# Enhancer connectome in primary human cells reveals target genes of disease-associated DNA elements

Maxwell R. Mumbach[1,2,12], Ansuman T. Satpathy[1,3,12], Evan A. Boyle[1,2,12], Chao Dai[1,12], Benjamin G. Gowen[4,5], Seung Woo Cho[1], Michelle L. Nguyen[6], Adam J. Rubin[7], Jeffrey M. Granja[1,2], Katelynn R. Kazane[4,5], Yuning Wei[1], Trieu Nguyen[8], Peyton G. Greenside[2], M. Ryan Corces[1], Josh Tycko[2], Dimitre R. Simeonov[6,9], Nabeela Suliman[2], Rui Li[1], Jin Xu[1], Ryan A. Flynn[1], Anshul Kundaje[2], Paul A. Khavari[7], Alexander Marson[5,6,8], Jacob E. Corn[4,5], Thomas Quertermous[9], William J. Greenleaf[1,2,8,11,12], Howard Y. Chang[1,7,12]

## AFFILIATIONS

[1] Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA 94305, USA
[2] Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA
[3] Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305, USA
[4] Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA
[5] Innovative Genomics Institute, University of California, Berkeley, CA, 94720, USA
[6] Department of Microbiology and Immunology, University of California, San Francisco, CA 94143, USA
[7] Program in Epithelial Biology, Stanford University School of Medicine, Stanford, CA, 94305, USA.
[8] Chan Zuckerberg Biohub, San Francisco, CA 94158, USA
[9] Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, California 94305, USA.
[10] Biomedical Sciences Graduate Program, University of California, San Francisco, CA 94143, USA
[11] Department of Applied Physics, Stanford University, Stanford, CA, 94305, USA
[12] These authors contributed equally
[13] These authors jointly directed this work

*To whom correspondence should be addressed. E-mail: howchang@stanford.edu and wjg@stanford.edu

**ABSTRACT**

The challenge of linking intergenic mutations to target genes has limited molecular understanding of diverse human diseases. Here, we show H3K27ac HiChIP generates high-resolution contact maps of active enhancers and target genes in rare primary human T cell subtypes and coronary artery smooth muscle cells. Differentiation of naïve T cells to either T helper 17 cells or regulatory T cells create subtype-specific enhancer-promoter interactions, specifically at regions of shared DNA accessibility. These data provide a principled means of assigning molecular functions to autoimmune and cardiovascular disease risk variants, linking hundreds of noncoding variants to putative gene targets. Target genes identified with HiChIP are further supported by CRISPR interference and activation at linked enhancers, by the presence of expression quantitative trait loci, and by allele-specific enhancer loops in patient-derived primary cells. The majority of disease-associated enhancers contact genes beyond the nearest gene in the linear genome, leading to a four-fold increase of potential target genes for autoimmune and cardiovascular diseases.

Gene expression programs are intimately linked to the hierarchical organization of the genome. In mammalian cells, each chromosome is organized into hundreds of megabase-sized topologically associated domains (TADs), which are conserved from early stem cells to differentiated cell types[1]. Within this invariant TAD scaffold, cell type-specific enhancer-promoter (E-P) interactions establish regulatory gene expression programs[2]. Standard methods require tens of millions of cells to obtain high-resolution interaction maps and confidently assign E-P contacts[3–5]. Thus, the principles that govern E-P conformation in disease-relevant patient samples are incompletely understood. This gap in understanding is particularly problematic for interpreting the molecular functions of inherited risk factors for common human diseases, which reside in intergenic enhancers or other non-coding DNA features in up to 90% of cases[6–9]. Such disease-relevant enhancers may not influence the expression of the nearest gene (often reported as the default target in the literature), and instead act in a cell-type specific manner on distant target genes residing up to hundreds of kilobases (kb) away[2,10–14]. Recently, systematic perturbations of regulatory elements in select gene loci have shown that effects of individual regulatory elements on gene activity can be predicted from the combination of (i) enhancer activity [marked by histone H3 lysine 27 acetylation (H3K27ac) level] and (ii) enhancer-target looping[5,15]. Here we leverage this insight to capture the combination of these two types of information genome-wide in a single assay, mapping the enhancer connectome in disease-relevant primary human cells.

## RESULTS

### H3K27ac HiChIP identifies functional enhancer interactions in low cell number samples

We recently developed HiChIP, a method for sensitive and efficient analysis of protein-centric chromosome conformation[16]. Cohesin HiChIP in GM12878 cells identified similar numbers of loops as *in situ* Hi-C (~10,000) with high correlation (R = 0.83), demonstrating that HiChIP captures loops with high sensitivity and specificity. Here, we evaluated the enhancer and promoter-associated mark H3K27ac[17–19] as a candidate factor to selectively interrogate E-P interactions genome-wide. We performed H3K27ac HiChIP in mouse embryonic stem (mES) cells to compare to cohesin HiChIP in the same cell type (**Supplementary Fig. 1a, Supplementary Table 1**)[16]. 3,552 of 4,191 H3K27ac HiChIP loops in mES cells were also identified by cohesin HiChIP. The H3K27ac-biased loops ($\log_2$ fold-change > 1 and FDR < 10%) spanned shorter distances than cohesin-biased loops, and were enriched for H3K27ac ChIP-seq peaks (78.9%; **Supplementary Fig. 1b-f, Supplementary Table 2**). Moreover, systematic titration of input material showed H3K27ac HiChIP retained high signal fidelity and reproducibility from 25 million to 50,000 cells as input material (loop signal correlation = 0.918 and loop overlap of 43%, comparable to in situ Hi-C replicate loop overlap of 53%; **Supplementary Figs. 2 and 3**). Therefore, H3K27ac HiChIP identifies high-

confidence chromatin loops focused around enhancer interactions from limited cell numbers.

