Activation-induced re-organization of chromatin in human T cells

Naiara G. Bediaga^{1,2}, Hannah D. Coughlan^{1,2}, Timothy M. Johanson^{1,2}, Alexandra L. Garnham^{1,2}, Gaetano Naselli¹, Jan Schröder^{1,2}, Liam G. Fearnley^{1,2}, Esther Bandala-Sanchez^{1,2}, Rhys S. Allan^{1,2}, Gordon K. Smyth^{1,3}, Leonard C. Harrison^{1,2*}.

¹The Walter and Eliza Hall Institute of Medical Research, Parkville, 3052, Australia. ²Department of Medical Biology, ³School of Mathematics and Statistics, The University of Melbourne, Parkville, 3010, Australia.

*Corresponding author: Leonard C. Harrison (harrison@wehi.edu.au)

ABSTRACT

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 genome organization orchestrate gene expression in response to cell perturbation. To characterize coordinate changes at different levels of chromatin architecture, we analysed chromatin accessibility, chromosome conformation and gene expression after activation of human T cells. T cell activation led to widespread changes in chromatin interactions and accessibility that were mostly shared between CD4⁺ and CD8⁺ T cells. Differential chromatin interactions were associated with upregulation or downregulation of linked target genes. Moreover, activation was associated with the formation of shorter chromatin interactions, partitioning of topologically associating domains (TADs) and acquisition of new TAD boundaries characterized by higher nucleosome occupancy, and lower chromatin accessibility and gene expression. These findings render an integrated and multiscale characterization of activation-induced re-organization of chromatin architecture underlying gene transcription in human T cells.

KEYWORDS: Human, T cell, activation, chromatin, accessibility, remodelling, genomic interactions, TAD, gene expression

BACKGROUND

Mammalian genomes are folded into highly organized hierarchical structures linked to function at each level ^[1]. 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 ^[2-4]. At an intermediate level, the genome is organized into protein-mediated loops that facilitate interaction between pairs of genomic sites such as promoters and enhancers distant along the linear genome. At a higher level, the genome is organized into self-interacting chromatin, called topologically associating domains (TADs) ^[5,6]. Disruption of TAD boundaries has been associated with developmental defects ^[7] 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 ^[8] and gene expression ^[9-11] 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, 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 marked reorganization of chromatin structure at the level of accessibility (measured by ATAC-seq)^[12-14] and, separately, conformation (measured by promoter capture-Hi-C) ^[15]. The present study is the first to analyse genome-wide chromosome conformation in conjunction with chromatin accessibility and whole transcriptome expression, to understand how coordinated changes in higher- and lower-order chromatin structures are linked to gene expression in response to T cell activation.

RESULTS

Chromatin accessibility and 3D genomic interactions are immune cell-type specific

Naïve and activated CD4⁺ T cells, naïve and activated CD8⁺ T cells, and mature B cells were isolated from two healthy human male donors. Each cell lineage from each donor was profiled using assay for transposase-accessible chromatin using sequencing (ATAC-seq) to generate genome-wide maps of chromatin accessibility, in situ Hi-C to interrogate 3D genomic interactions and RNA-seq for whole transcriptome expression. The consistent use of two independent biological replicates across all three genomic technologies was considered an important aspect of the study, and all bioinformatics analyses used statistically rigorous methods that assessed statistical significance between the cell populations relative to biological variability between the individual donors.

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, 1B, 1C). We observed a strong enrichment of ATAC-seq reads at transcription start sites (TSSs) genome-wide, including promoter-TSSs of the cell type-specific expressed genes *CD4*, *CD8A*, *IFNG*, *BCL11B* and *MS4A1*, an indication of the quality of the data ^[16] (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 ^[17] (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 (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 different cell types (Fig. 1D), in keeping with cell-type specificity of chromatin architecture ^[8]. 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) ^[18] we observed that activation (dimension 1) accounted 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 architecture and interactions, direct comparison of activated CD4⁺ and CD8⁺ T cells showed only a small number of chromatin architecture differences, viz. 864 differentially accessible regions and 69 differential interactions, indicating that activation has a similar effect on chromatin remodelling in CD4⁺ and CD8⁺ T cells.

T cell activation leads to widespread changes in chromatin accessibility

Activation increased accessibility of 8,484 (7.5 %) and 7,883 (6.9 %) ATAC-peaks in CD4⁺ and CD8⁺ T cells, respectively, while 2,961 (2.6 %) and 3,184 (2.8 %) peaks lost accessibility (Fig. 2A, Table S1). Compared to accessibility, T cell activation altered a higher proportion of expressed genes, with 4,894 (33.4%) and 4,727 (32.3%) being upregulated and 4,931 (33.7%) and 4,611 (31.5) downregulated in CD4⁺ and CD8⁺ T cells, respectively, relative to the resting state (Fig. 2A). Interestingly, despite differences in their effector functions, CD4⁺ and CD8⁺ T cells shared a large proportion of activation-induced chromatin accessibility (Fig. 2B). Using annotations from the UCSC database in the TxDb.Hsapiens.UCSC.hg38.knownGene package (v3.4.6) and the ChromHMM model in the Roadmap Epigenomics Project ^[19], we annotated the activation-induced accessibility changes and calculated their enrichment over the universe of ATAC-seq peaks using GAT ^[20]. Overall, approximately 75 % of the activation-induced peaks were located in intronic and intergenic regions and were significantly depleted for promoter regions (FDR < 0.05). Likewise, activation-induced peaks were enriched for ChromHMM predicted enhancer and other regulatory elements and were depleted in heterochromatin repressive marks and promoter-TSS regions compared to the universe of ATAC-seq peaks (Fig. S2A).

