



OPEN

## Multi-level remodelling of chromatin underlying activation of human T cells

Naiara G. Bediaga<sup>1,2,4</sup>, Hannah D. Coughlan<sup>1,2,4</sup>, Timothy M. Johanson<sup>1,2</sup>, Alexandra L. Garnham<sup>1,2</sup>, Gaetano Naselli<sup>1</sup>, Jan Schröder<sup>1,2</sup>, Liam G. Fearnley<sup>1,2</sup>, Esther Bandala-Sanchez<sup>1,2</sup>, Rhys S. Allan<sup>1,2</sup>, Gordon K. Smyth<sup>1,3</sup> & Leonard C. Harrison<sup>1,2</sup>✉

Remodelling of chromatin architecture is known to regulate gene expression and has been well characterized in cell lineage development but less so in response to cell perturbation. Activation of T cells, which triggers extensive changes in transcriptional programs, serves as an instructive model to elucidate how changes in chromatin architecture orchestrate gene expression in response to cell perturbation. To characterize coordinate changes at different levels of chromatin architecture, we analyzed chromatin accessibility, chromosome conformation and gene expression in activated human T cells. T cell activation was characterized by widespread changes in chromatin accessibility and interactions that were shared between activated CD4<sup>+</sup> and CD8<sup>+</sup> T cells, and with the formation of active regulatory regions associated with transcription factors relevant to T cell biology. Chromatin interactions that increased and decreased were coupled, respectively, with up- and down-regulation of corresponding target genes. Furthermore, activation was associated with disruption of long-range chromatin interactions and with partitioning of topologically associating domains (TADs) and remodelling of their TAD boundaries. Newly formed/strengthened TAD boundaries were associated with higher nucleosome occupancy and lower accessibility, linking changes in lower and higher order chromatin architecture. T cell activation exemplifies coordinate multi-level remodelling of chromatin underlying gene transcription.

Mammalian genomes are folded into highly organized hierarchical structures linked to function at each level<sup>1</sup>. At the primary level of chromatin structure, the nucleosome, a 147 base-pair DNA segment wrapped around an octamer of histone proteins, directly influences gene expression by dictating access of DNA to the transcriptional machinery<sup>2-4</sup>. At an intermediate level, the genome is organized into protein-mediated loops that facilitate 3-dimensional (3D) interactions between pairs of genomic sites such as promoters and enhancers distant within the genome. At a higher level, the genome is organized into self-interacting chromatin, called topologically associating domains (TADs)<sup>5,6</sup>. Disruption of TAD boundaries has been associated with developmental defects<sup>7</sup> but the functional significance of changes in TAD architecture are otherwise largely unknown. Further, chromosomes are organized into a gene-rich, transcriptionally active compartment (A) with open chromatin and active histone marks and a gene-poor, transcriptionally inactive compartment (B) with condensed chromatin and gene silencing histone marks. Within this overall organization, the interplay between chromatin structure<sup>8</sup> and gene expression<sup>9-11</sup> is cell-specific and mediated by transcription factors (TFs) and other DNA binding proteins the functions of which depend on chromatin accessibility.

The immune system evolved to respond to environmental stimuli and exhibits a high degree of phenotypic and functional plasticity in response to external cues. T cells have a central role in the adaptive immune system and are activated under different conditions to expand and differentiate into a variety of specialized and functionally distinct subsets. T cell activation is likely to be instructive of how coordinate changes in chromatin structure orchestrate gene expression programs that underlie rapid and often lifesaving responses. It was recently shown that T cell activation is associated with a marked remodelling of chromatin structure at the level of accessibility (measured by ATAC-seq)<sup>12-14</sup> and, separately, of conformation (measured by promoter capture-Hi-C)<sup>15</sup>. Here

<sup>1</sup>The Walter and Eliza Hall Institute of Medical Research, Parkville 3052, Australia. <sup>2</sup>Department of Medical Biology, The University of Melbourne, Parkville 3010, Australia. <sup>3</sup>School of Mathematics and Statistics, The University of Melbourne, Parkville 3010, Australia. <sup>4</sup>These authors contributed equally: Naiara G. Bediaga and Hannah D. Coughlan. ✉email: harrison@wehi.edu.au

we analyse genome-wide chromosome conformation in conjunction with chromatin accessibility and whole transcriptome expression to further understand how coordinated changes across different levels of chromatin structure are linked to gene expression in response to T cell activation.

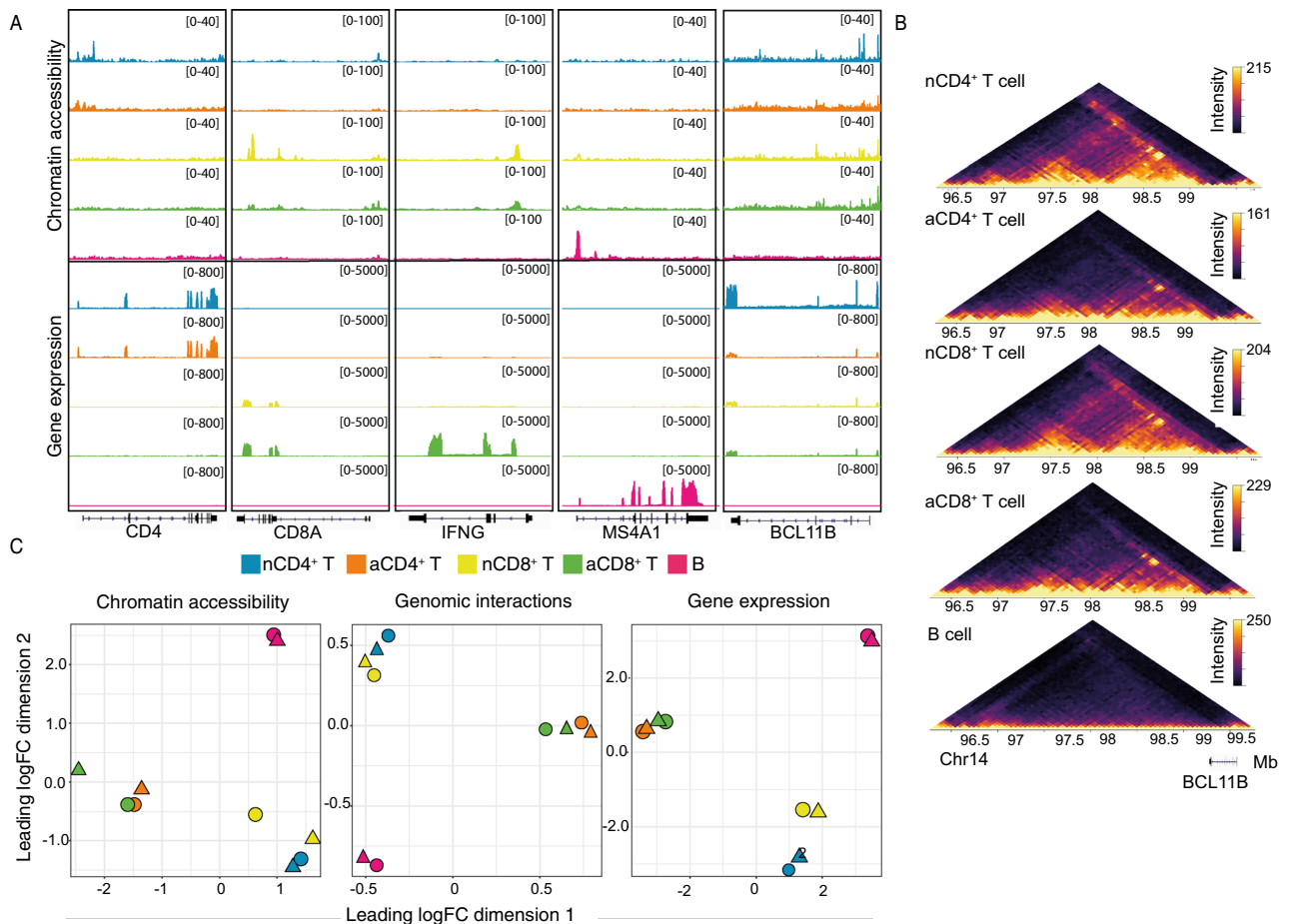
## Results

**Chromatin accessibility and 3D genomic interactions are immune cell-type specific.** Resting CD4<sup>+</sup> and CD8<sup>+</sup> T cells, and mature B cells, were isolated from two healthy human male donors. T cells were activated through the T cell receptor subunit CD3 and the co-stimulator CD28 by incubation with anti-CD3/CD28 antibody Dynabeads for 72 h. Each cell lineage was profiled by assay for transposase-accessible chromatin using sequencing (ATAC-seq) to generate genome-wide maps of chromatin accessibility, by in situ Hi-C for 3D genomic interactions and by RNA-seq for whole transcriptome expression. The consistent use of two independent biological replicates for each cell population across all three genomic technologies allowed us to assess activation-induced changes in an integrated and statistically rigorous way. For each technology, a set of relevant genomic features was identified and held fixed across each cell populations. For ATAC-seq, the features were identified peaks, for Hi-C the features were pairs of genomic windows and for RNA-seq the features were genes. Activation-induced changes were explored by comparing activated to resting T cells and the statistical significance of the changes was determined with the edgeR<sup>16</sup> and limma<sup>17</sup> software packages, controlling the false discovery rate relative to biological variability between the individual donors.

First, we confirmed the reproducibility and specificity of the data. ATAC-seq, Hi-C-seq and RNA-seq sequencing read densities at genes that define the cell phenotypes confirmed specificity for each cell type (Fig. 1A–C). We observed a strong enrichment of ATAC-seq reads at transcription start sites (TSSs) genome-wide, including promoter-TSSs of the cell type-specific genes *CD4*, *CD8A*, *IFNG*, *BCL11B* and *MS4A1*, which reflected the quality of the data<sup>18</sup> (Fig. 1A,B). T cells, but not B cells, exhibited robust enrichment of Hi-C genomic interactions at the *BCL11B* gene, known to be expressed in T but not B cells<sup>19</sup> (Fig. 1C). Conversely, B cells exhibited robust enrichment of Hi-C genomic interactions at the *BCL11A* gene, known to be expressed in B but not T cells (Supplementary Fig. S1). Unsupervised clustering of the samples by multidimensional scaling of the chromatin accessibility, chromatin interaction and gene expression data demonstrated distinct chromatin structure and gene expression signatures for the different cell types (Fig. 1D), as previously reported for chromatin architecture<sup>8</sup>. Samples clustered by activation status over dimension 1 and diverged by cell lineage over dimension 2. By calculating the proportion of variation explained by each dimension using the Glimma package (v1.6.0)<sup>20</sup>, activation (dimension 1) was found to account for the largest proportion of variation in chromatin accessibility, chromatin interactions and gene expression (34%, 27% and 49%, respectively), followed by lineage (dimension 2) (23%, 19% and 17%) (Fig. 1D). Consistent with the smaller contribution of T cell lineage to biological variance in chromatin accessibility and interactions, direct comparison of activated CD4<sup>+</sup> and CD8<sup>+</sup> T cells revealed only a small number of chromatin structure differences, viz. 864 differentially accessible regions and 69 differential interactions, indicating that chromatin remodelling in response to activation is similar in CD4<sup>+</sup> and CD8<sup>+</sup> T cells.