In order to capture (i) conformational change during T cell differentiation and (ii) cell type-specific chromatin contacts of autoimmune risk variants in protective and pathogenic T cell types, we performed H3K27ac HiChIP on primary human Naïve T cells ($CD4^+CD45RA^+CD25^-CD127^{hi}$), regulatory T cells ($T_{reg}$; $CD4^+CD25^+CD127^{low}$) and T helper 17 cells ($T_H17$; $CD4^+CD45RA^-CD25^-CD127^{hi}CCR6^+CXCR5^-$) directly isolated from donors (**Fig. 1a,b and Supplementary Fig. 4a**)[20,21]. $T_H17$ cells were sorted to include autoimmune disease-relevant pathogenic $T_H17$ cells and to exclude follicular helper T cells with a distinct surface phenotype and immune function (**Supplementary Fig. 4a**)[22–24]. Peripheral blood $CD4^+$ T cells were isolated from three healthy subjects, isolated by FACS, and subjected to H3K27ac HiChIP. HiChIP libraries from each subset were high quality; greater than 40% of the total reads sequenced represented unique paired-end tags (PETs) (**Supplementary Fig. 4b-d and Supplementary Table 1**). Furthermore, libraries exhibited high 1D signal enrichment at enhancers and promoters, and globally recapitulated publically available H3K27ac ChIP-seq datasets (74.7% overlap of ChIP-seq and 1D HiChIP peaks; **Fig. 1c**)[25]. Inspection of the interaction matrix at progressively higher resolution revealed chromatin compartments, TADs, and focal loops, as previously reported in high-resolution Hi-C and HiChIP analyses from cell lines (**Fig. 1b**)[4,16]. Importantly, H3K27ac HiChIP maps were capable of identifying focal interactions at 1 kb resolution, which is comparable to *in situ* Hi-C maps generated from 100-fold more cells and sequenced to 13-fold greater depth[4] (**Fig. 1b**).

Previous saturation perturbation screens demonstrated that functional enhancers can be identified by integrating H3K27ac ChIP-seq signal with chromosome conformation contact strength (Hi-C)[5]. Since H3K27ac HiChIP combines these two components into one assay, we reasoned that HiChIP signal, which we term Enhancer Interaction Signal (EIS), should identify functional regulatory elements. To validate this prediction, we first generated H3K27ac HiChIP maps in a chronic myelogenous leukemia cell line (K562) as a direct comparison to published high-resolution CRISPR interference (CRISPRi) screens[5]. We then examined the 3D enhancer landscape of the *MYC* and *GATA1* loci using virtual 4C (v4C) analysis, where a specific genomic position is set as an anchor viewpoint, and all interactions occurring with that anchor are visualized in 2D[16]. v4C analysis of the *MYC* promoter demonstrated that EIS in K562 cells captured all functional enhancers identified in the CRISPRi screen (**Fig. 2a**). Analysis of the *GATA1* locus demonstrated a similar agreement between both methods (**Fig. 2b**). Quantitatively, EIS in K562 cells was significantly correlated with CRISPRi score in the same cell type, whereas EIS in GM12878 (GM; B cell lymphoblast) cells was not correlated with K562 CRISPRi (Spearman's rho = 0.332 and 0.145; p-value = $9.25 \times 10^{-5}$ and 0.1246; **Fig. 2c**).

We found the enhancer landscapes of the *MYC* promoter to be highly cell-type specific. v4C analysis of the *MYC* promoter in GM and My-La ($CD4^+$ T cell leukemia) cells showed dramatically different regulatory interactions with the promoter compared to K562 cells (**Fig. 2d**). To validate EIS specificity, we performed CRISPRi experiments in GM cells using sgRNAs targeting enhancers identified in either GM or My-La HiChIP

maps as well as a positive control sgRNA targeting the *MYC* promoter and a negative control sgRNA targeting lambda phage sequence (**Fig. 2e**). As expected, we found that CRISPRi of GM, but not My-La, enhancers impacted *MYC* expression and cell growth in GM cells (**Fig. 2e**).

Finally, we focused on the *CD69* locus, where a high-resolution CRISPRa screen identified three enhancers located approximately 4 kb, 13 kb, and 37 kb upstream of the transcription start site[26]. These sites were also precisely identified by Naïve T cell H3K27ac HiChIP. Moreover, HiChIP identified four additional distal enhancers that were outside the region spanned by the sgRNA tiling array (**Fig. 2f and Supplementary Fig. 5**). To functionally validate these novel enhancers, we performed CRISPRa experiments in Jurkat cells with sgRNAs targeting these enhancers, the *CD69* promoter, the *KLRF2* promoter as a locus negative control, and a non-human genome-targeting negative control. We observed a significant increase in *CD69* RNA and protein levels in the four HiChIP enhancers compared to negative controls (**Fig. 2g and Supplementary Fig. 5**). Interestingly, two of the four identified novel enhancers were within promoter regions of distant genes. These findings are in line with previous reports that identified widespread distal gene regulatory functions of promoters genome-wide[27,28]. Altogether, these results suggest that H3K27ac HiChIP EIS identifies functional regulatory elements, and that enhancers that regulate a gene of interest can differ significantly between cell-types.