Several studies have shown that disease-associated SNPs frequently reside in cell-specific regulatory elements^[21-23]. 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 (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

activation-responsive elements compared to non-responsive elements (Fig. S2B). Moreover, enrichment was more pronounced in the subset of peaks that contained predicted enhancers by chromHMM (Fig. S2B).

To gain further insight into the activation-responsive chromatin elements in human T cells, we analysed activation-specific peaks for enrichment of known transcription factor binding (TFB) sites. Activation-induced peaks for both CD4⁺ and CD8⁺ T cells were specifically enriched for DNA motifs recognized by TFs involved in T cell development, activation or proliferation ^[10,24-28]. These included members of the bZIP (BATF, FOSL1, ATF3), NFκB (NFκB-p50, NFκB-p65), Zf (EGR2, EGR1), STAT (Stat5, Stat1, Stat3) and IRF (IRF4) families (Fig. S2C). Most changes in TFB site accessibility were accompanied by increased expression of the relevant TFs including those of the bZIP family, further supporting the regulatory nature of these activation-induced elements.

As whole transcriptome sequencing without poly(A) pre-selection allowed us to detect enhancer RNAs^[15], we could identify the subset of eRNAs that overlapped the set of human enhancer regions defined by the FANTOM consortium ^[29]. 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 activation-induced 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 (Fig. S3A).

To further explore the relationship between the activation-induced accessibility changes and expression of their potential target genes we applied the Genomic Regions Enrichment of Annotations Tool (GREAT) ^[30]. Using GREAT we associated our set of activation-induced peaks with their putative target genes and 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 (Fig. S3B). The set of target genes defined by GREAT comprised immune system-associated genes known to be upregulated upon T cell activation, including IL23R, IL12RB2, IL1R2, TNFRSF8, IL2, IL21 and IL13 among

others. Furthermore, functional enrichment analysis using GREAT showed that the activation-induced 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 revealed 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.

T cell activation induces extensive changes in chromatin interactions

Chromatin interactions bring distal regulatory elements and genes into spatial proximity to regulate transcription. To determine differential chromatin interactions between samples we used the diffHic pipeline^[31] which was designed for the detection of differential interactions from Hi-C data. It partitions the genome into bins and counts the number of read pairs mapping to each pair of bins. Each bin pair is then tested for significant changes in intensity between libraries, using the methods in the edgeR package^[32]. Upon activation, in CD4⁺ and CD8⁺ T cells, respectively, at 25kb resolution, we identified 17,965 and 15,496 differential chromatin interactions (DIs) with a significant increase in interaction counts which we refer to as 'gain' and 12,434 and 11,376 DIs with a significant decrease in interaction counts that we refer to as 'loss' (Tables S3A and S3B). Then to examine the relationship between genomic accessibility and interactions we selected DIs containing at least one activation-induced regulatory element in one anchor and a promoter of an expressed gene in the other. We found that accessibility was significantly increased for 'gain' interactions compared to 'loss' interactions following activation of either CD4⁺ and CD8⁺ T cell (Fig. 3A). Furthermore, when we examined if 'gain' or 'loss' chromatin interactions were associated with gene up- or down-regulation we found that gain interactions were linked to a higher percentage of genes upregulated while 'loss' interactions were linked to a higher percentage of genes downregulated (Chi-square: 35.0, p<0.05) (Fig. 3B). Thus, 'gain'

and ‘loss’ interactions are linked with upregulation and downregulation, respectively, 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 ^[1]). To examine TAD plasticity we first segmented the chromatin of naïve and activated T cells, and B cells, into TADs using TADbit ^[33]. Naïve CD4⁺ T, CD8⁺ T and B cells had 1,727, 1,750 and 1,602 TADs with a mean size of 1.40, 1.38 and 1.51 Mb, respectively, whereas activated CD4⁺ and CD8⁺ T cells had 2,772 and 2,863 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 compared to naïve T cells (TAD sizes: $t = 11.3$ and 10.7 for nCD4 vs. aCD4 and nCD8 vs. aCD8, respectively; $P\text{-value} < 2e-16$) (Fig. 4A). On average 62% of the TADs called in resting CD4⁺ T cells intersected with TADs in resting CD8⁺ T cells, with >75% reciprocal region overlap. Interestingly, this fell to 43.5 % when comparing the TAD overlap between resting and activated T cells. Further analysis of the TAD intersects revealed that while only 24% of the TADs called in resting CD4⁺ T cells overlapped two or more TADs in resting CD8⁺ T cells this increased to 44% when analysing the TAD overlap between resting and activated T cells (Fig. S4), thus suggesting that T cell-activation induces partitioning of TADs into finer chromatin domains.

Furthermore, we also observed that, analogous to TAD partitioning, chromatin interactions strengthened upon activation were on average significantly shorter than those that were weakened ($t = 9.3$; $P\text{-value} < 2e-16$), i.e. activation preferentially promoted 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 larger (Fig. 4C), thus suggesting that chromosome domains play a pivotal role in shaping the intra-TAD

chromatin loops. Moreover, TAD partitioning and rearrangement of chromatin interactions overlapped markers of T cell activation including CDH3, CDH1, IL23R and IL12RB2 (Fig. 5, Fig. S5).