**T cell activation leads to extensive changes in chromatin accessibility and gene expression.** Relative to the resting state, T cell activation increased accessibility of 11,368 (10.0%) and 12,282 (10.8%) ATAC-peaks in CD4<sup>+</sup> and CD8<sup>+</sup> T cells, respectively, while 3674 (3.2%) and 4,723 (4.1%) peaks lost accessibility (Fig. 2A, Supplementary Table S1). Activation altered a higher proportion of expressed genes, with 4894 (33.4%) and 4727 (32.3%) being up-regulated and 4931 (33.7%) and 4611 (31.5%) in CD4<sup>+</sup> and CD8<sup>+</sup> T cells, respectively (Fig. 2B). Interestingly, despite differences in their effector functions, activation-associated changes in chromatin accessibility and gene expression were highly concordant in CD4<sup>+</sup> and CD8<sup>+</sup> T cells, evidenced by their strong correlation (Fig. 2C,D) and large number of overlapping activation-associated chromatin accessibility (9083) and gene expression (7823) changes (Fig. 2E,F; 'aCD4<sup>+</sup> vs. nCD4<sup>+</sup> T cell' and 'aCD8<sup>+</sup> vs. nCD8<sup>+</sup> T cell' comparisons). Thus, differentially accessible (DA) peaks and differentially expressed (DE) genes up- and down-regulated upon activation of CD4<sup>+</sup> T cells were also up- and down-regulated, respectively, upon activation of CD8<sup>+</sup> T cells (Fig. 2C,D). Accessibility and gene expression differences between activated CD4<sup>+</sup> and CD8<sup>+</sup> T cells were also minimal (Fig. 2E,F; 'aCD4<sup>+</sup> vs. aCD8<sup>+</sup> T cell' comparison). Moreover, in interrogating the DA peaks for enrichment of known transcription factor motifs we observed that activation-associated accessibility changes in both CD4<sup>+</sup> and CD8<sup>+</sup> T cells were enriched for a similar set of DNA motifs recognized by TFs involved in T cell development, activation or proliferation<sup>10,21–25</sup> (Supplementary Fig. S2A). These included members of the bZIP (BATE, FOSL1, ATF3), ETS (Fli1, ETV1, ERG, GABPA), and Runt (RUNX1, RUNX2) families among others (Supplementary Fig. S2A). Comparison with ATAC-seq signatures (GSE118189) reported in human CD4<sup>+</sup> and CD8<sup>+</sup> T cells after 24 h activation<sup>10,21–25</sup>, revealed that 24 and 72 h activation-associated accessibility changes were also highly concordant (Supplementary Fig. S3A), had a large number of overlapping DA peaks (Supplementary Fig. S3B) and were enriched for a similar set of TF motifs (Supplementary Fig. S3C). Thus, both the earlier and later stages of CD4<sup>+</sup> and CD8<sup>+</sup> T cell activation would appear to be governed by similar chromatin regulatory and transcription factor programs.

Using annotations from the Refseq database and predicted regulatory states from ChromHMM in the Roadmap Epigenomics Project<sup>26</sup>, we annotated the activation-associated DA peaks and calculated their enrichment over the universe of ATAC-seq peaks using GAT<sup>27</sup>. Overall, approximately 77% of the activation-induced peaks were located in intronic and intergenic regions while the remaining were at promoters (defined from –1 kb to +100 bp of the TSS) (~14%) and exonic regions (~3%). Moreover, DA peaks showed strong enrichment for ChromHMM-predicted promoters and enhancer regions, and depletion in repressive marks (heterochromatin, quiescent DNA and repressed PolyComb) (Supplementary Fig. S2B).

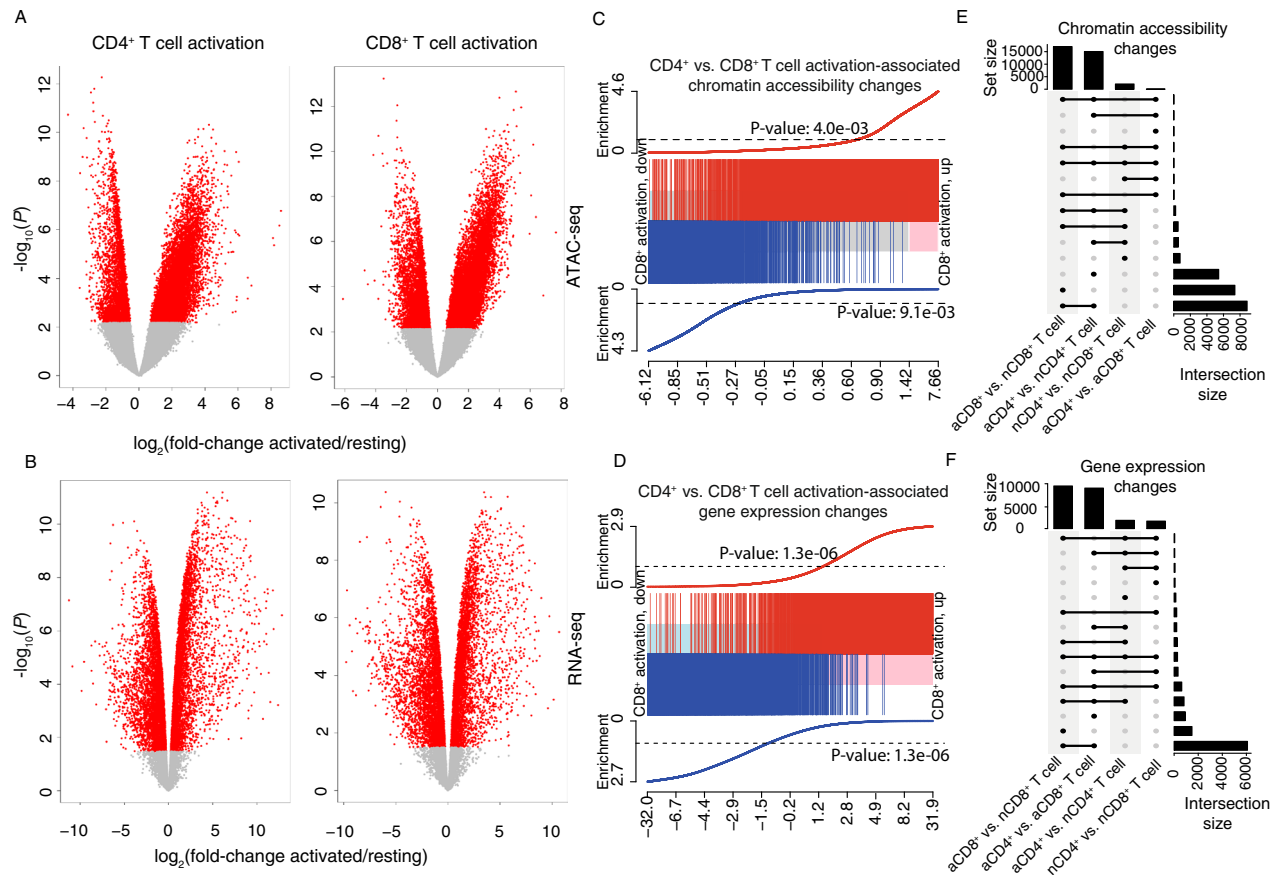


**Figure 1.** Chromatin structure and gene expression are immune cell type-specific. (A) Normalized read coverage plots of ATAC-seq and RNA-seq libraries at phenotype-defining genes including *CD4*, *CD8A*, *IFNG*, *MS4A1* and *BCL11B* loci in resting  $CD4^+$  ( $nCD4^+$ ), resting  $CD8^+$  ( $nCD8^+$ ), activated  $CD4^+$  ( $aCD4^+$ ), activated  $CD8^+$  ( $aCD8^+$ ) T and resting B (B) cells. (B) In-situ Hi-C contact matrices for a 1 Mb region on chromosome 14 that includes *BCL11B*, which is known to be expressed in T but not B cells. The top four Hi-C matrices display data from resting and activated T cells and show distinct genome organisation at the gene, while the bottom Hi-C matrix of B cells shows a lack of genome organisation at *BCL11B*. Color scale indicates number of reads per bin pair and has been scaled within each matrix to facilitate comparisons. (C) Multidimensional scaling (MDS) plots of log-CPM values with samples coloured by cell type and shaped by donors. The log-CPM values were corrected for the donor variable. Distances on the plot correspond to the leading fold-change, which is the average (root-mean-square) log<sub>2</sub>-fold-change for the 500 genes (RNA-seq), 5000 peaks (ATAC-seq) and 50,000, bin pairs (Hi-C) most divergent between each pair of samples.

Several studies have shown that disease-associated SNPs frequently reside in cell-specific regulatory elements<sup>28–30</sup>. We therefore tested whether regions of altered chromatin accessibility associated with T cell activation were enriched for disease-/trait-associated SNPs or those in linkage disequilibrium (LD) with them (Supplementary Table S2). After categorizing SNPs for 15 disease or trait classes, we observed a strong enrichment for immune system, neurological autoimmune, haematological and cancer SNPs in DA peaks compared to the universe of ATAC-seq peaks (Supplementary Fig. S2C). Moreover, enrichment was more pronounced in the subset of peaks that contained predicted enhancers by ChromHMM (Supplementary Fig. S2C).

Because whole transcriptome sequencing without poly(A) pre-selection allowed us to detect enhancer RNAs<sup>15</sup>, we could identify the subset of enhancer (e)RNAs that overlapped the set of human enhancer regions defined by the FANTOM consortium<sup>31</sup>. In order to investigate if variation in chromatin accessibility was related to changes in eRNA expression, and thus enhancer activity, we calculated the distribution of eRNA expression across the DA peaks. We observed that eRNA expression was on average significantly higher in the chromatin regions that gained accessibility upon activation than in those that lost accessibility (Supplementary Fig. S4A).

