signals at T cell developmental stages. Signals in each row were normalized between 0 and 1. **i.** Heat map showing the statistical test of enrichment of epigenetic features at each group of boundaries compared with conserved boundaries using Kolmogorov–Smirnov test. **j.** Genome browser view and contact matrix at the *Rag1/2* locus. Contact matrix was normalized according to the sequencing depth and triangles show the identified TADs in this region. Insulation score depicted as a line plot and colored according to their developmental stage as indicated in the figure legend. Genes in this region are shown with *Rag1* and *Rag2* genes highlighted in orange.



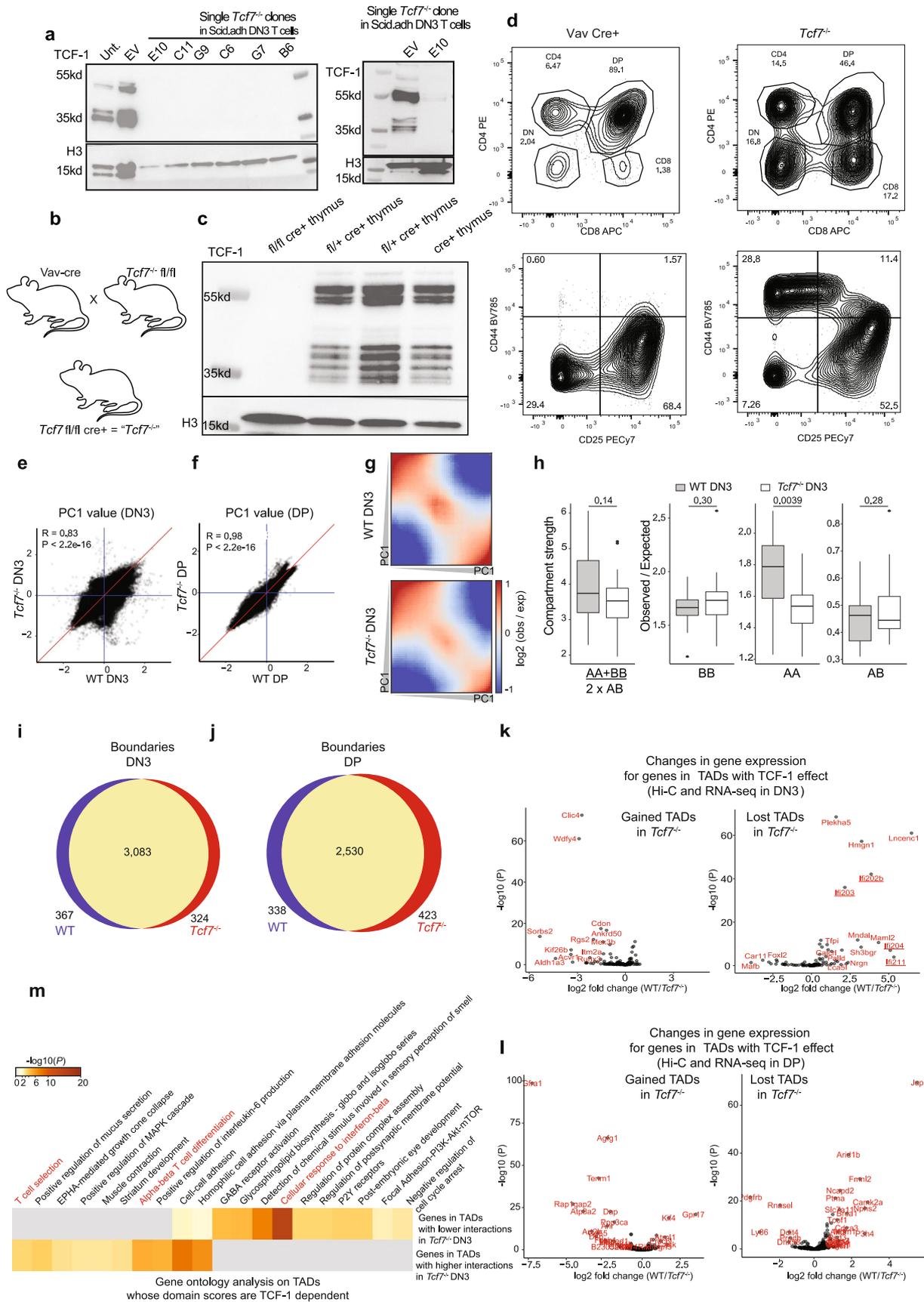
Extended Data Fig. 4 | See next page for caption.

**Extended Data Fig. 4 | TCF-1 and CTCF co-bound sites are evolutionarily conserved and insulated in non-T cells.** **a.** Summary plot and heat map of sequence conservation using phastCons score at CTCF only, CTCF+TCF-1, TCF-1-only peaks and random regions as control. **b.** Local pileup plot of interactions using Hi-C measurements in progenitor and mature T cells in humans at homologs CTCF and TCF-1 peaks. Hi-C measurements were from human publicly available T-ALL, ETP T-ALL and ultra-deep Hi-C in CD8<sup>+</sup> T cells from a healthy donor generated for this study. Human homologous regions for CTCF only, TCF-1 only as well as TCF-1+CTCF peaks in mice were found using liftOver. **c.** Local pileup plot of interactions at CTCF only, TCF-1+CTCF, TCF-1 only peaks in different mouse cell types, including mouse embryonic stem cell (mESC), fibroblast (NIH 3T3), cardiomyocyte, B cells, naïve CD4<sup>+</sup> and DP T cells. **d.** Local pileup plot of interactions in different human cell types from 4D Nucleome repository at the liftOver sites of CTCF only, TCF-1 only as well as CTCF+TCF-1 peaks from mice.