We next sought to identify features associated with the strengthening and weakening of TAD boundaries. We defined differential TAD boundaries, using the diffHic package^[31]. To identify these boundaries, we use the directionality index defined by Dixon et al.^[5]. For each region the directionality index was computed by counting the number of read pairs mapped between the target locus and an ‘upstream’ interval with higher genomic coordinates; counting the number of read pairs mapped between the target locus and a ‘downstream’ interval with lower genomic coordinates; and taking the normalized difference of the counts. We then used the statistical framework of edgeR to identify regions (boundaries) where upon activation the size or direction of the directionality statistic changes. Upon activation, 5,278 and 4,705 boundaries became significantly stronger and 6,600 and 7,312 weaker in CD4⁺ and CD8⁺ T cells, respectively. We then calculated the distribution of chromatin accessibility, nucleosome occupancy and gene expression for regions containing differential TAD boundaries in both resting and activated T cells. Interestingly, we observed that TAD boundaries that became significantly stronger upon activation were associated on average with lower chromatin accessibility, higher nucleosome occupancy and lower protein-coding gene expression (Fig. 6). In summary, T cell activation leads to global rearrangement of TAD boundaries, partitioning of TADs and disruption of long-range chromatin interactions, with strengthening of TAD boundaries associated with lower chromatin accessibility, higher nucleosome occupancy and lower gene expression.

Chromosome A/B compartments are not altered substantially by T cell activation

Lineage differentiation has been associated with rearrangement of chromosome A/B compartments^[34]. However, in response to T cell activation we observed only minor changes in the A/B compartments, with less than 4% of the genome switching compartments (Fig. 7A, 7B). 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 (Figure 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

In this study, we analysed chromatin accessibility, Hi-C chromatin interactions and the whole transcriptome in human CD4⁺, CD8⁺ T cells under resting (naïve) and activated conditions, and in B cells. Our study made use of biological replicates throughout, so that changes in chromatin accessibility, interaction intensity or gene expression were assessed using statistically rigorous methods relative to the biological variation observed between human individuals. Activation of T cells was associated with multi-level remodelling of chromatin structure, that was comparable between CD4⁺ and CD8⁺ T cells. It entailed 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. What we demonstrate in human T cells parallels what has been described in other systems^[12-14,35,36], namely that changes in chromatin accessibility in response to external cues identify active regulatory regions likely to be critical for transcriptional responses. Furthermore, we show that a T cell activation uncovers many additional regulatory regions likely to be associated with the immune response. T cell activation also induced thousands of differential chromatin interactions that were linked to the upregulation or down regulation of the corresponding target genes. At a higher level of chromosome architecture, T cell activation was linked with genome-wide rearrangement of TAD boundaries. Previous studies have either characterized activation-responsive chromatin accessibility^[13,14,35,36] or 3D organization of the genome^[8,13,15] in mouse and human immune cells. However, to the best of our knowledge, no study has previously integrated chromatin accessibility, nucleosome occupancy and 3D genome structure in relation to gene transcription in response to activation of human T cells. Ours is also the first study to investigate activation-induced

changes in chromatin architecture in the two main subpopulations of T cells, CD4⁺ and CD8⁺ T cells, enabling us to determine if their activation is linked to similar genome control mechanisms.

Although TADs were reported to be largely invariant across species and cell types ^[5,6], recent evidence ^[1] indicates they are more dynamic. In fact, we found that T cell activation was associated with global rearrangement of chromatin interactions involving changes in the strength of TAD boundaries and a finer compartment structure as indicated by the smaller size of TAD domains. 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 ^[37]. Polymer simulations of the chromatin fiber suggest that regions of “stiffness” can act as local insulators to decrease interaction frequencies ^[38]. 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 ^[39]. Our results in primary human cells support the view that long-range chromosomal organization and nucleosome occupancy are molecularly coupled and that regions of TAD boundary formation are primed before activation as indicated by the lower accessibility and higher nucleosome occupancy and 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 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-induced regulatory regions that will aid interpretation of chromatin accessibility data in human T cells.

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 CONTRIBUTION

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 INTEREST

The authors declare no financial or non-financial competing interests.