To further explore the relationship between the activation-associated DA peaks and expression of their potential target genes we applied the Genomic Regions Enrichment of Annotations Tool (GREAT)<sup>32</sup>. We observed that expression of genes linked to the peaks that gained accessibility was on average significantly higher than that of genes linked to peaks that lost accessibility (Supplementary Fig. S4B). The set of target genes defined by GREAT comprised immune system-associated genes known to be up-regulated upon T cell activation, including *TNFSF15*, *TNFRSF8*, *IL32*, *IL23R*, *IL12RB2*, *IL1R2*, *IL2*, *IL21* and *IL13* among others. Furthermore, functional



**Figure 2.** T cell activation induces extensive changes in chromatin accessibility and gene expression that are comparable between CD4<sup>+</sup> and CD8<sup>+</sup> T cells. **(A)** Volcano plots showing the magnitude of differential accessibility in activated CD4<sup>+</sup> and CD8<sup>+</sup> T cells. Each point represents a peak with significant changes highlighted in red (FDR < 0.05). **(B)** Volcano plots showing the magnitude of differential expression in activated CD4<sup>+</sup> and CD8<sup>+</sup> T cells. Each point represents a gene with significant changes in red (FDR < 0.05). **(C)** Barcode enrichment plot showing that DA peaks in CD4<sup>+</sup> T cells are similarly ranked in CD8<sup>+</sup> T cells. Peaks are ordered from left to right on the plot from most down-regulated to most up-regulated upon activation in CD8<sup>+</sup> T cells, forming a shaded middle horizontal bar, with log<sub>2</sub>-fold-changes shown by the x-axis. DA peaks significantly up- or down-regulated upon activation in CD8<sup>+</sup> T cells (FDR < 0.05) are marked by red and blue vertical bars, respectively. The red enrichment worm at the top of the plot shows the local density of red vertical bars while the blue enrichment worm at the bottom of the plot shows the local density for blue vertical bars. P-values show significance of the enrichments (Fry gene set tests). The enrichment pattern shows strong concordance of the CD4<sup>+</sup> and CD8<sup>+</sup> changes. **(D)** Barcode enrichment plot showing that DE genes in CD4<sup>+</sup> T cells are similarly ranked in CD8<sup>+</sup> T cells. **(E)** UpSet plot showing the number of DA peaks for each comparison (activated vs. resting CD8<sup>+</sup> T cells, activated vs. resting CD4<sup>+</sup> T cells, resting CD4<sup>+</sup> vs. resting CD8<sup>+</sup> T cells and activated CD4<sup>+</sup> vs. activated CD8<sup>+</sup> T cells) and their intersections. Vertical bars illustrate the total number of DA peaks for each comparison. Horizontal bars show the number of common DA peaks (intersection size) for a given set of comparisons (filled connected circles). **(F)** UpSet plot showing the number of DE genes for the same comparisons as in **(E)**.

enrichment analysis by GREAT showed that DA peaks were significantly enriched for a number of relevant Gene Ontology terms, including regulation of immune system process, regulation of response to stimulus or regulation of cellular processes, among others.

In summary, this integrated analysis reveals that T cell activation leads to extensive remodelling of chromatin accessibility and the formation of active regulatory chromatin regions associated with TFs relevant to T cell biology, enhancer activity and expression of genes critical in T cell responses, comparable between CD4<sup>+</sup> and CD8<sup>+</sup> T cells.

**T cell activation leads to genome-wide changes in chromatin interactions.** Chromatin interactions bring disparate regulatory elements and genes into spatial proximity to regulate transcription. To identify how the chromatin interactions change upon T cell activation we used the diffHic pipeline<sup>33</sup> in conjunction with edgeR. The pipeline partitions the genome into 25 kbp bins. Each pair of bins from the same chromosome represents the location of a potential chromatin interaction, with the intensity of the interaction measured by



the number of reads mapping to that bin-pair. Statistical tests for gain or loss of intensity were performed for each bin-pair. Upon activation, we found 17,965 and 15,496 interactions were gained in CD4<sup>+</sup> and CD8<sup>+</sup> T cells, respectively, and 12,434 and 11,376 interactions were lost (Supplementary Tables S3A and S3B). Across the whole genome, 103,741 bins had above-background interaction activity and more than half of these (56.3% in CD4<sup>+</sup> T cells and 52.4% in CD8<sup>+</sup> T cells) were involved in a least one significant differential interaction (DI). Similar to the chromatin accessibility and gene expression changes upon activation, activation-associated chromatin interaction changes were highly concordant between the CD4<sup>+</sup> and CD8<sup>+</sup> T cells, reflected by the strong positive correlation (Fig. 3A).

Examples of highly ranked DIs are shown in Supplementary Fig. S5. The figures show the interactions that are gained and lost on activation together with the associated genes that are DE and regions that are DA. We next identified a network of chromatin interactions connecting DA peaks with the promoters of expressed genes, which we refer to as gene-regulatory chromatin interactions. Regions with gene-regulatory chromatin interactions 'gained' were associated with a larger increase in chromatin accessibility compared to those with 'lost' or 'unchanged' interactions following activation of either CD4<sup>+</sup> or CD8<sup>+</sup> T cell (Fig. 3B). Furthermore, when we examined if 'gained' or 'lost' regulatory interactions were associated with up-regulation or down-regulation of gene expression we found that the proportion of 'gained' gene-regulatory interactions associated with up-regulated genes was significantly higher than that associated with 'lost' or 'unchanged' gene-regulatory interactions (Fisher's exact test  $p$ -value  $< 1.1 \times 10^{-11}$ ), while the proportion of 'lost' gene-regulatory interactions associated with down-regulated genes was also significantly higher than those linked to 'gained' or 'unchanged' gene-regulatory interactions (Fisher's exact test  $p$ -value  $< 5 \times 10^{-10}$ ) (Fig. 3C). Thus, 'gain' and 'loss' interactions are coupled, respectively, to up-regulation and down-regulation of corresponding target genes.

**T cell activation is associated with partitioning of genome topology.** Evidence has emerged of the dynamic nature of TAD and DNA loop formation, mainly from studies of mouse embryo development (reviewed in Ref.<sup>1</sup>). To examine TAD plasticity we first segmented the chromatin of resting and activated T cells, and B cells, into TADs using TADbit<sup>34</sup>. Naïve CD4<sup>+</sup> T, CD8<sup>+</sup> T and B cells had 1727, 1750 and 1602 TADs with a mean size of 1.40, 1.38 and 1.51 Mb, respectively; activated CD4<sup>+</sup> and CD8<sup>+</sup> T cells had 2772 and 2863 TADs with a mean size of 0.85 and 0.84 Mb, respectively (Fig. 4A). In activated T cells, TADs appeared to be partitioned, becoming smaller and more numerous than in resting T cells (Fig. 4A). We next explored the proportion of intersecting TADs among different TAD sets. TADs were termed intersecting if the reciprocal region overlap was higher than 75%. On average, 62% of the TADs intersected between resting CD4<sup>+</sup> T and CD8<sup>+</sup> T cells and 70% of TADs intersected between activated CD4<sup>+</sup> T and CD8<sup>+</sup> T cells. Interestingly, this overlap fell to an average of 43.5% when comparing resting and activated T cells. Further analysis of the TAD intersects revealed that while only 24% of the TADs called in resting CD4<sup>+</sup> T cells overlapped two or more TADs in resting CD8<sup>+</sup> T cells this increased to 44% when analysing the TAD overlap between resting and activated T cells (Supplementary Fig. S6). This suggests that T cell-activation induces partitioning of TADs into smaller chromatin domains.

We observed also that, analogous to TAD partitioning, chromatin interactions that strengthened upon activation were on average significantly shorter than those that were weakened ( $t = 9.3$ ;  $P$ -value  $< 2 \times 10^{-16}$ ), i.e. activation appeared to preferentially promote disruption of the longer-range interactions (Fig. 4B). Remarkably, intra-TAD chromatin interactions in TADs that became smaller upon activation showed a higher proportion of 'gain' interactions than those in TADs that remained unchanged or increased (Fig. 4C). This suggests that TADs have a pivotal role in shaping the intra-TAD chromatin loops. Examples of TAD partitioning and rearrangement of chromatin interactions are shown (Fig. 5 and Supplementary Fig. S7); significant differential interactions in response to activation are represented by red ('gained') and blue ('lost') arches, and locations of the TADs in resting (top) and activated (bottom) T cells are shown by yellow triangles.

We next sought to quantify how TADs changed upon T cell activation. Like most TAD caller algorithms, TADbit uses a binary approach whereby TAD boundaries are either present or completely absent, and thus TADbit is unable to identify subtle changes in genome architecture characterized by the quantitative strengthening or weakening of a boundary. Therefore, we used diffHic and edgeR to test for strengthening or weakening of each of the TAD boundaries. Of the 7260 and 7337 unique TAD boundaries called by TADbit, 3529 (46.0%) and 3755 (51.2%) were called differential by edgeR. Of these, 1596 and 1440 became significantly stronger and 1933 and 2315 weaker in CD4<sup>+</sup> and CD8<sup>+</sup> T cells, respectively. Thus, our findings indicate that T cell activation elicits a significant remodelling of the TAD boundaries.

We then calculated the distribution of chromatin accessibility, nucleosome occupancy and gene expression for the differential TAD boundaries in both resting and activated T cells. Interestingly, we observed that chromatin regions overlapping TAD boundaries that strengthened upon activation were associated on average with lower chromatin accessibility, higher nucleosome occupancy and lower protein-coding gene expression than those overlapping TAD boundaries weakened (Fig. 6). This trend was true for both resting and activated T cells, implying that chromatin regions associated with TAD boundary formation/disruption in T cells have intrinsic invariant features. In summary, T cell activation leads to global rearrangement of TAD boundaries, partitioning of TADs and disruption of long-range chromatin interactions, with weakening of TAD boundaries associated with regions of higher accessibility and gene expression.

**Chromosome A/B compartments are not altered substantially by T cell activation.** Lineage differentiation is associated with rearrangement of chromosome A/B compartments<sup>35</sup>. However, in response to T cell activation we observed only modest changes in the A/B compartments, with less than 824 (1.4%) and 476 (0.8%) of the genome switching compartments upon activation CD4<sup>+</sup> and CD8<sup>+</sup> T cells, respectively (Fig. 7A,B). As expected, the small proportion of regions that switched from A to B showed decreased accessibility and gene

expression, whereas regions that switched from B to A tended to show higher accessibility and gene expression (Fig. 7C). Changes across T and B cell types were also minimal. Overall, the pattern of change was subtle, indicating that T cell activation has minimal impact on the A and B compartments.

## Discussion

Previous studies have characterized either activation-responsive chromatin accessibility<sup>13,14,36,37</sup> or 3D organization of the genome<sup>8,13,15</sup> in mouse and human immune cells. However, to our knowledge no previous study has characterized chromatin accessibility, Hi-C chromatin interactions and the whole transcriptome in concert in response to activation of human CD4<sup>+</sup> and CD8<sup>+</sup> T cells. By using biological replicates throughout, changes in chromatin accessibility, interaction intensity and gene expression could be rigorously tested statistically, relative to the biological variation observed between human individuals. Activation of T cells was associated with multi-level remodelling of chromatin that was comparable between CD4<sup>+</sup> and CD8<sup>+</sup> T cells, indicating that their activation is regulated by similar genome control mechanisms. In response to T cell activation, we observed extensive changes in chromosome accessibility in regions enriched for regulatory elements and eRNAs, immune system-associated SNPs and TFBSs recognized by TFs involved in T cell development, activation or proliferation. These findings parallel those in other systems<sup>12–14,36,37</sup> and are consistent with the view that changes in chromatin accessibility in response to external cues identify active regulatory regions likely to be critical for transcriptional responses. Furthermore, we also show that T cell activation uncovers many additional regulatory regions likely to be associated with the immune response and induces thousands of differential chromatin interactions linked to up-regulation or down-regulation of the corresponding target genes. At a higher level of chromosome architecture, T cell activation was associated with genome-wide rearrangement of TAD boundaries.