**Landscape of enhancer interactions in primary human T cell subsets**

We examined global features of the enhancer connectome associated with cellular differentiation from Naïve T cells to either $T_H17$ cells or $T_{reg}$ cells. We identified a total of 10,706 high confidence loops in the union set of the three cell types (**Supplementary Table 2**). Analysis of loop read support between biological replicates demonstrated high reproducibility (**Supplementary Fig. 4c**), and ~91% of loop anchors were associated with either a promoter or enhancer[29], as expected, with a median distance of 130 kb (**Supplementary Fig. 6a,b**). Importantly, high-resolution E-P connectivity maps revealed several features that could not be discerned from 1D epigenomic data (i.e. H3K27ac ChIP-seq or ATAC-seq; **Fig. 3a**). These features included: (i) 'enhancer skipping': enhancers that lack strong EIS with their closest neighboring gene, or many neighboring genes, yet have strong EIS with a more distal target promoter, (ii) higher order structures such as 'enhancer cliques' (related to loop cliques[30]): multiple regulatory elements that have strong EIS with a single target promoter, (iii) promoter to promoter interactions[13,31], and (iv) 'enhancer switching': enhancers that exhibit differential EIS with a target promoter in a cell type-specific manner (**Fig. 3a**).

We found that EIS contacts were very cell type-specific, and gene regulation in distinct T cell populations was better understood using EIS signal, compared to 1D enhancer activity. After quantile-quantile normalization of matrix-balanced contact reads at high-confidence loops (correcting for false positives caused by 1D fragment visibility; **Materials and Methods**), we focused on the top and bottom 5% of EIS ranked by cell-type bias for each pair-wise comparison (**Supplementary Figs. 6c-g and 7,**

**Supplementary Tables 3-4**). We identified 6 clusters of EIS defined by interaction strength patterns across T cell subtypes: Clusters 1-3 were enriched in Naïve cells, Cluster 4 was enriched in $T_H17$ and $T_{reg}$ cells, and Clusters 5 and 6 were enriched in $T_H17$ and $T_{reg}$ cells, respectively (**Fig. 3b**). Cell type-specific enhancer loop anchors revealed genes encoding canonical T cell subtype TFs and effector molecules (**Fig. 3b, Supplementary Figs. 8 and 9**). Deeper v4C analysis of shared and cell type-specific loci pinpointed regulatory elements interacting with each gene promoter of interest as well as local conformational landscape changes (**Supplementary Figs. 8 and 9**). TF motifs located within cell type-specific loop anchors were enriched for TFs known to drive T cell subtype differentiation and nominated novel TFs involved in regulation (**Fig. 3c**). Furthermore, cell type EIS bias was associated with differential expression of genes located within corresponding EIS anchors for the same cell type (Spearman's rho = 0.242 and 0.207; p-value = $4 \times 10^{-15}$ and $2 \times 10^{-11}$; **Fig. 3d**).

Cell type-specific EIS may be driven by cell type-specific enhancer activation (based on H3K27ac ChIP-seq) or stable enhancer activation with cell type-specific looping (Hi-C) in a gene specific manner. We first examined H3K27ac ChIP-seq at differential EIS anchors and found that many biased H3K27ac HiChIP interactions also exhibited biased ChIP-seq signal, as expected. 58.5% of Naïve-biased loops contain at least one Naïve-biased ChIP-seq peak ($\log_2$ fold change > 1) located on the anchors. Similarly, 66.7% of $T_H17$-biased and 67.8% of $T_{reg}$-biased interaction anchors were cell type-specific in 1D (**Supplementary Fig. 10a**). Therefore, while on average ~64% of the differential EIS corresponded to change in 1D data, ~36% were likely also driven by change in 3D chromatin loop strength. To further assess the contribution of cell type-specific 3D signal to EIS, we examined HiChIP 1D signal at differential EIS anchors. We found that HiChIP 1D signal correlated better with ChIP-seq signal than EIS, with a higher likelihood of differential ChIP-seq signal overlapping differential HiChIP 1D signal compared to 3D, suggesting EIS bias is in part driven by 3D changes (**Supplementary Fig. 10b**).

As a complementary analysis, we asked whether the integration of reference cell line Hi-C data with primary T cell H3K27ac ChIP-seq could recapitulate HiChIP EIS in primary T cells. We first binned GM Hi-C loops with increasing primary T cell ChIP-seq signal at loop anchors and then determined the overlap of loops in each bin with loops derived from H3K27ac HiChIP. As expected, increased ChIP-seq signal at the Hi-C anchors led to increased overlap with the HiChIP loops. However, the overlap was lower in all T cell subtypes compared to the same analysis performed using GM HiChIP data. These observations demonstrate that cell-type specific 3D interactions can impact EIS independent of differences in 1D ChIP-seq signal (**Supplementary Fig. 10c**). Similarly, previously generated enhancer-promoter maps obtained from bulk unsorted T cells did not identify T cell subtype-specific interactions obtained using H3K27ac HiChIP. To assess the unique information obtained through cell type-specific interaction maps, we compared promoter Capture Hi-C maps in bulk $CD4^+$ T cells to H3K27ac HiChIP maps in Naïve, $T_H17$, and $T_{reg}$ cells[14]. We examined the overlap of promoter Capture Hi-C interactions with HiChIP interactions at increasing thresholds of cell type-specificity. Strikingly, the most cell type-specific loops in $T_H17$ and $T_{reg}$ (16-fold enriched)