299 METHODS

300 **Blood collection, cell subset isolation and T cell activation** Heparinised venous blood was collected at
 301 09.00 h from two healthy young adult males, who gave written informed consent. The research was
 302 carried out in accordance with the principles of the Declaration of Helsinki and the guidelines of Nature
 303 journals and was approved by the Walter and Eliza Hall Institute of Medical Research Human Research
 304 Ethics Committee (application 88/03). Peripheral blood mononuclear cells (PBMCs) were purified by
 305 Ficoll-Hypaque gradient centrifugation and cryopreserved in liquid N₂. Thawed PBMCs, ≥92% viable by
 306 acridine orange-ethidium bromide staining, were stained with anti-human αβ TCR (eBioscience, clone
 307 IP26, cat. no. 46-9986-42), anti-human CD4 (BD Pharmingen, clone RPA-T4, cat. no. 555349), anti-
 308 human CD45RA (BD Pharmingen, clone 5H9, cat. no. 556626), anti-human CD25 (BD Pharmingen, clone
 309 M-A251, cat. no. 557741), anti-human CD14 (BioLegend, clone 63D3, cat. no. 367104), anti-human
 310 CD16 (BD Pharmingen, clone 3G8, cat. no. 557758), anti-human HLA-DR (eBioscience, clone L243, cat.
 311 no. 48-9952-42), and anti-human CD19 (BioLegend, clone HIB19, cat. no. 302238). Naïve CD4⁺ T cells
 312 (CD14⁻CD16⁻TCRαβ⁺CD4⁺CD45RA⁺CD25⁻) and CD8⁺ T cells (CD14⁻CD16⁻TCRαβ⁺CD4⁺CD45RA⁺CD25⁻),
 313 and B cells (TCRαβ⁻, HLA-DR⁺CD19⁺), were flow-sorted on a FACSARIA (BD Biosciences). Naïve CD4⁺ and
 314 CD8⁺ T cells were cultured in Iscove's Modified Dulbecco's Medium containing 5% pooled, heat-
 315 inactivated, human serum, 100 nM non-essential aminoacids, 2mM of glutamine and 50 uM 2-
 316 mercaptoethanol (IP5) medium. T cells were activated through the T cell receptor subunit CD3 and the
 317 co-stimulator CD28 by the addition of Human T-Activator CD3/CD28 Dynabeads at a 1:2 bead:cell ratio
 318 (Life Technologies, cat. no. 111.31D) for 72 hours.

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

TapeStation. Total RNA-Seq libraries were 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^[40]. The number of fragments overlapping each Entrez gene were summarised using featureCounts^[41] 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).

ATAC-seq ATAC-seq was performed as previously described^[16]. 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 MgCl₂ and 0.03% Tween20). Immediately after lysis, nuclei were spun out at 500 g for 8 minutes at 4°C and the supernatant carefully removed. Nuclei were resuspended with Tn5 transposase reaction mix (25 µl 2X TD buffer, 2.5 µl Tn5 transposase, and 22.5 µl nuclease-free water) (Nextera DNA Library Prep Kit (Illumina, cat. no. FC-121-1030)). The transposase reaction was performed at 37°C for 30 minutes, 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 75bp paired-end reads.

In situ Hi-C In situ HiC was performed as previously published^[42]. Primary immune cell libraries were generated in biological duplicates. Libraries were sequenced on an Illumina NextSeq 500 to produce 81bp paired end reads. Between 129 million 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^[32] and limma v3.34.9^[43] 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)^[44]. All counts were then transformed to log2-CPM with associated precision weights using voom^[45]. Differential expression between all cell types was assessed

using linear models and robust empirical Bayes moderated t-statistics ^[46]. 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 naive 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 log2-CPM with precision weights using voom. Differential expression of the regions was then evaluated between the activated and naïve CD4⁺ and CD8⁺ 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) ^[47]. For each sample, mitochondrial reads, unmapped reads and low mappability (<30) reads were filtered out using Samtools (v1.6) function “view” ^[48]. After filtering, we had a median of 80 million (MAD +/- 13 million) reads per sample. Filtered ATAC-seq reads from naive and activated CD4⁺ and CD8⁺ 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) ^[49], 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 ^[41]. 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^[50] and limma v3.34.9^[43] software packages. The TMM method was applied to normalize compositional differences between libraries^[44]. 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 likelihood ratio tests (LRT) in edgeR. 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% and $\text{abs}(\log\text{FC}) \geq 1.5$ were defined as differentially accessible regions. Concordance between activation profiles was assessed by gene set tests using fry function in edgeR^[51].

Enrichment of transcription factor binding motifs The Homer suite v4.10^[52] 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', 3', promoter-TSS, exonic, intronic, TTS or intergenic using the Homer suite annotatePeaks.pl function and the default setting. We also found all instances of TFBSs with altered accessibility upon activation in the ATAC-seq peaks using the annotatePeaks.pl function, and options -m. Chromatin state(s) of the peaks were annotated using the ChromHMM calls for naive CD4⁺ and T CD8⁺ T cells and B cells from the Roadmap Epigenomics Project with BEDtools (function intersectBed)^[53]. Functionally-related subclasses were aggregated as follows: Promoter-TSS (TssA, PromU, PromD1 and PromD2), transcribed (Tx5', Tx, Tx3', TxWk), enhancers and other regulatory elements (TxReg, TxEnh5', TxEnh3', TxEnhW, EnhA1, EnhA2, EnhAF, EnhW1, EnhW2 and EnhAc), repetitive elements (ZNF/rep), heterochromatin/repressive (Het, ReprPC, Quies) and others (DNase, PromP, PromBiv).

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^[54]. The significance threshold was set up at FDR below 5%.

Prediction of transcriptional target genes. To identify potential targeted genes for the activation-induced peaks we performed enrichment analysis of gene annotations in the proximity of the activation-induced peaks using Genomic Regions Enrichment of Annotations Tool (GREAT) ^[30] 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 500kb). 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-induced peaks and classify them as potential transcription target genes.

Enrichment of GWAS loci GREGOR v1.4.0 ^[55] 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=1Mb and minimum neighbor number= 500. For the GWAS SNPs, we created an updated version of Table S2 in Maurano et al ^[22]. 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 (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 ^[22], 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 dyad centers were identified in the “nucmap_combined.bed” file, while nucleosome occupancy was defined with the “occ.bedgraph” files using the NucleoATAC v0.3.2.1 package ^[56].