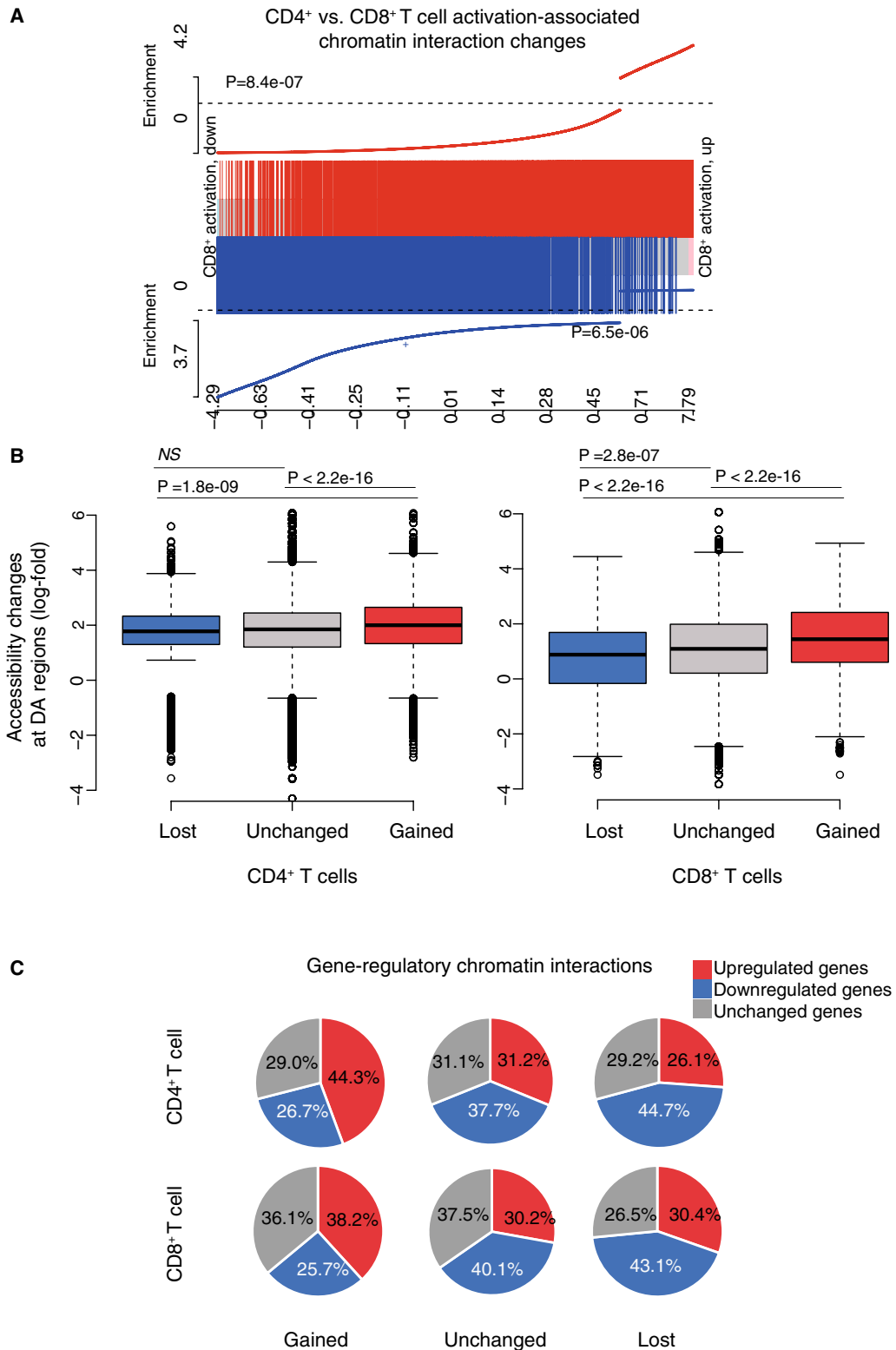
Although TADs were reported to be largely invariant across species and cell types<sup>5,6</sup> our findings indicate otherwise and, in accord with recent evidence<sup>1</sup>, reveal that TADs are dynamic. We found that T cell activation was associated with global rearrangement of chromatin interactions involving changes in the strength of TAD boundaries and a smaller compartment structure. Strengthened TAD boundaries displayed lower chromatin accessibility, higher nucleosome occupancy and lower gene expression both before and after activation. Nucleosome organization is known to dictate local chromatin folding by regulating internal factors including linker DNA length and linker-histone binding affinities, but whether nucleosome architecture influences higher level genome topology, i.e. TADs and DNA loops, remains unknown<sup>38</sup>. Polymer simulations of the chromatin fiber suggest that regions of “stiffness” can act as local insulators to decrease interaction frequencies<sup>39</sup>. In mammary epithelial MCF-10A cells, knockdown of BRG1, the major ATPase subunit of the SWI/SNF chromatin remodelling complex, globally altered long-range genomic interactions and decreased TAD boundary strength, suggesting that nucleosome occupancy around CTCF sites might contribute to the regulation of higher-order chromatin architecture<sup>40</sup>. Our results in primary human cells support the view that long-range chromosomal organization and nucleosome occupancy are molecularly coupled and suggest that chromatin regions associated with TAD boundary formation have invariant features reflected by lower accessibility and higher nucleosome occupancy in both resting and activated T cells. Further examination of other elements at differential TAD boundaries may provide valuable insights into the TAD formation process and its role in chromatin folding and transcriptional regulation.

Together, our findings connect multi-level features of chromatin structure to gene expression in response to T cell activation. In addition, we provide a genome-wide resource of T cell activation-associated regulatory regions that will aid interpretation of chromatin accessibility data in human T cells.

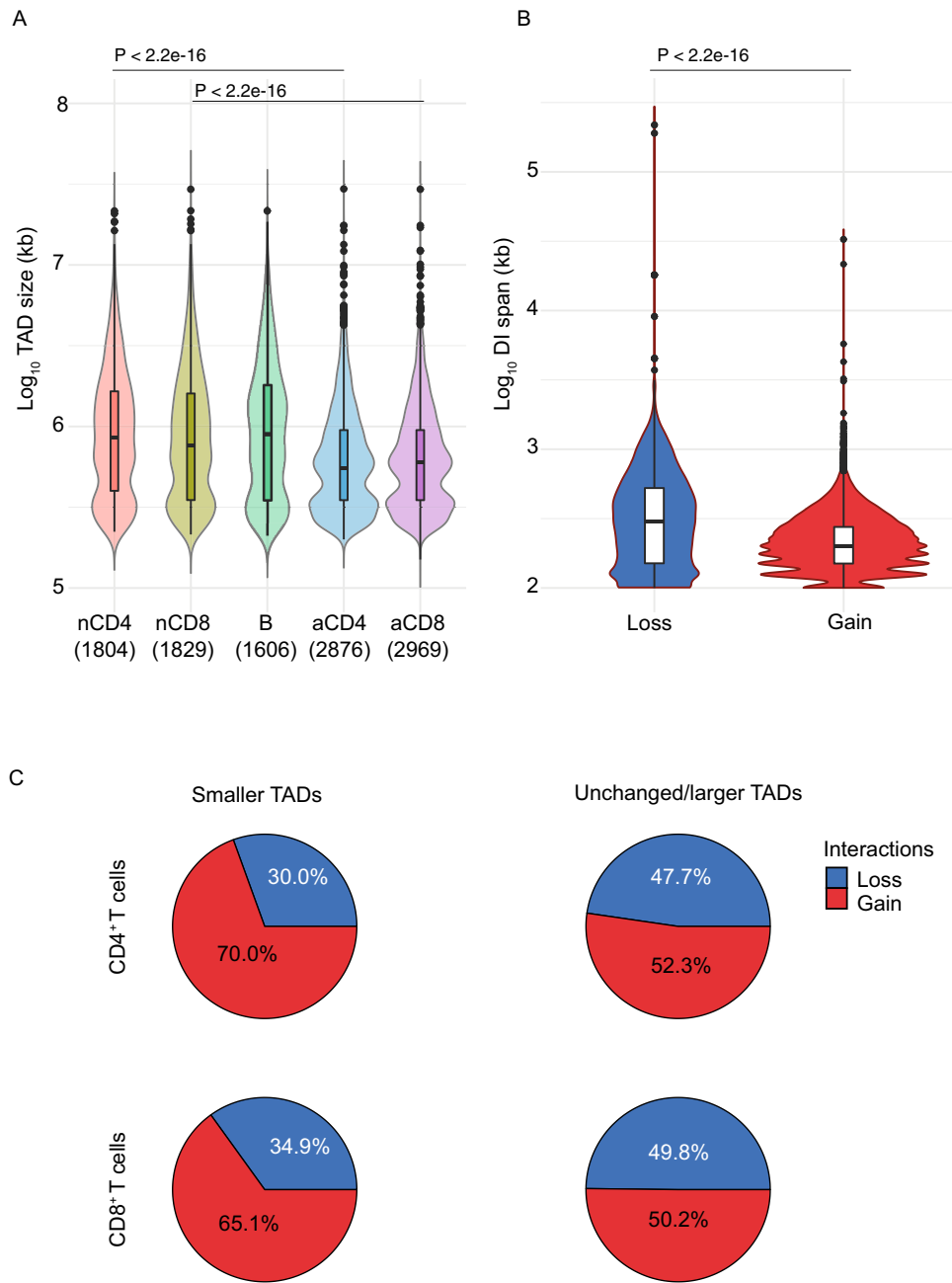
## Methods

**Blood collection, cell subset isolation and T cell activation.** Heparinised venous blood was collected at 09.00 h from two healthy young adult males, who gave written informed consent. The research was carried out in accordance with the principles of the Declaration of Helsinki and the guidelines of Nature journals and was approved by the Walter and Eliza Hall Institute of Medical Research Human Research Ethics Committee (application 88/03). Peripheral blood mononuclear cells (PBMCs) were purified by Ficoll-Hypaque gradient centrifugation and cryopreserved in liquid N<sub>2</sub>. Thawed PBMCs,  $\geq 92\%$  viable by acridine orange-ethidium bromide staining, were stained with anti-human  $\alpha\beta$  TCR (eBioscience, clone IP26, cat. no. 46-9986-42), anti-human CD4 (BD Pharmingen, clone RPA-T4, cat. no. 555349), anti-human CD45RA (BD Pharmingen, clone 5H9, cat. no. 556626), anti-human CD25 (BD Pharmingen, clone M-A251, cat. no. 557741), anti-human CD14 (BioLegend, clone 63D3, cat. no. 367104), anti-human CD16 (BD Pharmingen, clone 3G8, cat. no. 557758), anti-human HLA-DR (eBioscience, clone L243, cat. no. 48-9952-42), and anti-human CD19 (BioLegend, clone HIB19, cat. no. 302238). Naïve CD4<sup>+</sup> T cells (CD14<sup>-</sup> CD16<sup>-</sup> TCR $\alpha\beta$ <sup>+</sup> CD4<sup>+</sup> CD45RA<sup>+</sup> CD25<sup>-</sup>) and CD8<sup>+</sup> T cells (CD14<sup>-</sup> CD16<sup>-</sup> TCR $\alpha\beta$ <sup>+</sup> CD4<sup>+</sup> CD45RA<sup>+</sup> CD25<sup>-</sup>), and B cells (TCR $\alpha\beta$ <sup>-</sup>, HLA-DR<sup>+</sup>, CD19<sup>+</sup>), were flow-sorted on a FACSAria (BD Biosciences). Resting CD4<sup>+</sup> and CD8<sup>+</sup> T cells were cultured in Iscove's Modified Dulbecco's Medium containing 5% pooled, heat-inactivated, human serum, 100 nM non-essential aminoacids, 2 mM of glutamine and 50  $\mu$ M 2-mercaptoethanol (IP5) medium. T cells were activated through the T cell receptor subunit CD3 and the co-stimulator CD28 by the addition of Human T-Activator CD3/CD28 Dynabeads at a 1:2 bead:cell ratio (Life Technologies, cat. no. 111.31D) for 72 h.