demonstrated a low discovery rate in promoter Capture Hi-C T cells (11.83% in 415 loops and 13.83% in 373 loops, respectively; **Supplementary Fig. 10d**). Many of these subset-specific interactions included genomic loci encoding functionally important effector genes, such as *LRRC32*. The *LRRC32* locus contains $T_{reg}$-specific loops that are neither visualized in HiChIP maps from Naïve or $T_H17$ cells nor in bulk $CD4^+$ promoter Capture Hi-C maps (**Supplementary Fig. 10e**). Since primary human $T_H17$ and $T_{reg}$ cells are present in human blood with low frequency, it would also be challenging to generate subset-specific promoter Capture Hi-C maps with published promoter Capture Hi-C protocols. In summary, EIS is derived from a combination of 1D ChIP-seq and 3D interaction signal and cannot be accurately predicted from 3D maps in reference cell lines or unsorted primary cell datasets.

Cell type-specific EIS can occur at sites of shared chromatin accessibility. Paired chromatin accessibility profiles by Assay of Transposase-Accessible Chromatin by sequencing (ATAC-seq)[32] from each T cell subset revealed most cell type-specific loop anchors had equivalent chromatin accessibility across all three cell types (**Fig. 3e-g**). To illustrate this finding, we examined the *BACH2* promoter, which exhibits shared chromatin accessibility at enhancers, but increased EIS in Naïve cells (**Fig. 3e**). Globally, only 14.2% of Naïve-biased loops contain at least one Naïve-biased ATAC-seq peak ($\log_2$ fold change > 1) located on the anchors. In addition, 27.8% of $T_H17$-biased and 16.5% of $T_{reg}$-biased interaction anchors were cell type-specific for 1D accessibility. Furthermore, the majority of cell type-specific TF motifs were observed in shared ATAC-seq peaks within differential interactions, highlighting that these regions are functioning in T cell differentiation (**Fig. 3f-g**). Altogether, these results suggest that in highly related – yet functionally distinct – cell types, a portion of transcriptional control is achieved through differential chromosome looping, rather than differential chromatin accessibility. This finding is consistent with previous studies which demonstrated that T cell subset-specific TFs, such as Foxp3, act predominantly at pre-accessible chromatin sites to establish subset-specific gene expression[33].

## T cell subset-specific enhancer interactions link autoimmunity-associated DNA elements to target genes

The high specificity of EIS enabled us to identify putative target genes of autoimmune disease risk loci in functionally relevant T cell subsets. To achieve this, we used a previously described list of putatively causal variants associated with 21 autoimmune diseases, known as PICS SNPs, which were fine-mapped based on dense genotyping data[25]. First, we determined that PICS autoimmune SNPs were significantly enriched in T cell loop anchors, with specific autoimmune diseases showing greater than 5-fold enrichment compared to a shuffled control loop set (**Supplementary Fig. 11**). Next, we constructed a set of all possible connections between autoimmune risk SNPs and TSS within 1 Mb and measured the EIS for each SNP-TSS pair (**Fig. 4a**). We aggregated these signals to determine the overall interaction activity in each T cell subtype in each disease (**Fig. 4b**). Strikingly, we observed high interaction strength enrichments and cell type specificity in autoimmune disease SNPs, but low enrichment