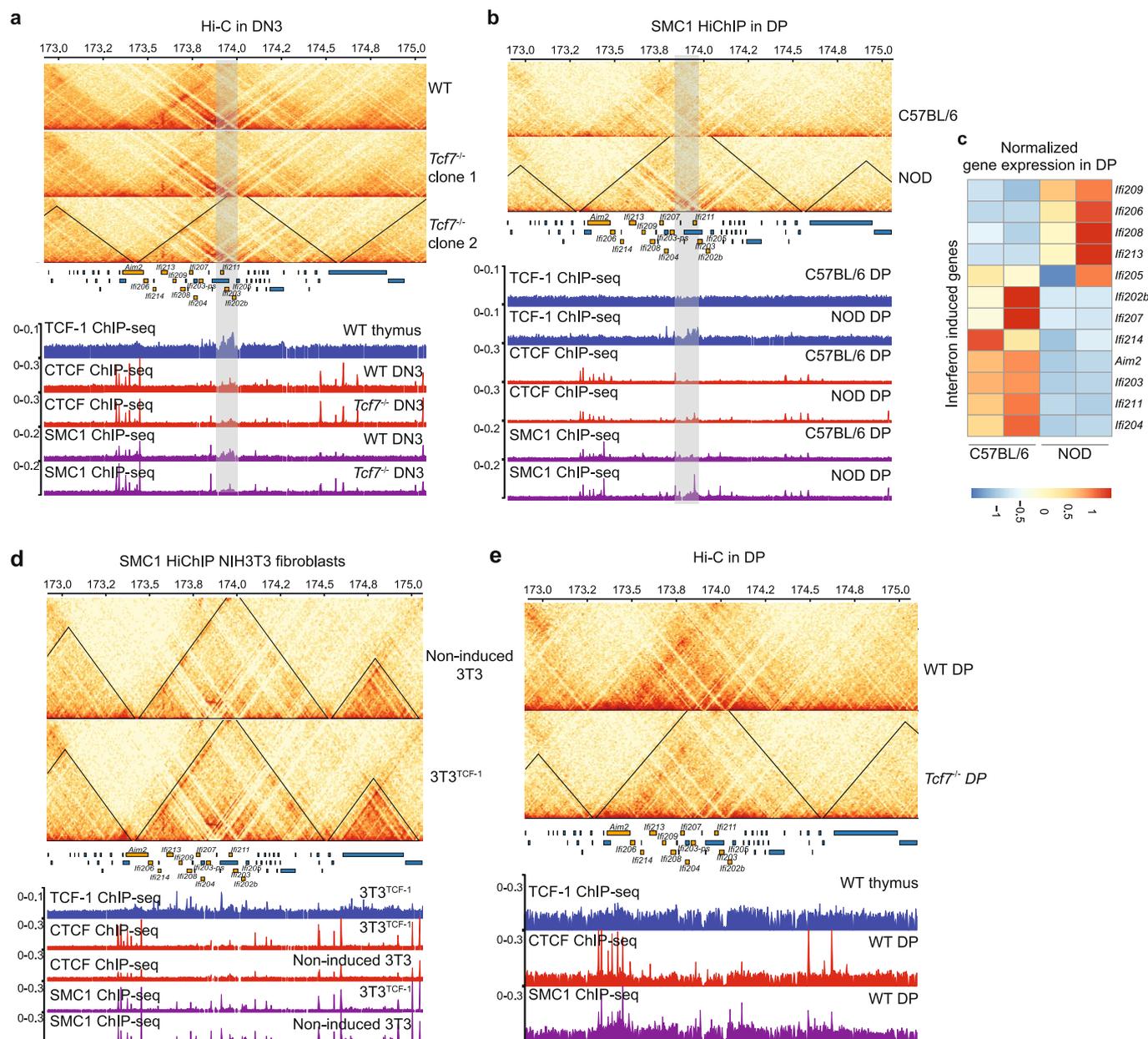
Extended Data Fig. 5 | See next page for caption.

**Extended Data Fig. 5 | Differential gene expression and CTCF binding after TCF-1 overexpression in fibroblast.** **a.** Volcano plot of differentially expressed genes in non-induced 3T3 and 3T3<sup>TCF-1</sup> fibroblasts. Log<sub>2</sub> fold change and *P* values are calculated with DESeq2. *P* depicts adjusted pvalue calculated by DESeq2. **b.** Volcano plot of differential CTCF binding in non-induced 3T3 and 3T3<sup>TCF-1</sup> fibroblasts. Log<sub>2</sub> fold change and *P* values are calculated with DESeq2. *P* depicts adjusted pvalue calculated by DESeq2. **c.** Scatter plot of PC1 values in non-induced 3T3 and 3T3<sup>TCF-1</sup> fibroblasts. The blue lines are where PC1 values are equal to zero and the red line is where PC1 values are equal in the two conditions. Pearson correlation and the two-sided significance level are shown, the PC1 values are significantly ( $P < 2.2 \times 10^{-16}$ ) correlated. **d.** Saddle plot showing the compartmentalization in non-induced 3T3 and 3T3<sup>TCF-1</sup> fibroblasts. The regions were sorted by PC1 value from B to A compartment and interactions between different regions are shown in the heat map. **e.** Boxplot (center, median; box limits, upper (75th) and lower (25th) percentiles; whiskers, 1.5x interquartile range; points, outliers) shows the compartment strength, BB interactions, AA interactions and AB interactions in wild-type and *Tcf7*<sup>-/-</sup> DN3s on all chromosomes (n=20). The statistical test was done using two-sided Student's t-test. (ns: not significant). **f.** Venn diagram showing the unique and shared boundaries non-induced 3T3 and 3T3<sup>TCF-1</sup> fibroblasts. **g.** Summary plot and heat map of TCF-1 ChIP-seq performed in NIH 3T3 cells at unique and common boundaries between and after TCF-1 induction. The three different sets of boundaries were normalized to the same height. **h.** Boxplots (center, median; box limits, upper (75th) and lower (25th) percentiles; whiskers, 1.5x interquartile range; points, outliers) showing the correlation of TCF-1 binding density in NIH 3T3 cells and increase in domain score after TCF-1 overexpression. TADs (n=2573) are grouped based on the increase of domain score and the boxplots show TCF-1 peak density in each group. *R* and *P* values are calculated with two-sided Pearson correlation. The results shows TCF-1 binding is significantly ( $P < 2.2 \times 10^{-16}$ ) associated with domain score increase. **i.** Illustration of our computational strategy to classify TCF-1-dependent 3D genome interactions originating from TCF-1 binding. Genomic regions were divided into bins and classified based on the number of TCF-1 peaks within 50 kb, as well as TCF-1 binding status in the bin. For every genomic bin, we therefore calculated the number of TCF-1 peaks in the 50 kbp neighborhood of the bin and additionally classified these bins into TCF-1 bound and unbound regions. **j.** Boxplots (center, median; box limits, upper (75th) and lower (25th) percentiles; whiskers, 1.5x interquartile range; points, outliers) show correlation of interactions at bins with different number of TCF-1 binding events within their 50 kb neighborhood. Each bin itself was further divided based on TCF-1 occupancy. **k.** The greatest increase in 3D interactions after TCF-1 expression in fibroblasts occurs when TCF-1 occupies both anchors and has high density at anchors' neighboring regions. We calculated the extent of increase in interactions between 10kbp bin pairs which were grouped based on the status of TCF-1 binding and the number of TCF-1 binding events in their 50kbp neighborhood. **l.** Violin plot and boxplot (center, median; box limits, upper (75th) and lower (25th) percentiles; whiskers, 1.5x interquartile range; points, outliers) showing the number of loops after TCF-1 overexpression in 3T3 at anchors that are bound by CTCF only, CTCF+TCF-1, TCF-1 only and random regions. The statistical test was done with two-sided Student's t-test. The results show that the gain of loops after TCF-1 overexpression are significantly ( $P < 2.2 \times 10^{-16}$ ) different between different sets of peaks.