**RNA-seq.** RNA was isolated using the miRNeasy Micro Kit (QIAGEN, cat. no. 217084). RNA libraries were prepared with an Illumina's TruSeq Total Stranded RNA kit with Ribo-zero Gold (Cat# RS-122-2001, Illumina) according to the manufacturer's protocol. rRNA-depleted RNA was purified and then reverse transcribed into cDNA using SuperScript II reverse transcriptase (Cat# 18064014, Invitrogen). Validation of the library preparations was performed using Agilent D1000 ScreenTape Assay on the Agilent 2200 TapeStation. Total RNA-Seq libraries were



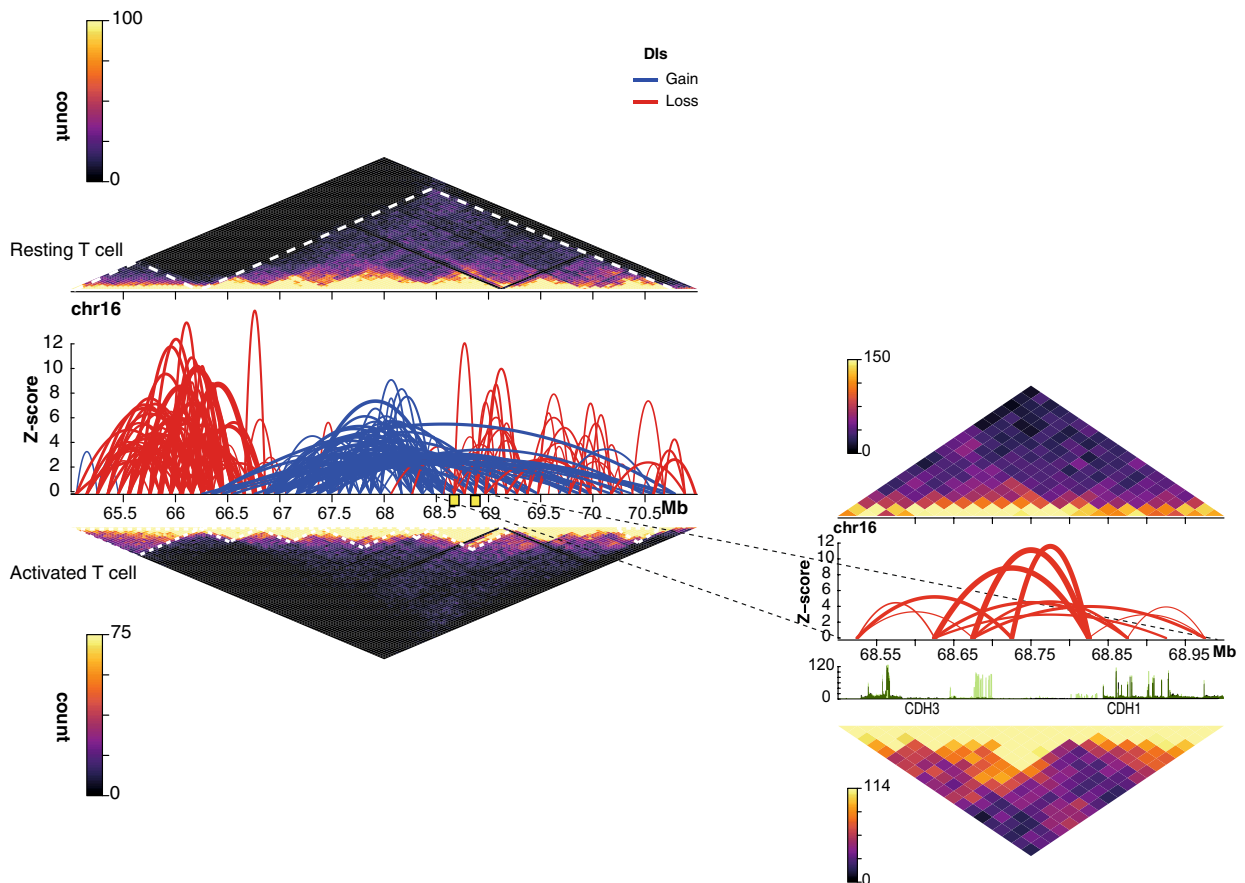
**Figure 3.** T cell activation induces genome-wide changes in gene-regulatory interactions. (A) Barcode enrichment plot showing that activation-associated DIs in CD4<sup>+</sup> T cells are similarly ranked in CD8<sup>+</sup> T cells. The plot is exactly analogous to Figs. 2C,D but for DI bin-pairs instead of DA peaks or DE genes. (B) Box plots showing that the increased accessibility at DA regions in gene-regulatory chromatin interactions ‘gained’ is higher than that at gene-regulatory chromatin interactions ‘unchanged’ or ‘lost’, following CD4<sup>+</sup> (left) and CD8<sup>+</sup> (right) T cell activation. Shown are the median (central horizontal line), interquartile range (boxes), values of the upper and lower quartiles (whiskers), outliers beyond 1.5 IQR (circles). Statistical comparisons were made with the unpaired Wilcoxon test. (C) Pie chart showing distribution of down-regulated, up-regulated or unchanged target genes linked to ‘lost’ (top) or ‘gained’ (bottom) of regulatory interactions..



**Figure 4.** T cell activation results in partitioning of genome topology. **(A)** Violin plots showing the distribution of TAD sizes in resting and activated T cells and B cells determined from summed libraries called with TADbit. Numbers in parenthesis depict the number of TADs called in each category. **(B)** Violin plots showing the distribution of the span of differential chromatin interactions (DIs) determined at 25 kbp resolution (FDR < 0.05) that are lost (decrease in logFC) or gained (increase in logFC) upon activation. Violin plots show median (horizontal line), interquartile range (boxes), values of the upper and lower quartiles (whiskers), outliers (circles) and kernel density estimation. Data for CD4<sup>+</sup> and CD8<sup>+</sup> T cells were merged as they showed a similar pattern. **(C)** Pie charts showing distribution of Loss and Gain chromatin interactions in TADs that become smaller in size (left) or remain unchanged/became larger in size (right) upon T cell activation. Statistical comparisons were made with the unpaired Wilcoxon test.

sequenced on the Illumina NextSeq 500 platform to produce 2 × 75 paired-end reads. All samples were aligned to the human genome, build hg38, using the Rsubread aligner v1.24.1<sup>41</sup>. The number of fragments overlapping each Entrez gene were summarised using featureCounts<sup>42</sup> and Rsubread's inbuilt hg38 annotation. Gene symbols were associated with Gene IDs using NCBI gene information ([ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/GENE\\_INFO](ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/GENE_INFO)).



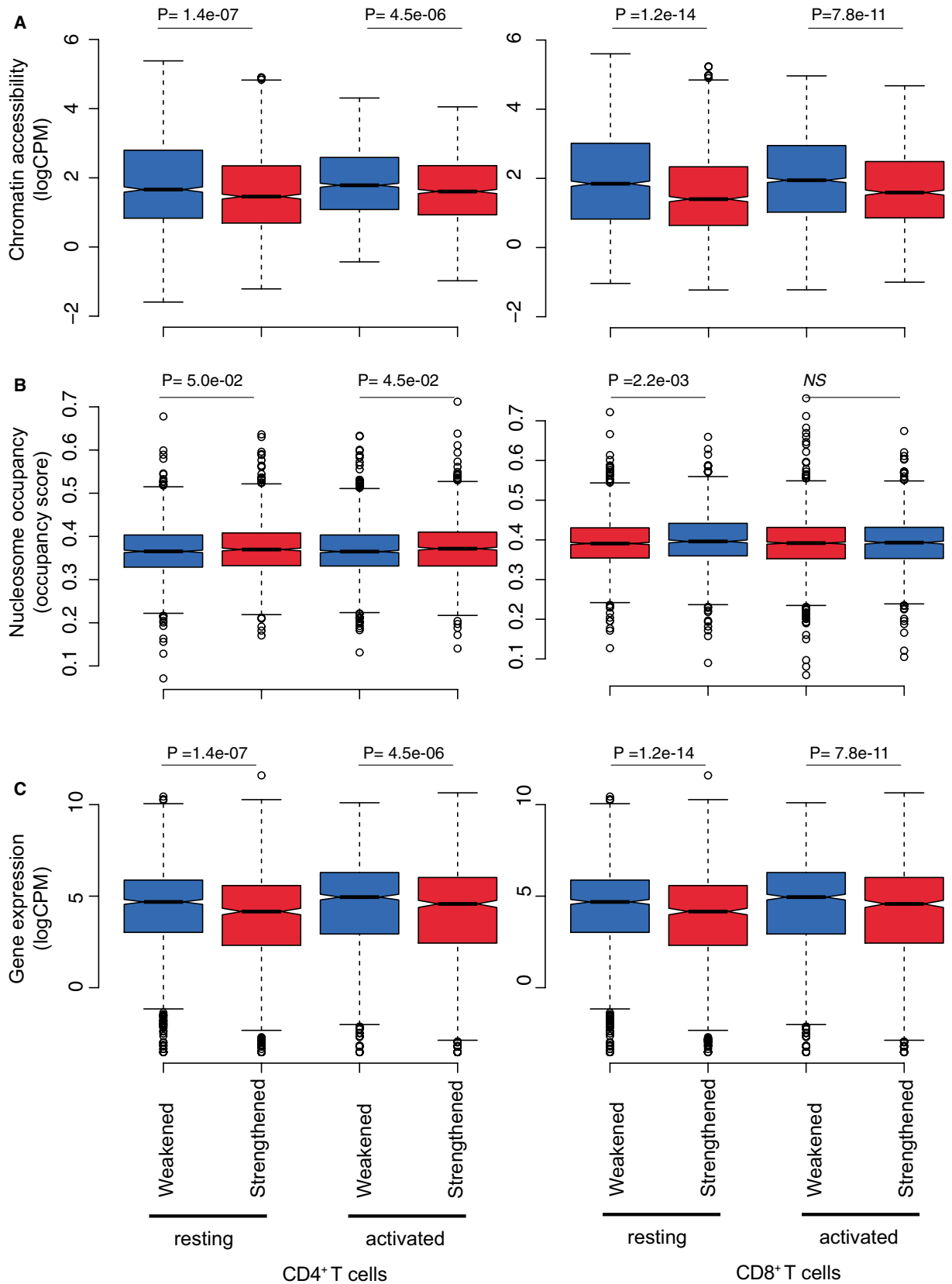


**Figure 5.** T cell activation results in partitioning of genome topology. In-situ Hi-C contact matrices plotted at 50 kbp resolution in resting and activated T cells over genes that have been associated with T cell activation such as CDH3 and CDH1. Colour scale of contact matrix indicates number of reads per bin pair. Locations of the TADs are marked by white dotted lines as called by TADbit. Significant differential interactions (DIs) as determined by the diffHic pipeline at 100 kbp (FDR < 0.05) are represented by arcs in the centre, where the vertical axis is the z-score ( $-\log_{10}$  p-value as calculated by edgeR). Each arc connects the two interacting bins and therefore identifies the relevant bin pair. Red and blue arcs represent strengthened ( $\log_{2}\text{FC} > 0$ ) and weakened ( $\log_{2}\text{FC} < 0$ ) DIs, respectively, in response to activation. The zoomed inset regions show in-situ Hi-C contact matrices and DIs at 25 kbp resolution (FDR < 0.05) and RNA sequencing coverage plots at the CDH3 and CDH1 loci. Light and dark green coverage plots represent activated and resting T cells, respectively.