308 and cell specificity in non-immune traits (**Fig. 4b**). To further visualize HiChIP bias in
309 shared or differential enhancers, we analyzed SNP-TSS interactions grouped by their
310 presence near H3K27ac ChIP-seq peaks (**Supplementary Fig. 12a,b**). Interestingly,
311 we observed a large number of active SNP-TSS pairs that were present in regulatory
312 regions that were shared between T effector cell types ($T_{reg}$ and $T_H17$), while relatively
313 less EIS signal was observed in SNPs located in cell-type specific enhancers,
314 supporting the concept that many autoimmune disease variants impact common T cell
315 effector/activation pathways[25,34]. Notably, SNPs present in enhancers shared across all
316 three cell types could still be distinguished by HiChIP bias (**Supplementary Fig. 12a,b**).
317 For example, although we could not detect cell type bias at risk loci for Alopecia Areata
318 using H3K27ac ChIP-seq (**Supplementary Fig. 12a,b** and ref. 3), H3K27ac HiChIP
319 identified increased SNP-TSS activity in $T_{reg}$ cells among shared T cell enhancers,
320 consistent with several studies identifying the crucial role of this cell type in disease
321 pathogenesis[35]. Importantly, autoimmune signal enrichments were not readily apparent
322 from 1D H3K27ac ChIP-seq peaks, aggregated ChIP-seq signal within the TAD
323 containing the SNP, nor cell line H3K27ac HiChIP datasets (**Fig. 4b and**
324 **Supplementary Fig. 12c**). Therefore, examining 3D disease variant interactions may
325 capture cell type biases more robustly than 1D epigenomic data. Finally, to validate our
326 findings with an orthogonal dataset, we performed SNP-TSS EIS analysis on an
327 overlapping set of autoimmune disease-associated SNPs obtained from the NHLBI
328 GRASP catalog and observed similar enrichments of specific T cell subsets
329 (**Supplementary Fig. 12d**).
330 We leveraged HiChIP to identify potential gene targets of intergenic SNPs, which
331 have classically been paired to the nearest neighboring gene. We overlapped the SNP-
332 TSS pairs with loops to call a discrete set of target pairs. We then performed differential
333 analysis on the SNP-TSS loops to ascertain bias for specific T cell subsets (**Fig. 4c**).
334 Overall, 3.7% and 9.4% of SNP-TSS pairs were cell type-biased in a comparison of
335 Naïve and Memory T cells, and 2.2% and 6% were cell type-biased in a comparison of
336 $T_{reg}$ cells and $T_H17$ cells, respectively. Examples of biased SNP-TSS pairs included
337 *FOXO1* in Naïve T cells (rs9603754), *BATF* (rs2300604) in Memory T cells, *CTLA4*
338 (rs10186048) in $T_{reg}$ cells, and *IL2* (rs7664452) in $T_H17$ cells (**Fig. 4c**). Next, we sought
339 to characterize the connectivity landscape of the SNP-TSS loops. We identified an
340 average of 1.75 gene targets per autoimmune SNP (ranging from 0 to over 10 target
341 genes), while non-immune traits did not demonstrate an increase in targets (0.33 genes
342 per SNP; **Supplementary Fig. 12e**). We examined the SNP-nearest gene associations
343 and compared these to the HiChIP gene target list. For 684 autoimmune intergenic
344 SNPs, we identified a total of 2,597 HiChIP target genes, representing a four-fold
345 increase in target genes for known disease SNPs (**Fig. 4d**). Only 367 (~14%) of all
346 targets were the nearest gene to the SNP, while approximately ~86% of SNPs skipped
347 at least one gene to reach a predicted target TSS (**Supplementary Fig. 12e**).
348 Furthermore, approximately ~45% of SNP to HiChIP target interactions had increased
349 signal compared to the same SNP to nearest gene, despite distance biases.
350

**Disease-associated enhancer-target gene interactions validated by eQTL and CRISPRi**

HiChIP enhancer-target gene interactions can be validated using previously identified point mutations that alter expression at distantly located genes in T cells—i.e. expression quantitative trait loci (eQTL)[36]. For example, the celiac disease-associated SNP rs2058660 impacts the expression of the inflammatory cytokine receptor genes *IL18RAP*, *IL18R1*, *IL1RL1*, and *IL1RL2*, which are known regulators of intestinal T cell differentiation and response[37]. HiChIP EIS revealed contacts between rs2058660 and each of these predicted gene promoters (**Supplementary Fig. 13a**). Similarly, the Crohn's disease risk variant rs6890268 and the multiple sclerosis (MS) risk variant rs12946510 impact the expression of *PTGER4* and *IKZF3*, respectively, and H3K27ac HiChIP also demonstrated clear contacts between these SNPs and their predicted promoter (**Supplementary Fig. 13a**). Globally, HiChIP contact signal was increased in eQTLs in T cells compared to a distance-matched background loop set (p-value < 2.2 x $10^{-16}$; **Fig. 4e)** or to eQTLs identified in an unrelated cell type (liver; p-value < 2.2 x $10^{-16}$). The overlap of HiChIP and eQTL loci provides support for chromosome interactions as a physical basis for distal eQTLs[10–12] and further validates the HiChIP approach to assign enhancer-target gene relationships.

We next sought to directly validate HiChIP SNP-gene targets using CRISPRi in My-La cells. First, we focused on three loci of interest in primary T cells and then confirmed that the SNP-TSS loops were also present in My-La cells (**Fig. 4f and Supplementary Fig. 13b**). We then targeted sgRNAs to these SNP-containing enhancers, as well as positive control sgRNAs to the HiChIP target gene promoters and a negative control sgRNA targeting lambda phage sequence. As expected, we observed a significant reduction of RNA levels in the HiChIP target genes upon CRISPRi of its SNP-containing enhancer (**Fig. 4f**).

**Fine-mapping disease-associated DNA variants with high-resolution H3K27ac HiChIP**

Since SNP-TSS HiChIP signal is capable of identifying target genes of candidate SNPs, we asked whether TSS-SNP HiChIP signals could also be used to nominate functional causal variants within haplotype blocks in a reciprocal manner. We first performed a proof-of-principle analysis using fine-mapped SNPs associated with inflammatory bowel disease (IBD)[38] or Type 1 Diabetes (T1D)[39] as well as high confidence PICS SNPs and examined EIS from putatively causal SNPs to all gene promoters within 300 kb. EIS from putatively causal SNPs to gene promoters was significantly higher than EIS from a distance-matched set of SNPs within the same LD block to gene promoters (p-value = 2.4 x $10^{-15}$, 8.7 x $10^{-8}$, 3.9 x $10^{-3}$ for IBD fine-mapped SNPs, T1D fine-mapped SNPs, and high confidence PICS, respectively; **Fig. 5a** and **Supplementary Fig. 14a**). Next, we assessed the fine-mapping ability of HiChIP EIS at individual loci of interest. First, we focused on IBD- and MS-associated SNPs neighboring the *PTGER4* and *SATB1* loci and performed v4C analysis anchored at the