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 ^[31]. Data were pre-processed and artefacts removed as per Johanson et al. 2018 ^[57].

Differential interaction (DI) analysis DIs between all five libraries were detected using the diffHic package ^[31] 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 Mbol 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 ^[58] 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 6-fold 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 25kbp) 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 ^[59,60] of the edgeR package (v3.20.9), which assesses statistical significance relative to biological variation between the replicate libraries. 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 ^[61], 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 ^[46] 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 25kbp) 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 ^[62]. Clustered DIs were used to report the number of differences between the libraries. The 25kbp 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 were detected with the TADbit v0.2.0.5 python based software ^[33]. Read pairs were counted into 50 kbp bin pairs (with bin boundaries rounded up to the nearest Mbol 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 ^[62]. Replicate contact matrices were summed. TADs were detected for each chromosome with the function find tad of the TADbit software. Only TADs boundaries with a score of 7 or higher were included in the results.

Detection of differential TAD boundaries Differential TAD boundaries between all five libraries were detected using the diffHic package ^[31]. To identify boundaries, the directionality index ^[5] for each genomic region was determined by counting the number of read pairs mapped between the genomic region 1 Mbp upstream and separately 1Mbp downstream then taking the normalized difference of the counts. This produces a log-fold-change for each sample at each genomic region. Differential TAD boundaries were determined to be genomic regions where the size or the direction of the logFC between up and down stream counts changed significantly between conditions. The domainDirections function was used to determine counts for 100 kbp genomic regions for 1 Mbp up- and downstream. A DGEList was constructed from the counts and library size. Low average abundance regions (calculated across all samples) of less than 0.8 were removed. Tests for differential TAD boundaries were performed again using the QL framework of the edgeR package as described in the differential interaction analysis section. A design matrix was constructed with cell type specific coefficients for the log-fold change between up- and downstream counts for each cell type. It also contains library specific and individual blocking factors. A P-value was computed for each 100kbp region using the QL F-test and adjusted for multiple testing using the Benjamini-Hochberg method (FDR below 5%). A differential TAD boundary was considered to be strengthened from one cell type to another if the absolute value of the log-fold change between up- and downstream has increased.

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. ^[63] using the HOMER HiC pipeline ^[52].

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 50000 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 0.6 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^[43] applied to the filtered and normalized log2-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^[64]. The inferno color palette from the viridis v0.5.1 package^[65] 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 access 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.

Data intersection Intersection between pairs of the genomic regions was performed using bedtools intersect (v2.19.1)^[53].

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.

FIGURES

Figure 1. Chromatin structure and gene expression are immune cell type-specific. Normalized read coverage plots of (A) ATAC-seq and (B) 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⁺) and activated CD8⁺ (nCD8⁺) T cells. (C) *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, while the bottom Hi-C matrix shows data from B cells. Color scale indicates number of reads per bin pair. (D) Multidimensional scaling (MDS) plots of log-CPM values over dimensions 1 and 2 with samples coloured by groups 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) log₂-fold-change for the 500, 5,000 and 50,000 genes (RNA-seq), peaks (ATAC-seq) or bins (Hi-C) most divergent between each pair of samples.

Figure 2. T cell activation induces genome-wide changes in chromatin and gene expression that are comparable between CD4⁺ and CD8⁺ T cells. (A) Volcano plots showing the effect of T cell activation for ATAC-seq (top) and RNA-seq (bottom), where the X-axis represents the log₂ fold-change and the Y-axis the significance ($-\log_{10}$ P-value based on moderated t-statistics) for the activated and resting comparisons. Differentially accessible regions (FDR < 0.05 and $|\log_2(\text{FC})| \geq 1.5$) and differentially expressed genes (FDR < 0.05) are coloured in red. (B) Barcode enrichment plot showing strong correlation of the CD4⁺ T activation-induced accessibility signature in the CD8⁺ T cell activation-induced accessibility. Regions are ordered on the plot from right to left from most positive log₂FC to most negative upon CD8⁺ T activation (x axis). Vertical bars indicate regions significantly up (red) or down (blue) upon CD4⁺ cell T activation. Significant enrichment was assessed by fry gene set tests.

547

548

549 **Figure 3. T cell activation induces extensive changes in gene-regulatory interaction** (A) Box plots
550 showing chromatin accessibility at regions of differential interaction – ‘loss’, ‘unchanged’ or ‘gain’ -
551 following T cell activation, at 25 kbp resolution (FDR<0.05). Shown are the median (central horizontal
552 line), interquartile range (boxes), values of the upper and lower quartiles (whiskers), outliers beyond
553 1.5 IQR (circles). Statistical comparisons were made with the unpaired Wilcoxon test: ***P-
554 value<0.0001. (B) Pie chart showing distribution of downregulated, upregulated or unchanged target
555 genes linked to ‘loss’ (top) or ‘gain’ (bottom) of regulatory interactions, for CD4⁺ (left) and CD8⁺ (right)
556 T cells.