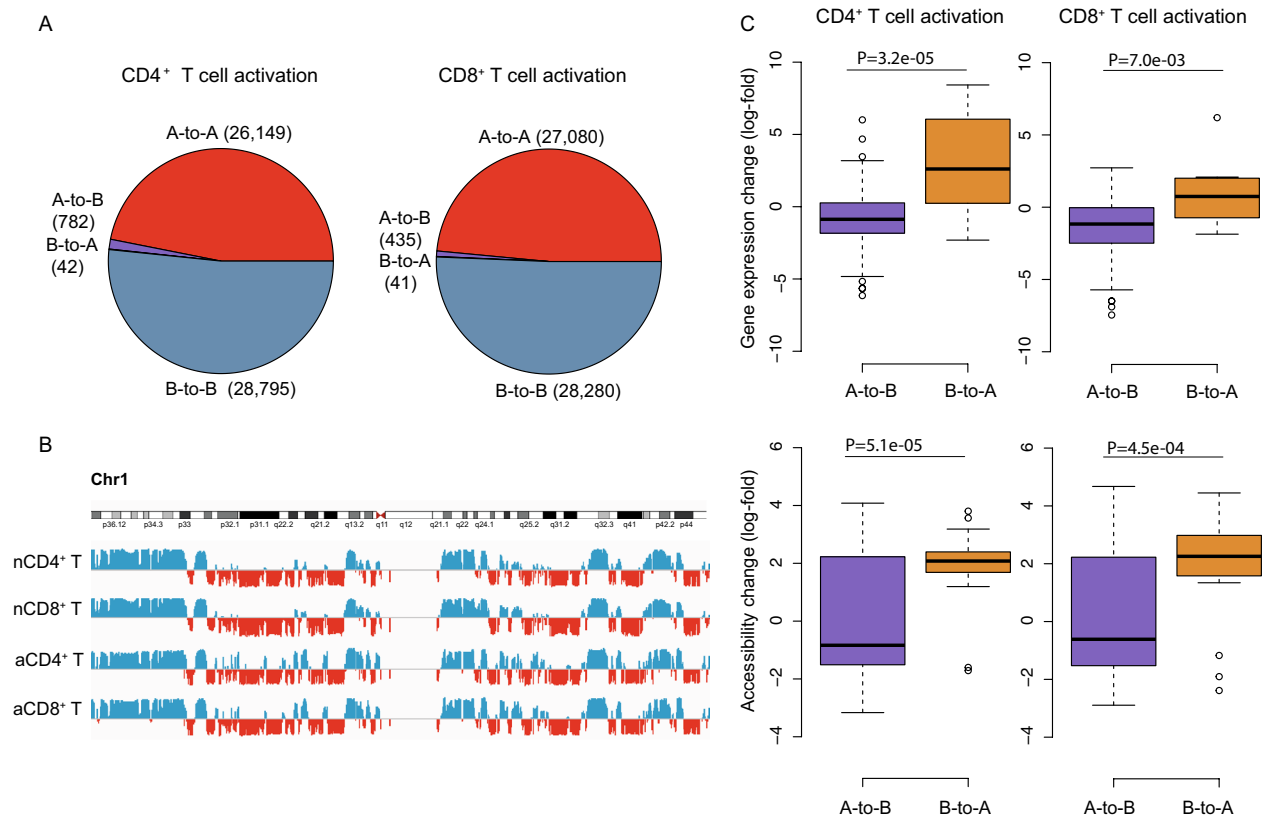
**ATAC-seq.** ATAC-seq was performed as previously described<sup>18</sup>. For each immune cell subset, 50,000 purified cells were lysed in cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub> and 0.03% Tween20). Immediately after lysis, nuclei were spun out at 500 g for 8 min at 4 °C and the supernatant carefully removed. Nuclei were resuspended with Tn5 transposase reaction mix (25  $\mu$ l 2X TD buffer, 2.5  $\mu$ l Tn5 transposase, and 22.5  $\mu$ l nuclease-free water) (Nextera DNA Library Prep Kit (Illumina, cat. no. FC-121-1030)). The transposase reaction was performed at 37 °C for 30 min, and DNA immediately purified with a Qiagen MinElute kit (QIAGEN, cat. no. 28204). ATAC-seq libraries were sequenced on an Illumina NextSeq 500 to produce 75 bp paired-end reads.

**In situ Hi-C.** In situ HiC was performed as previously published<sup>43</sup>. Primary immune cell libraries were generated in biological duplicates. Libraries were sequenced on an Illumina NextSeq 500 to produce 81 bp paired end reads. Between 129 and 280 million valid read pairs were generated per sample.

**Differential gene expression analysis.** Differential expression analyses were undertaken using the edgeR v3.20.9<sup>44</sup> and limma v3.34.9<sup>17</sup> software packages. Any gene which did not achieve a count per million mapped reads (CPM) greater than 1.5 in at least 2 samples were deemed to be unexpressed and subsequently filtered from the analysis. Additionally, all genes without current annotation were also removed. Compositional differences between libraries were normalized using the trimmed mean of log expression ratios method (TMM)<sup>45</sup>. All counts were then transformed to log<sub>2</sub>-CPM with associated precision weights using voom<sup>46</sup>. Differential expression between all cell types was assessed using linear models and robust empirical Bayes moder-



**Figure 6.** Increased strength in TAD boundary and chromatin interactions upon T cell activation is associated with nucleosome occupancy and gene expression. Box plots showing the distribution of the average (A) chromatin accessibility (logCPM), (B) nucleosome occupancy (NucleoATAC occupancy score) and (C) expression of protein-coding genes (logCPM) across differential TAD boundaries strengthened and weakened following activation of CD4<sup>+</sup> and CD8<sup>+</sup> T cells. Shown are the median (central horizontal line), interquartile range (boxes), values of the upper and lower quartiles (whiskers), outliers beyond 1.5 IQR (circles). Statistical comparisons were made with unpaired Wilcoxon tests.



**Figure 7.** Chromosome A/B compartments are not substantially altered by T cell activation. **(A)** Pie charts showing the distribution of chromatin regions in the four different group of compartment changes: stable (A to A, B to B) or flipping (A to B and B to A). This was defined by pairwise comparison between activated and resting T cells. **(B)** PCA (principal component analysis) scores in resting and activated T cells along chromosome 1 representing the stability of the A/B compartments. PCA scores were calculated as described in “Methods”. **(C)** Boxplots showing the distribution of chromatin accessibility and gene expression at regions that exhibit A/B compartment shifts upon T cell activation. Shown are the median (central horizontal line), interquartile range (boxes), values of the upper and lower quartiles (whiskers), outliers beyond 1.5 IQR (circles). Statistics were determined with the unpaired Wilcoxon test.

ated t-statistics<sup>47</sup>. To increase precision, the linear models incorporated a correction for a donor batch effect. P-values were adjusted to control the FDR below 5% using the Benjamini and Hochberg method.

**eRNA analysis.** Reads overlapping the non-exonic ATAC-seq peaks were summarised using featureCounts in the resting and activated CD4 and CD8 samples. To avoid inflation of expression estimates, the library size for each sample was set to the total number of reads aligned to the genome for that sample. All regions that failed to achieve a CPM greater than 0.5 in at least 2 samples were considered to be unexpressed and were therefore filtered from the analysis. The TMM method was then applied to normalize compositional differences between libraries and the data transformed to log<sub>2</sub>-CPM with precision weights using voom. Differential expression of the regions was then evaluated between the activated and naïve CD4<sup>+</sup> and CD8<sup>+</sup> T cell samples using linear models and robust empirical Bayes moderated t-statistics. P-values were adjusted to control the FDR below 5% using the Benjamini and Hochberg method.

**ATAC-seq data pre-processing and peak calling.** 75 bp ATAC-seq reads were aligned to the human genome assembly (hg38) using Bowtie2 v2.2.5 (bowtie2 -p 4 -X 2000)<sup>48</sup>. For each sample, mitochondrial reads, unmapped reads and low mappability (<30) reads were filtered out using Samtools (v1.6) function “view”<sup>49</sup>. After filtering, we had a median of 80 million (MAD+/- 13 million) reads per sample. Filtered ATAC-seq reads from resting and activated CD4<sup>+</sup> and CD8<sup>+</sup> T cell as well as B cells from two donors were merged using samtools function merge, and peaks were called on the merged bam file using MACS2 (v2.1.0)-callpeak (with parameters -nomodel, -extsize 200, and -shift 100)<sup>50</sup>, such that there were 113,689 peaks after excluding peaks mapping outside the main chromosome contigs. ATAC-seq reads overlapping the peaks were summarized using featureCounts<sup>42</sup>. Peaks in blacklisted genomic regions as defined by ENCODE for hg38 were removed.

**Differential accessibility analysis.** Differential accessibility analysis was undertaken using the edgeR v3.20.9<sup>16</sup> and limma v3.34.9<sup>17</sup> software packages. The TMM method was applied to normalize compositional

differences between libraries<sup>45</sup>. A mean-dependent trend was fitted to the negative binomial dispersions with the estimateDisp function and, and differential accessibility between all cell types was assessed using the quasi-likelihood (QL) framework in edgeR<sup>51,52</sup>, which assesses statistical significance relative to biological variation between the replicate libraries. As is the differential expression analysis, linear models incorporated a correction for a donor batch effect. P-values were adjusted for multiple testing using the Benjamini–Hochberg method. Peaks with a FDR below 5% were defined as differentially accessible regions. Barcode enrichment plots were drawn with limma's barcodeplot function and enrichment p-values were obtained using the fry gene set testing function in edgeR<sup>53</sup>.

For comparison of our ATAC-seq data with the previously published 24 h human T cell activation accessibility signature (GSE118189)<sup>12</sup>, raw data were pre-processed as above and only peaks common to both datasets were considered using the bedtools intersect function (version 2.19.1)<sup>54</sup>. ATAC-seq reads overlapping the intersecting list of peaks (76,812 peaks) were summarized again using featureCounts.

**Enrichment of transcription factor binding motifs.** The Homer suite v4.10<sup>55</sup> was used to determine transcription factor enrichment within ATAC peaks, using the findMotifsGenome.pl function (with parameters hg38 and –size given).