395 gene promoters. We calculated EIS signal at 1 kb resolution and identified specific
396 regions within the linkage disequilibrium (LD) blocks that contained the highest EIS to
397 the target promoters, positioning the likely causal SNPs within these regions (**Fig. 5b**
398 and **Supplementary Fig. 14b**). For example, at the *PTGER4* locus (**Fig. 5b**), the ~160
399 kb genomic interval spanned by LD SNPs in association with Crohn's disease is refined
400 to two bins of 3kb and 4kb, which both contain PICS SNPs.
401 We asked whether complex disease-associated loci containing more than one
402 gene could be fine-mapped using HiChIP. We focused on two disease-associated
403 enhancers in between the *STAT1* and *STAT4* gene promoters (**Fig. 5c**). These two
404 genes encode transcription factors with distinct and canonical roles in immune
405 regulation. Signal transducer and activator of transcription 1 (STAT1) is critical for type I
406 IFN and IFNγ signaling, whereas STAT4 is activated downstream of IL-12 signaling to
407 induce $T_H1$ differentiation and IFNγ expression[40]. We first investigated bias of these
408 enhancers to *STAT1* and *STAT4* and found that, despite comparable linear distance
409 and 1D signal at the promoters, the enhancers were significantly biased to interact with
410 *STAT4*. Next, we fine-mapped the disease associated SNPs within this locus using 1 kb
411 resolution EIS from the *STAT4* promoter, and narrowed down candidate functional
412 variants within the two enhancers (**Fig. 5c**). In summary, HiChIP EIS can nominate
413 functional causal variants within haplotype blocks, and two-way analysis of target gene
414 identification from an enhancer of interest and high-resolution interaction maps of that
415 enhancer with its target gene can be used to fine-map disease-associated loci
416 containing several candidate genes.
417
418 **Allelic target gene bias of cardiovascular disease risk variants in patient smooth**
419 **muscle cells**
420
421 Finally, we asked whether this approach could be applied broadly to other
422 categories of human disease, and whether we could directly test SNP-TSS associations
423 using allele-specific HiChIP. We generated high-resolution E-P maps from primary
424 human coronary artery smooth muscle cells (HCASMC), which can be used to inform
425 variants linked to cardiovascular diseases[41]. First, to validate cell type specificity, we
426 examined the *TCF21* gene promoter, a transcription factor required for the
427 differentiation of epicardial progenitor cells to HCASMC[42] and observed enrichment in
428 HCASMC EIS relative to Naïve T cells (**Fig. 6a**). We next examined the 9p21.3 locus,
429 which harbors risk associations with several cardiovascular disorders, including
430 coronary artery disease (CAD), myocardial infarction, and stroke[43–45]. We found that the
431 promoters of all three genes in the locus interact with one another and with CAD variant-
432 containing enhancers located approximately 100 kb upstream of the *CDKN2B* promoter
433 (**Supplementary Fig. 15**). We then generated SNP-TSS target lists using CAD SNPs
434 identified in the CARDIoGRAMplusC4D study[46]. We again performed differential
435 analysis on the SNP-TSS loops to ascertain bias for HCASMC versus Naïve T cells
436 (**Fig. 6b**). Overall, 75.1% of biased HCASMC SNP-TSS pairs were CAD SNPs, while
437 only 5.5% of Naïve T cell biased SNP-TSS pairs were CAD SNP-TSS loops,
438 demonstrating cell-type specificity of SNP-TSS interactions. Next, we examined the

connectivity of the HCASMC SNP-TSS contacts and identified 1,062 gene targets, of which only 120 (~11%) mapped to the nearest gene. Furthermore, approximately 89% skipped at least one gene to reach a predicted target TSS, and 64% of SNPs were mapped to more than a single gene target.

We took advantage of genome phasing information in HCASMC to measure E-P interactions at allele-specific CAD SNPs, allowing us to examine the functional consequence of a risk variant compared to its alternative allele in the same nucleus. First, 4.2% of high confidence loops in HCASMC with no observed mapping bias in the anchors exhibited significant allelic bias (FDR < 0.05, **Fig. 6c**), consistent with frequency of allelic imbalance of RNA expression and prior evidence of allele-specific regulation of specific E-P interactions[49,50]. We leveraged this global E-P allelic bias to examine the effect of a risk variant compared to its control alternative allele for a set of CAD-associated SNP-target gene pairs (**Fig. 6d**)[49]. We found that many risk alleles disrupt enhancer-target gene interactions, but a subset of pathogenic SNPs increased enhancer-target gene interaction. At CAD risk variant rs1537373 in the 9p21.3 locus, the risk allele (T) showed increased EIS to the *CDKN2A* promoter as well as an additional enhancer within the lncRNA *ANRIL* relative to the reference allele (G) (**Fig. 6e**). We further observed increased EIS of the CAD risk variant rs4562997 to an additional *SMAD3* enhancer 10 kb downstream of the TSS (**Fig. 6e**). The ability to resolve enhancer connectomes of the risk and reference alleles in the same nucleus demonstrates that the mutated base in the risk allele suffices to alter enhancer looping in *cis* in disease-relevant primary cells.