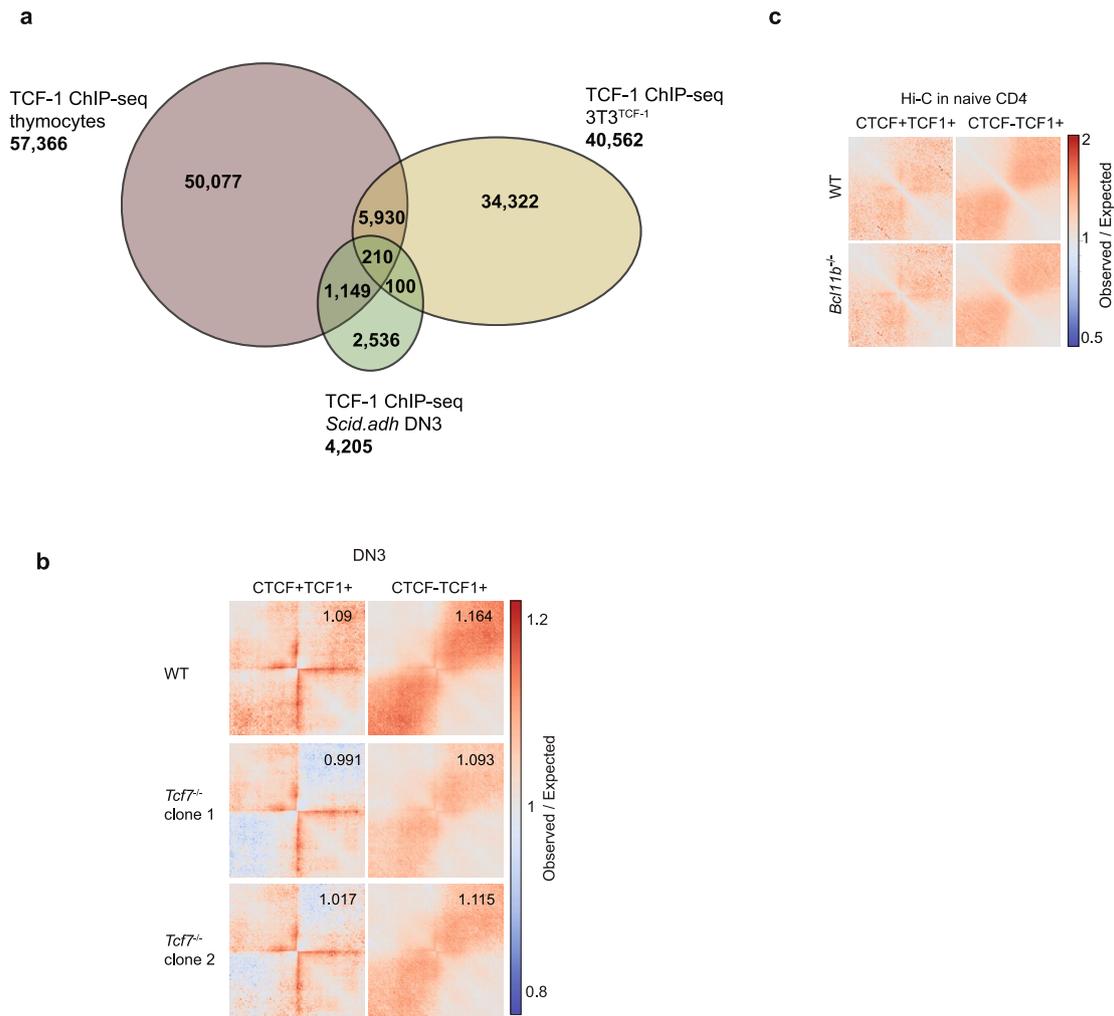


Extended Data Fig. 6 | See next page for caption.