**Annotation of ATAC peaks.** Peaks were annotated as 5' UTR, 3' UTR, promoter-TSS, exonic, intronic, TTS, non-coding or intergenic using the Homer suite annotatePeaks.pl function and the default setting. Chromatin state(s) of the DA peaks were annotated using the ChIP-seq-defined ChromHMM states from the Roadmap Epigenomics Project and following the method in Corces et al.<sup>56</sup> In brief, 15 state models were downloaded from the chromatin state learning site for the 'Primary T helper naïve cells from peripheral blood' (E038) and 'Primary T CD8<sup>+</sup> naïve cells from peripheral blood' (E047) ([https://egg2.wustl.edu/roadmap/web\\_portal/chr\\_state\\_learning.html](https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html)). We then identified the regions of each ChromHMM state that were overlapped by an DA peak. To determine the significance of these overlaps for each ChromHMM state, we compared the total length of DA peaks covered by the given ChromHMM state to the expected background determined by the total length of the universe of ATAC-seq peaks (13,688 peaks) covered by the ChromHMM state using a binomial test in R.

**Enrichment for chromatin states and gene annotation.** Enrichment of ATAC-seq peaks for chromatin states as well as for gene annotations was calculated using GAT v1.3.4<sup>27</sup>. The significance threshold was set up at FDR below 5%.

**Prediction of transcriptional target genes.** To identify potential targeted genes for the activation associated accessibility changes we performed enrichment analysis of gene annotations in the proximity of the DA peaks using Genomic Regions Enrichment of Annotations Tool (GREAT)<sup>32</sup> against the whole genome as background. GREAT links genomic regions with genes by defining a 'regulatory domain' for each gene in the genome. Gene regulatory domains were defined with the "Basal plus extension" association rules (proximal 5 kb upstream and 1 kb downstream from the TSS, plus distal extended to the nearest gene's basal domain but not more than 500 kb). Significantly enriched gene sets were then selected by FDR < 0.05 for binomial tests to identify the regulatory domains with the densest clusters of activation-associated DA peaks and classify them as potential transcription target genes.

**Enrichment of GWAS loci.** GREGOR v1.4.0<sup>57</sup> was used for enrichment analysis of disease-trait associated SNPs in the ATAC-seq peaks. GREGOR calculates enrichment relative to MAF, TSS-distance and number of LD neighbor-matched null SNP sets using the GREGOR parameters:  $r^2$  threshold = 0.8, LD window size = 1 Mb and minimum neighbor number = 500. For the GWAS SNPs, we created an updated version of Supplementary Table S2 in Maurano et al.<sup>29</sup>. The GWAS SNP set used for analysis was derived from the NHGRI GWAS Catalog, downloaded on August 2, 2017. The catalog contained 41,304 entries at the time of download. We excluded SNPs mapping outside the main chromosome contigs, including the "random" chromosome fragments, SNPs without coordinates in the GRCh37/hg19 human genome assembly. There were 40,929 unique SNP disease/trait combinations that represented 34,421 unique SNP IDs (Supplementary Table S2). Of these, 19,075 were in non-coding regions. Coding regions were defined by the CCDS Project (downloaded from the UCSC genome browser at <http://hgdownload.cse.ucsc.edu/goldenPath/hg38/database/ccdsGene.txt.gz> on August 4, 2017). As in Maurano et al.<sup>29</sup>, we also grouped SNPs into classes of similar diseases or traits but some of the categories were updated so they could better reflect the new and extended list of SNPs. Categories comprised: aging related; immune system; cancer; cardiovascular diseases and traits; metabolic disorder; drug metabolism; hematological parameters; kidney, lung, or liver; miscellaneous; serum metabolites; neurological/behavioral; neurological/autoimmune; parasitic or bacterial disease; quantitative traits; radiographic (primarily bone density); viral disease.

**Nucleosome occupancy.** Nucleosome occupancy was defined with the "occ.bedgraph" files using the NucleoATAC v0.3.2.1 package<sup>58</sup>.

**In situ Hi-C data pre-processing.** Each sample was aligned to the hg38 genome using the presplit\_map.py script in the diffHic package v1.10.0<sup>33</sup>. Data were pre-processed and artefacts removed as per Johanson et al.<sup>59</sup>.

**Differential interaction (DI) analysis.** DIs between all five libraries were detected using the diffHic package<sup>33</sup> at two different resolutions, 100 kbp and 25 kbp. Read pairs were counted into 25 or 100 kbp bin

pairs (with bin boundaries rounded up to the nearest MboI restriction site) using the squareCounts function. This yielded a matrix of read pair counts for each bin pair in each library. All bins with counts less than 5 were discarded along with bins on the sex chromosomes. Bins containing blacklisted genomic regions as defined by ENCODE for hg38<sup>60</sup> were also removed. Filtering of bin-pairs was performed using the filterDirect function, where bin pairs were only retained if they had average interaction intensities more than sixfold higher than the background ligation frequency. The ligation frequency was estimated from the inter-chromosomal bin pairs from a 2 Mbp bin-pair count matrix. Bins on the first diagonal of the interaction space are also removed with the filterDiag function.

For the retained bin pairs, counts were normalized between libraries using a LOESS-based approach to account for abundance-dependent biases. This was performed using the normOffsets function to obtain a matrix of offsets with bin pairs less than 100 kbp (for the 25 kbp) or 150 kbp (for the 100 kbp) from the diagonal normalized separately from other bin pairs. Tests for differential interactions were performed using the quasi-likelihood (QL) framework<sup>51,52</sup> of the edgeR package (v3.20.9). The design matrix was constructed using a layout that specified the cell lineage to which each library belonged and which individual the cells were sampled from. Using the counts and offsets for all bin pairs, a mean-dependent trend was fitted to the negative binomial dispersions with the estimateDisp function. A generalized linear model (GLM) was fitted to the counts for each bin pair<sup>16</sup>, and the QL dispersion was estimated from the GLM deviance with the glmQLFit function. The QL dispersions were then squeezed toward a second mean-dependent trend, using a robust empirical Bayes strategy<sup>47</sup> to share information between bin pairs. A P-value was computed for each bin pair using the QL F-test, representing the evidence against the null hypothesis, i.e., no difference in counts between groups. P-values were adjusted for multiple testing using the Benjamini–Hochberg method. If a bin pair had a FDR below 5%, it was defined as a DI. To reduce redundancy in the results, DIs adjacent in the interaction space were aggregated into clusters using the diClusters function to produce clustered Dis. DIs were merged into a cluster if they overlapped in the interaction space, to a maximum cluster size of 500 kbp (for the 25 kbp) or 1Mbp (for the 100 kbp) to mitigate chaining effects. The significance threshold for each bin pair was defined such that the cluster-level FDR was controlled at 5%. Cluster statistics were computed using the combineTests and getBestTest functions from the csaw package v1.12.0<sup>61</sup>. Clustered DIs were used to report the number of differences between the libraries. The 25 kbp unclustered DIs were used for overlap analysis and integration with other data types. Contact matrices were created from the libraries using the inflate function in diffHic for various bin sizes with no filtering. Contact matrices from biological replicates were summed.

**Detection of TADs.** TAD breakpoints were detected with the TADbit v0.2.0.5 python based software<sup>34</sup>. Read pairs were counted into 50 kbp bin pairs (with bin boundaries rounded up to the nearest MboI restriction site) using the squareCounts function of diffHic with no filter. This yielded a count matrix containing a read pair count for each bin pair in each library. The count matrix was converted into a contact matrix for each somatic chromosome with the inflate function of the InteractionSet package<sup>61</sup>. Replicate contact matrices were summed to obtain one matrix for each cell type. The TADbit tool *find\_tad* was run on the raw counts specifying normalized = FALSE. Only TADs boundaries with a score of 7 or higher were included in the results.

**Detection of differential TAD boundaries (DTBs).** Changes in TAD boundaries between activated and resting cells were assessed with the diffHic and edgeR packages using the approach described in Chapter 8 of the diffHic User's Guide. This approach adapts the statistical strategy recently developed for differential methylation<sup>62</sup> to identify TAD boundaries that are significantly strengthened or weakened between cell types. The strength of each TAD boundary was assessed based on the upstream vs downstream intensity contrast at that genomic loci, defined as the ratio of upstream to downstream Hi-C read pairs anchored at that genomic region. edgeR was used to test whether the ratio increased or decreased for each boundary upon T cell activation. This method directly assesses differential boundary strength relative to biological variation without needing to make TADs calls in individual samples. Upstream and downstream read counts were determined for the same 50 kbp genomic regions as used for the TADbit analysis. An upstream (downstream) read was defined as one with anchors in the 50 kbp region and 0–1 Mbp upstream (downstream) of the region. Low abundance regions with average log<sub>2</sub>-counts per million below 0.8 were removed. An appropriate edgeR linear model was fitted to the counts for all the T cell samples. Tests for differential TAD boundaries were performed using the QL framework of the edgeR package as described in the DI analysis section above. Regions with FDR below 0.05 were considered to be significantly different.

**Detection of A/B compartments.** A/B compartments were identified at a resolution of 50 kbp using the method outlined by Lieberman-Aiden et al.<sup>63</sup> using the HOMER HiC pipeline<sup>55</sup>.

After processing with the diffHic pipeline libraries were converted to the HiC summary format using R. Then input tag directions were created for each library with the makeTagDirectory function specifying the genome (hg38) and restriction enzyme cut site (GATC). Biological replicates tag directories for each cell type were summed. The runHiCpca.pl function was used on each library with -res 50,000 and specifying the genome (hg38) to perform PCA to identify compartments. To identify changes in A/B compartments between libraries, the getHiCcorrDiff.pl function was used to directly calculate the difference in correlation profiles. We identified regions with an A-to-B or B-to-A compartment flip that showed a correlation profile lower than 0 and regarded them as differential compartments.

**Visualization of the data and plots.** Plots were performed using R and ggplot. Multi-dimensional scaling (MDS) plots were constructed using the plotMDS function in the limma v3.34.9 package<sup>17</sup> applied to the



filtered and normalized log<sub>2</sub>-counts-per-million values for each library. The removeBatchEffect function of the limma package was used to correct for effect of the individual and batch in the data. Plaid plots were constructed using the contact matrices and the plotHic function from the Sushi v1.22.0 R package<sup>64</sup>. The inferno color palette from the viridis v0.5.1 package<sup>65</sup> was used and the range of color intensities in each plot was scaled according to the library size of the sample, to facilitate comparisons between plots from different samples. DI arcs were plotted with the plotBedpe function of the Sushi package. The z-score shown on the vertical axis was calculated as  $-\log_{10}$  (P-value). RNA-seq coverage was plotted with the plotBedgraph function of the Sushi package and Integrative Genomics Viewer, IGV. UpSet plots were generated using the UpSet function in the ComplexHeatmap package in R<sup>66</sup>.

**Data intersection.** Intersection between pairs of the genomic regions was performed using bedtools intersect (v2.19.1)<sup>54</sup>.

### Data availability

Raw and processed Hi-C, ATAC-seq, and RNA-seq data from this study are available from the NCBI Gene Expression Omnibus as series GSE126117 (reviewer token yxupgogepbsvrqp).