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⁺ and CD8⁺ T cells were merged as they showed a similar pattern. (C) Pie chart showing distribution of Loss and Gain chromatin interactions in TADs that become smaller in size (left) or remained unchanged/became larger in size (right) upon T cell activation. Statistical comparisons were made with the unpaired Wilcoxon test: ***P-value<0.0001

Figure 5. T cell activation results in partitioning of genome topology. *In-situ* Hi-C contact matrices plotted at 100 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 indicates number of reads per bin pair. Locations of the TADs as called by TADbit at these regions are indicated below each contact matrix. Significant differential interactions at 100 kbp (FDR<0.05) are represented by arches where the z-score is $-\log_{10}(\text{p-value})$. Red and blue lines represent gain and loss of interactions, respectively, in response to activation. The zoomed inset regions show *in-situ* Hi-C contact matrices, significant differential interactions at 25 kbp resolution (FDR<0.05) and RNA sequencing coverage plots at the CDH3 and CDH1 loci. Turquoise and black coverage plots represent activated and resting T cells, respectively.

Figure 6. Increased strength in TAD boundary and chromatin interactions upon T cell activation is associated with nucleosome occupancy and gene expression Violin plots showing the distribution of (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⁺ and CD8⁺ T cells. Violin plots show median (horizontal line), interquartile range (boxes), values of the upper and lower quartiles (whiskers), outliers (circles) and kernel density estimation. Statistical comparisons were made with the unpaired Wilcoxon test: ***P-value<0.0001.

Figure 7. Chromosome A/B compartments are not altered substantially 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 naive 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. ***P-value<0.0001.

605 SUPPLEMENTARY FILES

606

607 **Figure S1. Chromatin structure and gene expression are immune cell type-specific** (A) *In-situ* Hi-C
608 contact matrices for a 2.5 Mb region on chromosome 2 that includes the BCL11A locus at 50 kbp
609 resolution. The top four Hi-C matrices display data from resting and activated T cells, while the bottom
610 Hi-C matrix shows data from B cells. Color scale indicates number of reads per bin pair. (B) Bar plots
611 showing the expression of BCL11A across resting and activated CD4⁺ and CD8⁺ T cells, and B cells.

Figure S2. T cell activation-induced accessibility changes are enriched for enhancers and other regulatory elements and GWAS SNPs. (A) Bar plots showing the distribution of the predicted chromatin states (ChromHMM) in subsets of peaks. (B) Bar plots showing enrichment of GWAS SNPs in subsets of ATAC-seq peaks. Shown are the relative enrichment of GWAS (observed/ expected number of overlapping SNPs) (Y axis) for each SNP category (X axis). Red asterisks denote SNP categories for which subsets of ATAC-seq peaks showed significant enrichment by GREGOR.(C) Bar plots showing the top-ranked TFBS motifs that are enriched in activation-induced accessibility changes in CD4⁺ and CD8⁺ T cells, where the X-axis represent the significance of the enrichment ($-\log_{10}$ P-value, cumulative binomial test), and Y-axis each of top-ranked motifs.

Figure S3. Variation of chromatin accessibility at the activation-induced peaks is associated with changes in expression of eRNAs and target genes. Distribution of the expression of (A) eRNAs and (B) predicted enhancer target genes across the activation-induced peaks that gained and lost accessibility is represented as box plots (central bar represents the median with boxes indicating interquartile range, values of the upper and lower quartiles are represented in with whiskers and outliers with circles). Statistical comparisons were made with the unpaired Wilcoxon test: ***P-value<0.0001.

Figure S4. T cell activation results in partitioning of genome topology. Pie chart representation of the distribution of (A) TADs in resting CD4⁺ T cells overlapping 0,1, 2 or ≥ 3 TADs in resting CD8⁺ T cells, (B) TADs in resting CD8⁺ T cells overlapping 0,1, 2 or ≥ 3 TADs in resting CD4⁺ T cells (C) TADs in resting CD4⁺ T cells overlapping 0,1, 2 or ≥ 3 TADs in activated CD4⁺ T cells, and (D) TADs in resting CD8⁺ T cells overlapping 0,1, 2 or ≥ 3 TADs in activated CD8⁺ T cells.

Figure S5. T cell activation results in partitioning of genome topology. In-situ Hi-C contact matrices at 100kbp in resting and activated T cells over genes that have been associated with T cell activation such as IL23R and IL12RB2. Colour scale indicates number of reads per bin pair. Locations of the TADs as called by TADbit at these regions are indicated below each contact matrix. Significant differential interactions at 100 kbp (FDR<0.05) are represented by arches where the z-score is $-\log_{10}(\text{p-value})$. Red and blue lines represent gain and loss of interactions, respectively, in response to activation. The zoomed inset regions show in-situ Hi-C contact matrices, significant differential interactions at 25 kbp resolution (FDR<0.05) and RNA sequencing coverage plots for the IL23R and IL12RB2 loci. Turquoise and black coverage plots represent activated and resting T cells respectively.