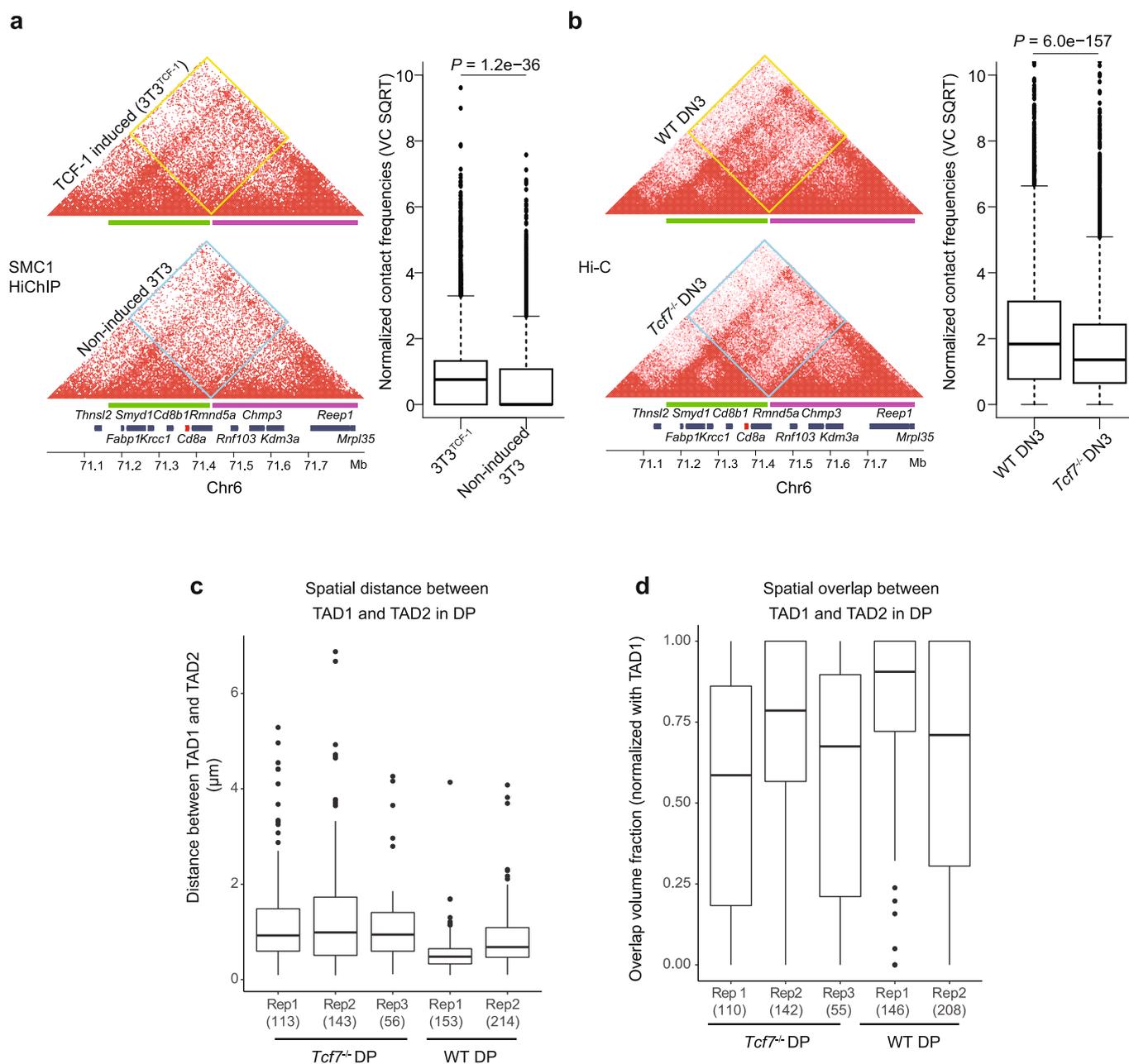
**Extended Data Fig. 6 | 3D genome reorganization after TCF-1 deletion in DN3s and DPs.** **a.** Western blot of TCF-1 in DN3s in which TCF-1 was disrupted using CRISPR/Cas9. Single cell clones in bold indicate those utilized for further experiments. Experiments were carried out twice and similar results were obtained. **b.** Illustration of breeding strategy for disruption of *Tcf7* in mice by crossing *Vav-Cre* with *Tcf7* floxed mice, generating mice in which TCF-1 is conditionally ablated in all hematopoietic cells. **c-d.** Western blot of TCF-1 in thymocytes from homozygous fl/fl cre+ (*Tcf7*<sup>-/-</sup>), heterozygous fl/+ cre+ (N=2) and cre+ mice. **(c)** Flow cytometric analysis of thymocytes from *Vav cre+* control and *Tcf7*<sup>-/-</sup> experimental mice. Top panel shows CD4<sup>+</sup> and CD8<sup>+</sup> populations and is pre-gated on lymphocytes, live and single cells. Bottom panel shows DN populations DN1-DN4 (upper left to lower left quadrant clockwise) as measured based on CD44 and CD25 expression and is pre-gated on lymphocytes, live, single cells and lineage negative cells **(d)**. **e-f.** Scatter plot showing A/B compartment distribution using correlation of PC1 values between wildtype and *Tcf7*<sup>-/-</sup> DN3s **(e)** and DPs **(f)**. Blue lines are PC1 equals zero and red line is where PC1 are equal in two conditions. Pearson correlation coefficients and *P*-values are shown. **g.** Saddle plot shows the compartmentalization in wildtype and *Tcf7*<sup>-/-</sup> DN3s. The regions were sorted by PC1 value from B to A compartment and interactions between different regions are shown in the heat map. One Hi-C experiment was performed in the wildtype condition and two Hi-C experiments were performed in two distinct *Tcf7*<sup>-/-</sup> DN3 clones. **h.** Data are shown as boxplots (center, median; box limits, upper (75th) and lower (25th) percentiles; whiskers, 1.5× interquartile range; points, outliers) showing the compartment strength, BB interactions, AA interactions and AB interactions in wildtype and *Tcf7*<sup>-/-</sup> DN3s (n=20 mouse chromosomes). The statistical test was done using two-sided Student's t-test. (ns: not significant, \*: *P* < 0.05). **i-j.** Venn diagram showing the unique and shared boundaries between wildtype and *Tcf7*<sup>-/-</sup> DN3s **(i)** and DP **(j)**. **k-l.** Scatter plot showing changes in gene expression located in the top 50 TADs that lost domain score in DN3 **(k)** and DP **(l)** T cells. Genes that were significantly (*P* < 0.05, abs(log<sub>2</sub>FoldChange) > 0.5) differentially expressed were shown in red. **m.** Gene-ontology analysis of genes in the top 50 TADs that gained domain score and the top 50 TADs that lost domain score after TCF-1 deletion in DN3s using metaScape. T cell-specific ontology terms are marked in red.