Received: 1 December 2019; Accepted: 15 December 2020

Published online: 12 January 2021

### References

- Hansen, A. S., Cattoglio, C., Darzacq, X. & Tjian, R. Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus* **9**, 20–32. <https://doi.org/10.1080/19491034.2017.1389365> (2018).
- Richmond, T. J. & Davey, C. A. The structure of DNA in the nucleosome core. *Nature* **423**, 145–150. <https://doi.org/10.1038/nature01595> (2003).
- Henikoff, S. & Smith, M. M. Histone variants and epigenetics. *Cold Spring Harb. Perspect. Biol.* **7**, a019364. <https://doi.org/10.1101/cshperspect.a019364> (2015).
- Achinger-Kawecka, J., Taberlay, P. C. & Clark, S. J. Alterations in three-dimensional organization of the cancer genome and epigenome. *Cold Spring Harb. Symp. Quant. Biol.* **81**, 41–51. <https://doi.org/10.1101/sqb.2016.81.031013> (2016).
- Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380. <https://doi.org/10.1038/nature11082> (2012).
- Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385. <https://doi.org/10.1038/nature11049> (2012).
- Lupianez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025. <https://doi.org/10.1016/j.cell.2015.04.004> (2015).
- Javierre, B. M. *et al.* Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384. <https://doi.org/10.1016/j.cell.2016.09.037> (2016).
- Gorkin, D. U., Leung, D. & Ren, B. The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell* **14**, 762–775. <https://doi.org/10.1016/j.stem.2014.05.017> (2014).
- Dekker, J. & Mirny, L. The 3D genome as moderator of chromosomal communication. *Cell* **164**, 1110–1121. <https://doi.org/10.1016/j.cell.2016.02.007> (2016).
- Stadhouders, R. *et al.* Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nat. Genet.* **50**, 238–249. <https://doi.org/10.1038/s41588-017-0030-7> (2018).
- Calderon, D. *et al.* Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat. Genet.* <https://doi.org/10.1038/s41588-019-0505-9> (2019).
- Gate, R. E. *et al.* Genetic determinants of co-accessible chromatin regions in activated T cells across humans. *Nat. Genet.* **50**, 1140–1150. <https://doi.org/10.1038/s41588-018-0156-2> (2018).
- Soskic, B. *et al.* Chromatin activity at GWAS loci identifies T cell states driving complex immune diseases. *Nat. Genet.* **51**, 1486–1493. <https://doi.org/10.1038/s41588-019-0493-9> (2019).
- Burren, O. S. *et al.* Chromosome contacts in activated T cells identify autoimmune disease candidate genes. *Genome Biol.* **18**, 165. <https://doi.org/10.1186/s13059-017-1285-0> (2017).
- McCarthy, D. J., Chen, Y. S. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297. <https://doi.org/10.1093/nar/gks042> (2012).
- Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47. <https://doi.org/10.1093/nar/gkv007> (2015).
- 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. <https://doi.org/10.1038/nmeth.2688> (2013).
- Avram, D. & Califano, D. The multifaceted roles of Bcl11b in thymic and peripheral T cells: Impact on immune diseases. *J. Immunol.* **193**, 2059–2065. <https://doi.org/10.4049/jimmunol.1400930> (2014).
- Su, S. *et al.* Glimma: Interactive graphics for gene expression analysis. *Bioinformatics* **33**, 2050–2052. <https://doi.org/10.1093/bioinformatics/btx094> (2017).
- Karin, M. & Greten, F. R. NF- $\kappa$ B: Linking inflammation and immunity to cancer development and progression. *Nat. Rev. Immunol.* **5**, 749–759. <https://doi.org/10.1038/nri1703> (2005).
- O’Shea, J. J., Lahesmaa, R., Vahedi, G., Laurence, A. & Kanno, Y. Genomic views of STAT function in CD4<sup>+</sup> T helper cell differentiation. *Nat. Rev. Immunol.* **11**, 239–250. <https://doi.org/10.1038/nri2958> (2011).
- Li, P. *et al.* BATF-JUN is critical for IRF4-mediated transcription in T cells. *Nature* **490**, 543–546. <https://doi.org/10.1038/nature11530> (2012).
- Murphy, T. L., Tussiwand, R. & Murphy, K. M. Specificity through cooperation: BATF-IRF interactions control immune-regulatory networks. *Nat. Rev. Immunol.* **13**, 499–509. <https://doi.org/10.1038/nri3470> (2013).
- Kurachi, M. *et al.* The transcription factor BATF operates as an essential differentiation checkpoint in early effector CD8<sup>+</sup> T cells. *Nat. Immunol.* **15**, 373–383. <https://doi.org/10.1038/ni.2834> (2014).
- Ernst, J. & Kellis, M. ChromHMM: Automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216. <https://doi.org/10.1038/nmeth.1906> (2012).
- Heger, A., Webber, C., Goodson, M., Ponting, C. P. & Lunter, G. GAT: A simulation framework for testing the association of genomic intervals. *Bioinformatics* **29**, 2046–2048. <https://doi.org/10.1093/bioinformatics/btt343> (2013).

28. Farh, K. K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343. <https://doi.org/10.1038/nature13835> (2015).
29. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195. <https://doi.org/10.1126/science.1222794> (2012).
30. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94**, 559–573. <https://doi.org/10.1016/j.ajhg.2014.03.004> (2014).
31. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461. <https://doi.org/10.1038/nature12787> (2014).
32. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501. <https://doi.org/10.1038/nbt.1630> (2010).
33. Lun, A. T. & Smyth, G. K. diffHic: A Bioconductor package to detect differential genomic interactions in Hi-C data. *BMC Bioinform.* **16**, 258. <https://doi.org/10.1186/s12859-015-0683-0> (2015).
34. Serra, F., Bàu, D., Filion, G. & Marti-Renom, M. A. Structural features of the fly chromatin colors revealed by automatic three-dimensional modeling. *bioRxiv*, 036764 (2016).
35. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336. <https://doi.org/10.1038/nature14222> (2015).
36. Yoshida, H. *et al.* The cis-regulatory atlas of the mouse immune system. *Cell* **176**, 897–912. <https://doi.org/10.1016/j.cell.2018.12.036> (2019).
37. Scott-Browne, J. P. *et al.* Dynamic changes in chromatin accessibility occur in CD8(+) T cells responding to viral infection. *Immunity* **45**, 1327–1340. <https://doi.org/10.1016/j.immuni.2016.10.028> (2016).
38. Bascom, G. & Schlick, T. Linking chromatin fibers to gene folding by hierarchical looping. *Biophys. J.* **112**, 434–445. <https://doi.org/10.1016/j.bpj.2017.01.003> (2017).
39. Fudenberg, G. *et al.* Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049. <https://doi.org/10.1016/j.celrep.2016.04.085> (2016).
40. Barutcu, A. R., Maass, P. G., Lewandowski, J. P., Weiner, C. L. & Rinn, J. L. A TAD boundary is preserved upon deletion of the CTCF-rich *Firre* locus. *Nat. Commun.* <https://doi.org/10.1038/s41467-018-03614-0> (2018).
41. Liao, Y., Smyth, G. K. & Shi, W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* **47**, e47. <https://doi.org/10.1093/nar/gkz114> (2019).
42. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930. <https://doi.org/10.1093/bioinformatics/btt656> (2014).
43. Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680. <https://doi.org/10.1016/j.cell.2014.11.021> (2014).
44. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140. <https://doi.org/10.1093/bioinformatics/btp616> (2010).
45. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* <https://doi.org/10.1186/gb-2010-11-3-r25> (2010).
46. Law, C. W., Chen, Y. S., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* <https://doi.org/10.1186/gb-2014-15-2-r29> (2014).
47. Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S. & Smyth, G. K. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann. Appl. Stat.* **10**, 946–963. <https://doi.org/10.1214/16-AOAS920> (2016).
48. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359. <https://doi.org/10.1038/nmeth.1923> (2012).
49. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> (2009).
50. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* <https://doi.org/10.1186/gb-2008-9-9-r137> (2008).
51. Lund, S. P., Nettleton, D., McCarthy, D. J. & Smyth, G. K. Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. *Stat. Appl. Genet. Mol. Biol.* <https://doi.org/10.1515/1544-6115.1826> (2012).
52. Chen, Y., Lun, A. T. & Smyth, G. K. From reads to genes to pathways: Differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Res* **5**, 1438. <https://doi.org/10.12688/f1000research.8987.2> (2016).
53. Wu, D. *et al.* ROAST: Rotation gene set tests for complex microarray experiments. *Bioinformatics* **26**, 2176–2182. <https://doi.org/10.1093/bioinformatics/btq401> (2010).
54. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842. <https://doi.org/10.1093/bioinformatics/btq033> (2010).
55. 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. <https://doi.org/10.1016/j.molcel.2010.05.004> (2010).
56. Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science* <https://doi.org/10.1126/science.aav1898> (2018).
57. Schmidt, E. M. *et al.* GREGOR: Evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics* **31**, 2601–2606. <https://doi.org/10.1093/bioinformatics/btv201> (2015).
58. Schep, A. N. *et al.* Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome Res.* **25**, 1757–1770. <https://doi.org/10.1101/gr.192294.115> (2015).
59. Johanson, T. M. *et al.* Genome-wide analysis reveals no evidence of trans chromosomal regulation of mammalian immune development. *PLoS Genet.* <https://doi.org/10.1371/journal.pgen.1007431> (2018).
60. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74. <https://doi.org/10.1038/nature11247> (2012).
61. Lun, A. T. & Smyth, G. K. csaw: A Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res.* **44**, e45. <https://doi.org/10.1093/nar/gkv1191> (2016).
62. Chen, Y., Pal, B., Visvader, J. E. & Smyth, G. K. Differential methylation analysis of reduced representation bisulfite sequencing experiments using edgeR. *F1000Res* **6**, 2055. <https://doi.org/10.12688/f1000research.13196.2> (2017).
63. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293. <https://doi.org/10.1126/science.1181369> (2009).
64. Sushi: Tools for visualizing genomics data (2015).
65. Garnier, S. viridis: Default Color Maps from 'matplotlib'. (2018).
66. Gu, Z. G., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849. <https://doi.org/10.1093/bioinformatics/btw313> (2016).

## Acknowledgements

We thank Dr. Stephen Wilcox (Genomics Hub, WEHI) for generating the sequencing data and Drs. Natasha Jansz and Marnie Blewitt for their advice on the ATAC-seq protocol. This work was supported by Australian

National Health and Medical Research Council (NHMRC) Program Grant (APP1037321), Research Fellowships (H.D.C, T.M.J, L.F, G.K.S, R.S.A, L.C.H) and Centre for Research Excellence Grant (APP 1078106) (N.G.B.). The work was made possible through Victorian State Government Operational Infrastructure Support and NHMRC Research Institute Infrastructure Support Scheme.

### Author contributions

L.C.H. and N.G.B. designed the project and wrote the manuscript. G.N. and T.M.J. performed the RNA-seq, ATAC-seq and Hi-C experiments. N.G.B., H.D.C., A.L.G., J.S. performed the computational analyses. G.K.S., R.S.A., E.B.S. and L.F. contributed to data interpretation and manuscript editing.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-020-80165-9>.

**Correspondence** and requests for materials should be addressed to L.C.H.

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

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021