## DISCUSSION

Here, we developed an approach to define the high-resolution landscape of E-P regulation in primary human cells. We find that E-P contacts are highly dynamic in related cell types and often involve genomic elements with shared accessibility. Accordingly, many complex features of the 3D enhancer connectome cannot simply be predicted from 1D, which demonstrates that mapping conformation in primary cells can identify novel regulatory connections underlying gene function in human disease. We take advantage of this principle to chart the connectivity of autoimmune and cardiovascular GWAS SNPs and link SNPs to hundreds of potential target genes. Although non-genic SNPs have previously been paired with their closest neighboring gene, we find that the majority of these variants can engage in long-distance interactions, including skipping several promoters to predicted target genes, connecting to multiple genes, or acting in concert with enhancer cliques to contact a single gene. Further use of this approach will help to clarify hidden mechanisms of human disease that are driven by genetic perturbations in non protein-coding DNA elements, which can now be linked to their cognate gene targets in primary cells.

## AUTHOR CONTRIBUTIONS

M.R.M, A.T.S., W.J.G., and H.Y.C. conceived the project. M.R.M., A.T.S., J.T., and R.L. performed all genomics assays with help from T.N., M.R.C., N.S., and R.A.F. A.T.S. performed all sorting for experiments. B.G.G., S.W.C., M.R.M., M.L.N., K.R.K., and D.R.S., performed all CRISPR validation experiments. E.A.B., C.D., M.R.M., and J.X. analyzed HiChIP data. J.G., A.T.S., and Y.W. analyzed ATAC-seq data. A.J.R. and P.G.G. analyzed GWAS SNPs in HiChIP data. A.K., P.A.K., A.M., J.E.C., T.Q., W.J.G., and H.Y.C guided experiments and data analysis. M.R.M, A.T.S, E.A.B, C.D., W.J.G, and H.Y.C. wrote the manuscript with input from all authors.

## ACKNOWLEDGEMENTS

We thank members of the Chang and Greenleaf laboratories for helpful discussions and Justin Tumey for artwork. This work was supported by National Institutes of Health (NIH) P50HG007735 (H.Y.C. and W.J.G.) and U19AI057266 (W.J.G), Human Frontier Science Program (W.J.G.), Rita Allen Foundation (W.J.G.), and the Scleroderma Research Foundation (H.Y.C). M.R.M. and E.A.B. acknowledge support from the National Science Foundation Graduate Research Fellowship. A.T.S. is a Cancer Research Institute Irvington Fellow supported by the Cancer Research Institute. B.G.G. was supported by an IGI-AstraZeneca Postdoctoral Fellowship. M.R.C. is supported by a grant from The Leukemia & Lymphoma Society Career Development Program. J.E.C. was supported by the Li Ka Shing Foundation and the Heritage Medical Research Institute. W.J.G. and A.M. are Chan Zuckerberg Biohub investigators. A.M. serves as an advisor to Juno Therapeutics and PACT Therapeutics and the Marson lab has received sponsored research support from Juno Therapeutics and Epinomics. A.M. and J.E.C. are founders of Spotlight Therapeutics. Sequencing was performed by the Stanford Functional Genomics Facility (NIH S10OD018220). We thank Agilent Technologies for generating oligo pools for cloning of the CRISPRa gRNAs. We thank the UC Berkeley High Throughput Screening Facility and Flow Cytometry Facility. H.Y.C. and W.J.G. are founders of Epinomics and members of its scientific advisory board.

511 **FIGURE LEGENDS**
512
513 **Fig. 1 HiChIP identifies high-resolution chromosome conformation in primary**
514 **human T cells. (A)** Primary T cell H3K27ac HiChIP experimental outline. **(B)** (Left)
515 FACS strategy for Naïve, $T_H17$, and $T_{reg}$ cells from total peripheral blood $CD4^+$ T cells.
516 Number represents percent of cells within that gate. (Right) Knight-Ruiz (KR) matrix-
517 balanced interaction maps for Naïve, $T_H17$, and $T_{reg}$ cells at 500 kb, 25 kb, and 5 kb
518 resolution, and raw interaction maps at 1 kb resolution, centered on the *KLF2*, *RBPJ*,
519 and *LRRC32* loci. **(C)** HiChIP 1D and 3D signal enrichment at the *RORC* locus in $T_H17$
520 over Naïve T cells.
521
522 **Fig. 2 Validation of H3K27ac HiChIP regulatory element identification with**
523 **CRISPR interference and activation. (A)** Interaction profile of the *MYC* promoter in
524 K562 H3K27ac HiChIP at 10 kb resolution. K562 H3K27ac ChIP-Seq is from ENCODE.
525 CRISPRi-validated regulatory regions in K562 cells are indicated[5]. **(B)** Interaction profile
526 of the *GATA1* promoter in K562 H3K27ac HiChIP at 1 kb resolution. CRISPRi-validated
527 regulatory regions in K562 cells are indicated[5]. **(C)** Correlation of *MYC* K562 H3K27ac
528 HiChIP signal with max CRISPRi score within the HiChIP 10 kb window. (**D**) Interaction
529 profiles of the *MYC* promoter in GM12878 and My-La H3K27ac HiChIP at 10 kb
530 resolution. T cell H3K27ac ChIP-seq and ATAC-seq are from Naïve T cells. **(E)** (Top)
531 CRISPRi validation in GM12878 cells of GM12878 and My-La-biased *MYC* enhancers.
532 (Bottom) *MYC* RNA levels by qRT-PCR and cell growth rates in CRISPRi GM12878
533 cells targeted to cell type-biased enhancers, the *MYC* promoter, and a non-targeting
534 negative control (n = 3). * corresponds to p-value < 0.05, ** to p-value < 0.01, *** to p-
535 value < 0.001, **** to p-value < 0.0001, and n.s. to not significant. **(F)** Interaction profile
536 of the *CD69* promoter in Jurkat H3K27ac HiChIP at 5 kb and 1 kb resolutions. The 1 kb
537 profile is focused on the window of the CRISPRa tiling screen. CRISPRa-validated
538 regulatory regions in Jurkat cells are indicated[26]. **(G)** (Top) CRISPRa validation in Jurkat
539 cells of *CD69* distal enhancers. (Bottom) *CD69* RNA and protein levels in CRISPRa
540 Jurkat cells targeted to distal enhancers, the *CD69* promoter, the *KLRF2* promoter as a
541 locus negative control, and a non-targeting negative control (n = 2).
542
543 **Fig. 3 Dynamic 3D enhancer landscapes in T cell differentiation. (A)**
544 Conformational features observed by H3K27ac HiChIP. **(B)** HiChIP EIS in 913
545 differential interactions identified in T cell subtypes. Interactions are clustered by cell-
546 type specificity. **(C)** Cell-type specific motif identification from ATAC-seq peaks in biased
547 EIS anchors. **(D)** EIS bias quartiles for Naïve to $T_H17$ and Naïve to $T_{reg}$ differentiation,
548 with corresponding differential RNA gene expression rankings. **(E)** Proportion of ATAC-
549 seq peaks within HiChIP differential interaction anchors that are cell-type specific ($log_2$
550 fold change > 1) or shared across all three subtypes. **(F)** Interaction profile of the
551 *BACH2* promoter at 5 kb resolution, demonstrating shared accessibility signal at Naïve-
552 biased EIS. **(G)** Global correlation of EIS and ATAC-seq fold-change in different T cell
553 subset pairwise comparisons.
554