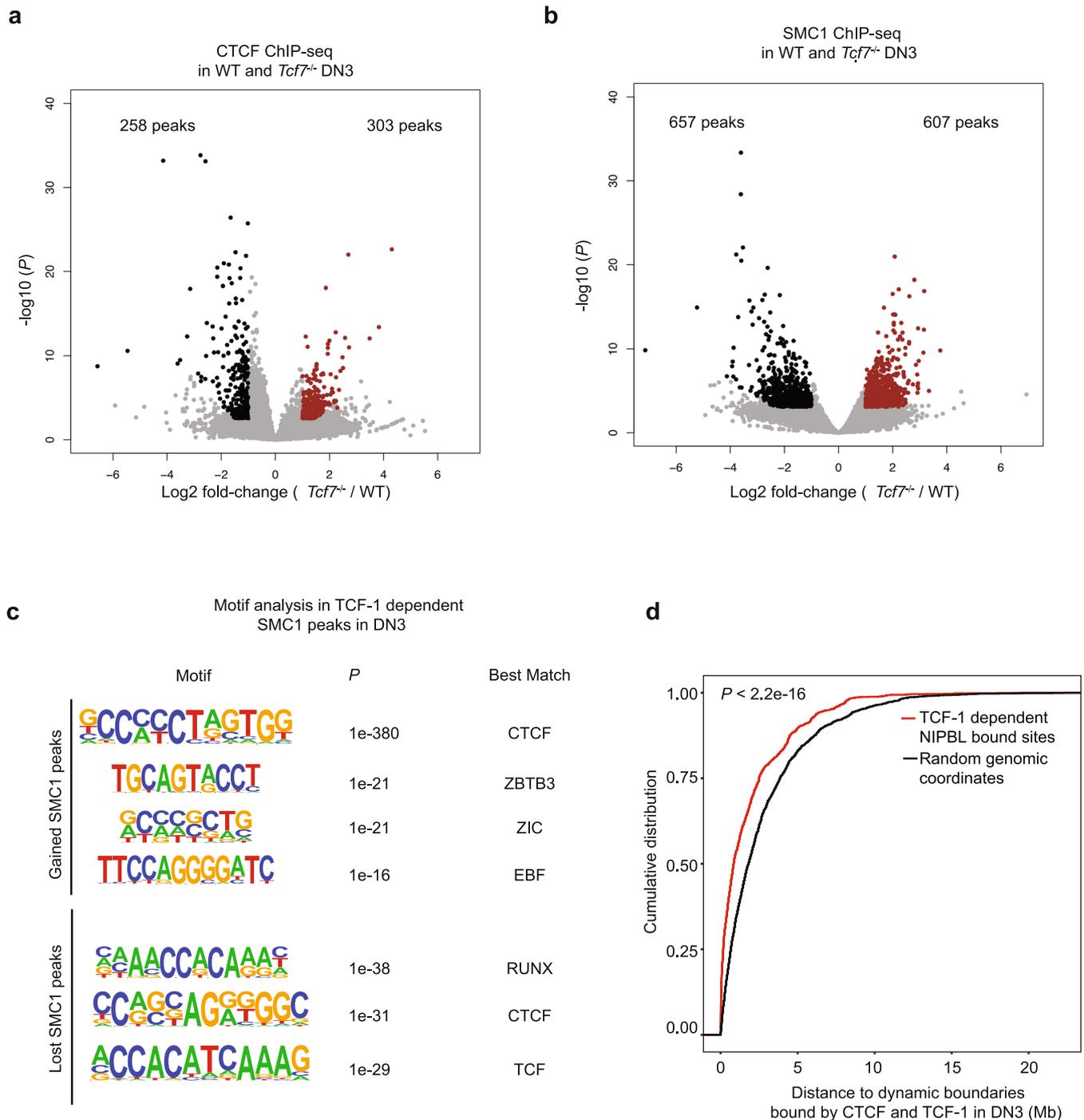
**Extended Data Fig. 7 | Reorganization of interferon genes by TCF-1. a,b,d,e.** Contact matrix and genome browser views of ultra-deep Hi-C measurements at interferon-induced gene cluster (genes marked in orange) using two independent clones for deleting TCF-1 using CRISPR/Cas9 in DN3 (**a**), comparing SMC1 HiChIP in DPs of C57BL/6 and NOD mice (**b**). We used Stripenn to quantify stripiness of the stripe originating from the TCF1 cluster, which showed that NOD has a stripe score of 2.14, while C57BL/6 is -2.3. SMC1 HiChIP in TCF-1 induction in fibroblasts (**d**) and ultra-deep Hi-C in wildtype and *Tcf7<sup>-/-</sup>* DPs (**e**). The browser view focuses on a TAD that lose intra-TAD interactions after TCF-1 deletion in both replicates, which also have a cluster of TCF-1 binding (gray highlighted) in wildtype DN3s. The CTCF and SMC1 binding in both conditions are also shown. **c.** Heat map showing the row normalized gene expression level from DPs in NOD and C57BL/6 of the interferon induced genes in the TAD with different intra-TAD interactions in C57BL/6 and NOD mice. Orange bars represent genomic coordinates of interferon genes.



**Extended Data Fig. 8 | TCF-1 ChIP-seq comparison in DPs and DN3s. a.** Venn diagram of overlapping and unique bound regions by TCF-1 ChIP-seq thymocytes, DN3s and 3T3<sup>TCF-1</sup>. **b.** Local pileup plot of long-range interactions using ultra-deep Hi-C in wildtype and two *Tcf7*<sup>-/-</sup> DN3 clones at a subset of CTCF+TCF-1 co-bound, as well as TCF-1-only peaks that gain insulation in *Tcf7*<sup>-/-</sup> DN3s. The two classes of peaks were defined using TCF-1 and CTCF ChIP-seq in thymocytes and DPs, respectively as in Fig. 1. These peaks were further selected based on the decrease of average interactions between the upstream and downstream of the peaks by at least 0.05 after TCF-1 deletion. The numbers are the average observed/expected interactions in the upper right square, which is the interactions between the upstream and downstream of the peaks. **c.** Local pileup interactions in wildtype and *Bcl11b* deficient naïve CD4 T<sup>+</sup> cells at the CTCF and TCF-1 co-bound, as well as TCF-1 only peaks.



**Extended Data Fig. 9 | Oligopaint 3D FISH corroborates the role of TCF-1 on chromatin interactions at the *Cd8* locus. a-b.** Contact matrix plot of Hi-C data sets at the *Cd8a-Cd8b1* locus in NIH3T3 fibroblasts (**a**) and DN3s (**b**) ( $n=7221$  genomic interactions). Inter-domain interactions were quantified by boxplots. Two sided Paired wilcoxon ranksum test was performed. Oligopaint probes for TAD1 (green) and TAD2 (magenta) are depicted. Two-sided Wilcoxon rank-sum test  $P$ -values are shown for boxplots. **c-d.** Box plot showing spatial distance (**c**) and spatial overlap (**d**) between TAD1 and TAD2 compared between wildtype ( $N=2$ ) and TCF-1-deficient ( $N=3$ ) of DPs in biological mouse replicates. Spatial distance and overlap were calculated as in (Fig. 5d, e). Number of cells used for imaging analysis per mouse is depicted in parenthesis. **a-d** Data are shown as boxplots (center, median; box limits, upper (75th) and lower (25th) percentiles; whiskers, 1.5x interquartile range; points, outliers).