664 Table S1: Differential accessibility analysis comparing naïve to activated CD4⁺ and CD8⁺ T cells.

665 Table S2: List of disease- and trait-associated variants

666 Tables S3A and S3B: (A) Differential interactions comparing naïve to activated CD4⁺ (A) and CD8⁺ (B) T
667 cells.

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

REFERENCES

- 1 Hansen, A. S., Cattoglio, C., Darzacq, X. & Tjian, R. Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus* **9**, 20-32, doi:10.1080/19491034.2017.1389365 (2018).
- 2 Richmond, T. J. & Davey, C. A. The structure of DNA in the nucleosome core. *Nature* **423**, 145-150, doi:10.1038/nature01595 (2003).
- 3 Henikoff, S. & Smith, M. M. Histone variants and epigenetics. *Cold Spring Harb Perspect. Biol.* **7**, a019364, doi:10.1101/cshperspect.a019364 (2015).
- 4 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, doi:10.1101/sqb.2016.81.031013 (2016).
- 5 Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380, doi:10.1038/nature11082 (2012).
- 6 Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385, doi:10.1038/nature11049 (2012).
- 7 Lupianez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025, doi:10.1016/j.cell.2015.04.004 (2015).
- 8 Javierre, B. M. et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369-1384 e1319, doi:10.1016/j.cell.2016.09.037 (2016).
- 9 Gorkin, D. U., Leung, D. & Ren, B. The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell* **14**, 762-775, doi:10.1016/j.stem.2014.05.017 (2014).
- 10 Dekker, J. & Mirny, L. The 3D genome as moderator of chromosomal communication. *Cell* **164**, 1110-1121, doi:10.1016/j.cell.2016.02.007 (2016).
- 11 Stadhouders, R. et al. Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nat. Genet.* **50**, 238-249, doi:10.1038/s41588-017-0030-7 (2018).

715 12 Calderon, D. et al. Landscape of stimulation-responsive chromatin across diverse human
716 immune cells. *Nat. Genet.* **51**, 1494–1505, doi:10.1038/s41588-019-0505-9 (2019).

717 13 Gate, R. E. et al. Genetic determinants of co-accessible chromatin regions in activated T cells
718 across humans. *Nat. Genet.* **50**, 1140-1150, doi:10.1038/s41588-018-0156-2 (2018).

719 14 Soskic, B. et al. Chromatin activity at GWAS loci identifies T cell states driving complex
720 immune diseases. *Nat. Genet.* **51**, 1486-1493, doi:10.1038/s41588-019-0493-9 (2019).

721 15 Burren, O. S. et al. Chromosome contacts in activated T cells identify autoimmune disease
722 candidate genes. *Genome Biol.* **18**, 165, doi:10.1186/s13059-017-1285-0 (2017).

723 16 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of
724 native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins
725 and nucleosome position. *Nat. Methods* **10**, 1213-1218, doi:10.1038/nmeth.2688 (2013).

726 17 Avram, D. & Califano, D. The multifaceted roles of Bcl11b in thymic and peripheral T cells:
727 impact on immune diseases. *J Immunol* **193**, 2059-2065, doi:10.4049/jimmunol.1400930 (2014).

728 18 Su, S. et al. Glimma: interactive graphics for gene expression analysis. *Bioinformatics* **33**,
729 2050-2052, doi:10.1093/bioinformatics/btx094 (2017).

730 19 Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization.
731 *Nat. Methods* **9**, 215-216, doi:10.1038/nmeth.1906 (2012).

732 20 Heger, A., Webber, C., Goodson, M., Ponting, C. P. & Lunter, G. GAT: a simulation framework
733 for testing the association of genomic intervals. *Bioinformatics* **29**, 2046-2048,
734 doi:10.1093/bioinformatics/btt343 (2013).

735 21 Farh, K. K. et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants.
736 *Nature* **518**, 337-343, doi:10.1038/nature13835 (2015).

737 22 Maurano, M. T. et al. Systematic localization of common disease-associated variation in
738 regulatory DNA. *Science* **337**, 1190-1195, doi:10.1126/science.1222794 (2012).

739 23 Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of
740 18 human traits. *Am. J. Hum. Genet.* **94**, 559-573, doi:10.1016/j.ajhg.2014.03.004 (2014).

741 24 Karin, M. & Greten, F. R. NF-kappaB: linking inflammation and immunity to cancer
742 development and progression. *Nat. Rev. Immunol.* **5**, 749-759, doi:10.1038/nri1703 (2005).

743 25 O'Shea, J. J., Lahesmaa, R., Vahedi, G., Laurence, A. & Kanno, Y. Genomic views of STAT
744 function in CD4+ T helper cell differentiation. *Nat. Rev. Immunol.* **11**, 239-250, doi:10.1038/nri2958
745 (2011).

746 26 Li, P. et al. BATF-JUN is critical for IRF4-mediated transcription in T cells. *Nature* **490**, 543-546,
747 doi:10.1038/nature11530 (2012).

748 27 Murphy, T. L., Tussiwand, R. & Murphy, K. M. Specificity through cooperation: BATF-IRF
749 interactions control immune-regulatory networks. *Nat. Rev. Immunol.* **13**, 499-509,
750 doi:10.1038/nri3470 (2013).

751 28 Kurachi, M. et al. The transcription factor BATF operates as an essential differentiation
752 checkpoint in early effector CD8+ T cells. *Nat Immunol* **15**, 373-383, doi:10.1038/ni.2834 (2014).

753 29 Andersson, R. et al. An atlas of active enhancers across human cell types and tissues. *Nature*
754 **507**, 455-461, doi:10.1038/nature12787 (2014).

755 30 McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat.*
756 *Biotechnol.* **28**, 495-501, doi:10.1038/nbt.1630 (2010).

757 31 Lun, A. T. & Smyth, G. K. diffHic: a Bioconductor package to detect differential genomic
758 interactions in Hi-C data. *BMC Bioinformatics* **16**, 258, doi:10.1186/s12859-015-0683-0 (2015).