**Extended Data Fig. 10 | Recruitment of cohesin at TCF-1 binding sites.** **a.** Volcano plot shows the differential CTCF binding in wildtype and *Tcf7*<sup>-/-</sup> DN3s. Red indicates stronger binding after TCF-1 deletion. Log<sub>2</sub> fold change and *P* values are calculated with DESeq2. *P* depicts adjusted pvalue calculated by DESeq2. **b.** Volcano plot showing the differential SMC1 binding in wildtype and *Tcf7*<sup>-/-</sup> DN3s. Log<sub>2</sub> fold change and *P* values were calculated with DESeq2. *P* depicts adjusted pvalue calculated by DESeq2. Red indicates stronger binding after TCF-1 deletion. **c.** Enriched motifs from HOMER at the gained or lost SMC1 peaks after TCF-1 deletion in DN3 cells using random background. The most significantly enriched motifs and associated *P* values are shown. *P* values were calculated using hypergeometric test. **d.** Cumulative distribution of genomic distance between 2,042 TCF-1 dependent SMC1 and NIPBL co-bound enhancers or 2,042 randomly selected genomic regions to 988 dynamic boundaries with TCF-1 and CTCF co-binding. Kolmogorov-Smirnov test *P*-value is shown for cumulative distribution plot.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

1. In the analysis of HiC/HiChIP data, HiC-Pro (version 2.11.1) was used to process the HiC and HiChIP data and generate ValidPairs for each sample. Then these ValidPairs files were further converted to .hic and .cool files. Juicebox (version 1.11.08) was used to visualize the contact maps and cooltools (version 0.4.0) was used to do compartment and TADs level analysis. juicer\_tools (version 1.13.02) was used to call the loops from HiC data, while hichipper (version 0.7.8b0) was used to identify the loops in HiChIP data.
2. In the analysis of ChIP-seq data, bowtie2 (version 2.3.4.1) was used for alignment. Reads aligned to the mitochondrial genome or chrY as well as reads mapped to multiple genomic loci were discarded from downstream analyses. Bigwig files were generated by bedtools (version v2.27.1) genomcov and wigToBigWig normalizing tracks to tags-per-million. For peak calling, macs2 (version 2.2.6) was used. Input sample for each strain was prepared by the same approach without immunoprecipitation and used as input control for peak calling. DESeq2 (version 1.30.1) was used to compare the binding strength in two conditions.
3. In the analysis of RNA-seq data, RNA-seq samples were aligned by STAR (version 2.5.0a\_alpha). HTSeq (version v0.6.1) facilitated counting RNA-seq reads on Gencode vM11 gene models. DESeq2 was subsequently applied on gene counts to identify genes differentially expressed.
4. We used HiExplorer (version 3.6) to visualize the interactions, insulation score, genes and ChIP-seq signals of selected regions.
5. R version 3.6.2 was used to generate boxplot, scatter plot, ecdf plot, bar plot and heatmap (ggplot2 version 3.3.3)
6. python 3.7 and packages including matplotlib, seaborn were used to plot venn diagramm (matplotlib\_venn) and heatmaps.
7. Coolpup.py (version 0.9.1) was used to generate 3D pileup plots.
8. The pileup analysis of 1D features including insulation score, directional index, ChIP-seq and phastCons score was done with deeptools.
9. Metascape was used for gene set enrichment analysis. (<https://metascape.org/gp/index.html#>)
10. FlowJo (version 10.6.1) was used analysis flow cytometric data.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

### 1. Publicly available data

- HiC data during T cell development	GSE79422	
- TCF-1 ChIP-seq in thymus	GSE46662	
- CTCF ChIPseq in DP T cells	GSE141853	
- Smc1 HiChIP in DP T cells	GSE141853	
- Histone modifications during T cell development		GSE31235
- ATAC-seq during T cell development		
- RNA-seq during T cell development	GSE79422	
- Mouse B cell hic	4DNFISA93XFU	
- Mouse cardiomyocyte hic	GSE96693	
- Mouse mESC hic	4DNFIC21MG3U	
- Human hESC hic	4DNFI7JNCNFB	
- Human IMR90 hic	4DNFIJTOIGOI	
- Human olfactory sensory neurons hic	4DNFI7V61PXA	
- Human HCT116 hic	4DNFIV3PIEQJ	
- Human K562 hic	4DNFI244AS29	
- Human HFF hic	4DNFIWMLVHWW	

### 2. List of figures that have associated raw data

- Extended Data Figure : 7a, 7c

### 3. Data generated in this study

- HiC, HiChIP, ChIPseq, RNAseq GSE178348

The data generated in this study is publicly available on GEO: GSE178348

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample sizes of two biological replicates demonstrated the differences between experimental groups and were reproducible and significant, No statistical method was used to determine sample size.
Data exclusions	No data were excluded from the analyses.
Replication	All experiments were independently replicated at least twice.
Randomization	No experiment in this study required randomization.
Blinding	No experiment in this study required blinding.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials &amp; experimental systems

## Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

## Antibodies used

anti-CTCF (Millipore; Cat# 07-729; RRID:AB\_441965) 1:100  
 anti-SMC1 (Bethyl Laboratories; Cat# A300-055A; RRID:AB\_2192467) 1:100  
 anti-TCF1 (CST; Cat# 2206S; RRID:AB\_2199300) 1:10  
 normal Rabbit IgG (CST; Cat# #2729; RRID:AB\_1031062) 1:100  
 anti-NIPBL (Bethyl Laboratories; A301-779A; RRID:AB\_1211232) 1:100  
 Biotin anti-mouse CD8a (53-6.7) (BioLegend; Cat# 100703, RRID:AB\_312742) 1:200  
 Histone H3 (acetyl K27) antibody;Abcam;ab4729; RRID:AB\_2118291 1:100  
 Anti-Histone H3 Antibody, Unconjugated;CST;Cat# 9715, RRID:AB\_331563 1:2000  
 PE anti-mouse CD4 (RM4-4);BioLegend;Cat# 116005; RRID:AB\_313690 1:200  
 APC anti-mouse CD8 (53-6.7);BioLegend;Cat# 100711; RRID:AB\_312750 1:200  
 PeCy7 anti-mouse CD25 (PC61);BioLegend;Cat# 102015; RRID:AB\_312864 1:350  
 Brilliant Violet 785™ anti-mouse/human CD44 (IM7);BioLegend;Cat# 103041; RRID:AB\_11218802 1:400  
 APC Streptavidin;BioLegend;Cat# 405207 1:200  
 Biotin anti-mouse Ly-6G/Ly-6C (Gr-1) (RB6-8C5);BioLegend;Cat# 108403; RRID:AB\_313368 1:200  
 Biotin anti-mouse NK1.1 (PK136);BioLegend;Cat# 108703; RRID:AB\_313390 1:200  
 Biotin anti-mouse CD11b (M1/70);BioLegend;Cat# 101203, RRID:AB\_312786 1:200  
 Biotin anti-mouse Ter119 (TER-119);BioLegend;Cat# 116203, RRID:AB\_313704 1:200  
 Biotin anti-mouse B220 (RA3-6B2);BioLegend;Cat# 103203, RRID:AB\_312988 1:200  
 Anti-rabbit IgG, HRP-linked Antibody;CST;Cat# 7074, RRID:AB\_2099233 1:2000

## Validation

All antibodies have been validated by their manufacturer and in citations by other studies. For flow cytometric antibodies from BioLegend: "Specificity testing of 1-3 target cell types with either single- or multi-color analysis (including positive and negative cell types)." Anti-Smc1 and anti-NIPBL antibody have been validated by the manufacturer Bethyl Laboratories: "All Bethyl Laboratories® antibodies are validated to meet our strict performance standards." The Anti-CTCF antibody is validated by the manufacturer: MilliporeSigma: "validated for use in immunostaining, blotting, purification, ChIP, flow cytometry and multiplexed assays." The anti-TCF1 antibody was validated for species reactivity in mouse and human by the manufacturer and is utilized in ChIP-seq experiments in multiple other studies.

## Eukaryotic cell lines

Policy information about [cell lines](#)

## Cell line source(s)

NIH 3T3 - ATCC  
 HEK293T - ATCC  
 Scid.adh - gift from Dr. Warren Pear's lab

## Authentication

Cell lines were not authenticated.

## Mycoplasma contamination

Mycoplasma contamination were tested periodically in all cell lines, no mycoplasma contamination was detected.

Commonly misidentified lines  
(See [ICLAC](#) register)

Commonly misidentified cell lines were not used.

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

## Laboratory animals

Vav-iCre transgenic mice (Strain #008610) and Tcf7eGFP Mice (strain # 030909) were purchased from Jackson Laboratory. "Tcf7-/-" mice were generated by breeding Tcf7eGFP with Vav-iCre mice. The F1 generation was backcrossed to Tcf7eGFP mice to reach homozygous floxed Cre+ experimental mice. Vav-iCre+ control mice were generated by crossing Vav-iCre mice with C57BL6/J (strain # #000664) and used as controls in all experiments. Experimental and control mice were 6-10 weeks old of either sex.

## Wild animals

The study did not involve any wild animals.

## Field-collected samples

The study did not involve any field-collected samples.

## Ethics oversight

All mice were bred and housed in an American Association for the Accreditation of Laboratory Animal Care (AAALAC) accredited vivarium at the University of Pennsylvania. All husbandry and experimental procedures were performed according to the protocol reviewed and approved by the Institutional Animal Care and Use Committee (IACUC).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## ChIP-seq

## Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

## Data access links

May remain private before publication.

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE178348>

## Files in database submission

Norm3T3.CTCF.peaks.rep1.bed  
 Norm3T3.CTCF.peaks.rep1.bw  
 Norm3T3.CTCF.peaks.rep2.bed  
 Norm3T3.CTCF.peaks.rep2.bw  
 Norm3T3.Smc1.peaks.bed  
 Norm3T3.Smc1.peaks.bw  
 TCF13T3.CTCF.peaks.rep1.bed  
 TCF13T3.CTCF.peaks.rep1.bw  
 TCF13T3.CTCF.peaks.rep2.bed  
 TCF13T3.CTCF.peaks.rep2.bw  
 TCF13T3.Smc1.peaks.bed  
 TCF13T3.Smc1.peaks.bw  
 TCF13T3.TCF1.peaks.bed  
 TCF13T3.TCF1.peaks.bw  
 DN3EV.CTCF.peaks.rep1.bed  
 DN3EV.CTCF.peaks.rep1.bw  
 DN3EV.CTCF.peaks.rep2.bed  
 DN3EV.CTCF.peaks.rep2.bw  
 DN3EV.Smc1.peaks.rep1.bed  
 DN3EV.Smc1.peaks.rep1.bw  
 DN3EV.Smc1.peaks.rep2.bed  
 DN3EV.Smc1.peaks.rep2.bw  
 DN3EV.TCF1.peaks.rep1.bed  
 DN3EV.TCF1.peaks.rep1.bw  
 DN3EV.TCF1.peaks.rep2.bed  
 DN3EV.TCF1.peaks.rep2.bw  
 DN3KO.CTCF.peaks.rep1.bed  
 DN3KO.CTCF.peaks.rep1.bw  
 DN3KO.CTCF.peaks.rep2.bed  
 DN3KO.CTCF.peaks.rep2.bw  
 DN3KO.Smc1.peaks.rep1.bed  
 DN3KO.Smc1.peaks.rep1.bw  
 DN3KO.Smc1.peaks.rep2.bed  
 DN3KO.Smc1.peaks.rep2.bw

Genome browser session  
(e.g. [UCSC](#))

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

## Methodology

## Replicates

2

## Sequencing depth

Experiment	Total number of reads	Uniquely mapped reads	Length	Paired/Single
Normal NIH-3T3 cells - CTCF Rep1	25,648,526	15,711,761	38	Paired
Normal NIH-3T3 cells - CTCF Rep2	23,351,022	14,216,621	38	Paired
TCF1 induced NIH-3T3 cells - CTCF Rep1	29,290,524	14,801,112	38	Paired
TCF1 induced NIH-3T3 cells - CTCF Rep2	23,216,006	140,40,781	38	Paired
Normal NIH-3T3 cells - Smc1	50,727,576	36,476,045	38	Paired
TCF1 induced NIH-3T3 cells - Smc1	26,467,091	20,266,340	38	Paired
TCF1 induced NIH-3T3 cells - TCF1	38,904,421	23,410,074	38	Paired
Wild type DN3 cells - CTCF Rep1	14,331,832	9,032,356	38	Paired
Wild type DN3 cells - CTCF Rep2	17,194,919	10,945,053	38	Paired
TCF1 KO DN3 cells - CTCF Rep1	24,244,742	15,443,950	38	Paired
TCF1 KO DN3 cells - CTCF Rep2	12,788,342	8,256,537	38	Paired
Wild type DN3 cells - Smc1 Rep1	34,835,832	23,121,630	38	Paired
Wild type DN3 cells - Smc1 Rep2	32,481,963	21,461,530	38	Paired
TCF1 KO DN3 cells - Smc1 Rep1	33,872,606	23,124,663	38	Paired

	TCF1 KO DN3 cells - Smc1 Rep2	31,967,993	21,689,990	38	Paired																																				
	Wild type DN3 cells - TCF1 Rep1	32,965,771	19,211,104	38	Paired																																				
	Wild type DN3 cells - TCF1 Rep2	34,854,961	19,831,232	38	Paired																																				
Antibodies	anti-CTCF (Millipore; Cat# 07-729; RRID:AB_441965) 1:100 anti-SMC1 (Bethyl Laboratories; Cat# A300-055A; RRID:AB_2192467) 1:100 anti-TCF1 (CST; Cat# 2206S; RRID:AB_2199300) 1:10 anti-NIPBL (Bethyl Laboratories; A301-779A; RRID:AB_1211232) 1:100																																								
Peak calling parameters	macs2 callpeak -c inputfile--nolambda--nomodel--keep-dup all -p 0.00001																																								
Data quality	<table border="1"> <thead> <tr> <th>Experiment</th> <th>Total number of peaks</th> </tr> </thead> <tbody> <tr> <td>Normal NIH-3T3 cells - CTCF Rep1</td> <td>38,369</td> </tr> <tr> <td>Normal NIH-3T3 cells - CTCF Rep2</td> <td>41,236</td> </tr> <tr> <td>TCF1 induced NIH-3T3 cells - CTCF Rep1</td> <td>68,306</td> </tr> <tr> <td>TCF1 induced NIH-3T3 cells - CTCF Rep2</td> <td>64,645</td> </tr> <tr> <td>Normal NIH-3T3 cells - Smc1</td> <td>105,033</td> </tr> <tr> <td>TCF1 induced NIH-3T3 cells - Smc1</td> <td>95,293</td> </tr> <tr> <td>TCF1 induced NIH-3T3 cells - TCF1</td> <td>39,489</td> </tr> <tr> <td>Wild type DN3 cells - CTCF Rep1</td> <td>42,214</td> </tr> <tr> <td>Wild type DN3 cells - CTCF Rep2</td> <td>53,729</td> </tr> <tr> <td>TCF1 KO DN3 cells - CTCF Rep1</td> <td>62,353</td> </tr> <tr> <td>TCF1 KO DN3 cells - CTCF Rep2</td> <td>55,149</td> </tr> <tr> <td>Wild type DN3 cells - Smc1 Rep1</td> <td>70,438</td> </tr> <tr> <td>Wild type DN3 cells - Smc1 Rep2</td> <td>77,587</td> </tr> <tr> <td>TCF1 KO DN3 cells - Smc1 Rep1</td> <td>88,886</td> </tr> <tr> <td>TCF1 KO DN3 cells - Smc1 Rep2</td> <td>96,416</td> </tr> <tr> <td>Wild type DN3 cells - TCF1 Rep1</td> <td>4,978</td> </tr> <tr> <td>Wild type DN3 cells - TCF1 Rep2</td> <td>5,356</td> </tr> </tbody> </table>					Experiment	Total number of peaks	Normal NIH-3T3 cells - CTCF Rep1	38,369	Normal NIH-3T3 cells - CTCF Rep2	41,236	TCF1 induced NIH-3T3 cells - CTCF Rep1	68,306	TCF1 induced NIH-3T3 cells - CTCF Rep2	64,645	Normal NIH-3T3 cells - Smc1	105,033	TCF1 induced NIH-3T3 cells - Smc1	95,293	TCF1 induced NIH-3T3 cells - TCF1	39,489	Wild type DN3 cells - CTCF Rep1	42,214	Wild type DN3 cells - CTCF Rep2	53,729	TCF1 KO DN3 cells - CTCF Rep1	62,353	TCF1 KO DN3 cells - CTCF Rep2	55,149	Wild type DN3 cells - Smc1 Rep1	70,438	Wild type DN3 cells - Smc1 Rep2	77,587	TCF1 KO DN3 cells - Smc1 Rep1	88,886	TCF1 KO DN3 cells - Smc1 Rep2	96,416	Wild type DN3 cells - TCF1 Rep1	4,978	Wild type DN3 cells - TCF1 Rep2	5,356
Experiment	Total number of peaks																																								
Normal NIH-3T3 cells - CTCF Rep1	38,369																																								
Normal NIH-3T3 cells - CTCF Rep2	41,236																																								
TCF1 induced NIH-3T3 cells - CTCF Rep1	68,306																																								
TCF1 induced NIH-3T3 cells - CTCF Rep2	64,645																																								
Normal NIH-3T3 cells - Smc1	105,033																																								
TCF1 induced NIH-3T3 cells - Smc1	95,293																																								
TCF1 induced NIH-3T3 cells - TCF1	39,489																																								
Wild type DN3 cells - CTCF Rep1	42,214																																								
Wild type DN3 cells - CTCF Rep2	53,729																																								
TCF1 KO DN3 cells - CTCF Rep1	62,353																																								
TCF1 KO DN3 cells - CTCF Rep2	55,149																																								
Wild type DN3 cells - Smc1 Rep1	70,438																																								
Wild type DN3 cells - Smc1 Rep2	77,587																																								
TCF1 KO DN3 cells - Smc1 Rep1	88,886																																								
TCF1 KO DN3 cells - Smc1 Rep2	96,416																																								
Wild type DN3 cells - TCF1 Rep1	4,978																																								
Wild type DN3 cells - TCF1 Rep2	5,356																																								
Software	Bowtie2(version 2.3.4.1), macs2(version 2.2.6)																																								