759 32 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential
760 expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140,
761 doi:10.1093/bioinformatics/btp616 (2010).

762 33 Serra, F., Baù, D., Filion, G. & Marti-Renom, M. A. Structural features of the fly chromatin
763 colors revealed by automatic three-dimensional modeling. *bioRxiv*, 036764 (2016).

764 34 Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation.
765 *Nature* **518**, 331-336, doi:10.1038/nature14222 (2015).

766 35 Yoshida, H. et al. The cis-regulatory atlas of the mouse immune system. *Cell* **176**, 897-912
767 e820, doi:10.1016/j.cell.2018.12.036 (2019).

768 36 Scott-Browne, J. P. et al. Dynamic changes in chromatin accessibility occur in CD8(+) T Cells
769 Responding to Viral Infection. *Immunity* **45**, 1327-1340, doi:10.1016/j.immuni.2016.10.028 (2016).

770 37 Bascom, G. & Schlick, T. Linking chromatin fibers to gene folding by hierarchical looping.
771 *Biophys. J.* **112**, 434-445, doi:10.1016/j.bpj.2017.01.003 (2017).

772 38 Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**,
773 2038-2049, doi:10.1016/j.celrep.2016.04.085 (2016).

774 39 Barutcu, A. R., Maass, P. G., Lewandowski, J. P., Weiner, C. L. & Rinn, J. L. A TAD boundary is
775 preserved upon deletion of the CTCF-rich Firre locus. *Nat. Commun.* **9**, doi:ARTN 1444
776 10.1038/s41467-018-03614-0 (2018).

777 40 Liao, Y., Smyth, G. K. & Shi, W. The R package Rsubread is easier, faster, cheaper and better
778 for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* **47**, e47,
779 doi:10.1093/nar/gkz114 (2019).

780 41 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for
781 assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930,
782 doi:10.1093/bioinformatics/btt656 (2014).

783 42 Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of
784 chromatin looping. *Cell* **159**, 1665-1680, doi:10.1016/j.cell.2014.11.021 (2014).

785 43 Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and
786 microarray studies. *Nucleic Acids Res.* **43**, e47, doi:10.1093/nar/gkv007 (2015).

787 44 Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression
788 analysis of RNA-seq data. *Genome Biology* **11**, doi:ARTN R25
789 10.1186/gb-2010-11-3-r25 (2010).

790 45 Law, C. W., Chen, Y. S., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model
791 analysis tools for RNA-seq read counts. *Genome Biology* **15**, doi:ARTN R29

792 10.1186/gb-2014-15-2-r29 (2014).

793 46 Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S. & Smyth, G. K. Robust hyperparameter
794 estimation protects against hypervariable genes and improves power to detect differential
795 expression. *Ann. Appl. Stat.* **10**, 946-963, doi:10.1214/16-AOAS920 (2016).

796 47 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**,
797 357-359, doi:10.1038/nmeth.1923 (2012).

798 48 Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-
799 2079, doi:10.1093/bioinformatics/btp352 (2009).

800 49 Zhang, Y. et al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, doi:ARTN R137
801 10.1186/gb-2008-9-9-r137 (2008).

802 50 McCarthy, D. J., Chen, Y. S. & Smyth, G. K. Differential expression analysis of multifactor RNA-
803 Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288-4297,
804 doi:10.1093/nar/gks042 (2012).

805 51 Wu, D. et al. ROAST: rotation gene set tests for complex microarray experiments.
806 *Bioinformatics* **26**, 2176-2182, doi:10.1093/bioinformatics/btq401 (2010).

807 52 Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-
808 regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576-589,
809 doi:10.1016/j.molcel.2010.05.004 (2010).

810 53 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic
811 features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

812 54 Heger, A., Webber, C., Goodson, M., Ponting, C. P. & Lunter, G. GAT: a simulation framework
813 for testing the association of genomic intervals. *Bioinformatics* **29**, 2046-2048,
814 doi:10.1093/bioinformatics/btt343 (2013).

815 55 Schmidt, E. M. et al. GREGOR: evaluating global enrichment of trait-associated variants in
816 epigenomic features using a systematic, data-driven approach. *Bioinformatics* **31**, 2601-2606,
817 doi:10.1093/bioinformatics/btv201 (2015).

56 Schep, A. N. et al. Structured nucleosome fingerprints enable high-resolution mapping of
chromatin architecture within regulatory regions. *Genome Res.* **25**, 1757-1770,
doi:10.1101/gr.192294.115 (2015).

57 Johanson, T. M. et al. Genome-wide analysis reveals no evidence of trans chromosomal
regulation of mammalian immune development. *Plos Genet.* **14**, doi:ARTN e1007431
10.1371/journal.pgen.1007431 (2018).

58 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature*
489, 57-74, doi:10.1038/nature11247 (2012).

59 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. **11**, doi:10.1515/1544-6115.1826 (2012).

60 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, doi:10.12688/f1000research.8987.2 (2016).

61 McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-
Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288-4297,
doi:10.1093/nar/gks042 (2012).

62 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, doi:10.1093/nar/gkv1191 (2016).

63 Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding
principles of the human genome. *Science* **326**, 289-293, doi:10.1126/science.1181369 (2009).

64 Sushi: Tools for visualizing genomics data (2015).

65 Garnier, S. viridis: Default Color Maps from 'matplotlib'. (2018).