**Fig. 4 HiChIP identifies cell type-specificity in autoimmune diseases and pairs GWAS SNPs with target genes. (A)** Generation of a loop set between all autoimmune SNPs and gene promoters within a 1 Mb region. **(B)** H3K27ac ChIP and HiChIP signal bias in T cell subtypes for SNP-TSS pairs. For each bin, PICS SNPs are tagged by H3K27ac only in the concordant cell type for the bias tested. SNPs are grossly divided into associations with autoimmune diseases or control, non-immune traits. **(C)** EIS Bias of SNP-TSS loops (with nearest gene annotated) in $T_H17$ and $T_{reg}$ subsets versus Naïve, and $T_H17$ versus $T_{reg}$. **(D)** Number of HiChIP gene targets versus nearest gene predictions for all looping nongenic autoimmune SNPs as well as SNPs for specific diseases. **(E)** Global validation of HiChIP SNP gene targets. Synthetic SNP-TSS pairs were generated from each $CD4^+$ eQTL SNP to its associated gene and compared to both a distance-matched shuffled SNP-TSS pair and a liver eQTL SNP-TSS pair. **(F)** HiChIP target gene RNA levels by qRT-PCR in CRISPRi My-La cells targeted to SNP-containing enhancers of interest, as well as positive control sgRNAs to the HiChIP target promoters and a non-targeting negative control (n = 3). * corresponds to p-value < 0.05, ** to p-value < 0.01, *** to p-value < 0.001, **** to p-value < 0.0001, and n.s. to not significant.

**Fig. 5 Fine-mapping of GWAS variants in LD using H3K27ac HiChIP. (A)** Global validation of HiChIP signal at putatively causal SNPs versus corresponding SNPs in LD ($r^2 \geq 0.8$) for $T_H17$ cells. SNP-TSS pairs were generated from published fine-mapped datasets, compared to a distance-matched SNP-TSS pair set in the same LD block. **(B)** Interaction profile of the *PTGER4* promoter, and a 1 kb resolution visualization of the SNP-containing enhancer of interest. LD SNPs ($r^2 \geq 0.8$) correspond to GRASP SNPs (genome-wide significance p-value < $10^{-8}$). The highlighted SNP was identified in both the high confidence PICS and GRASP datasets. **(C)** Interaction profiles of the *STAT1* and *STAT4* promoters, with 1 kb resolution visualizations of the SNP-containing enhancers of interest. Highlighted are 1D signal contributions at the *STAT1* and *STAT4* promoters. Highlighted SNPs are PICS closest to focal EIS to *STAT4*.

**Fig. 6 HiChIP identifies allelic bias to target genes for cardiovascular disease risk variants. (A)** Interaction profile of the *TCF21* gene promoter for H3K27ac HiChIP of HCASMC and Naïve T cells. **(B)** EIS bias between HCASMC and Naïve T cells in a union set of CARDIoGRAMplusC4D CAD and PICS autoimmune SNP-TSS loops. **(C)** Q-Q plot of allelic EIS imbalance in high confidence loops. Allelic mapping biased loops were identified through simulation and removed prior to EIS analysis. **(D)** EIS bias between CAD risk variants and their alternative alleles to eQTL associated target genes. **(E)** Allele-specific HiChIP interaction profiles at the 9p21.3 and *SMAD3* loci at 10 kb resolution in order to examine the functional consequence of a risk variant compared to its alternative allele.

## MATERIALS AND METHODS

### Human Subjects

This study was approved by the Stanford University Administrative Panels on Human Subjects in Medical Research, and written informed consent was obtained from all participants.

### Availability of Sequencing Data

Raw and processed data available at NCBI Gene Expression Omnibus, accession number GSE101498.

T cell ATAC and HiChIP datasets can be visualized in the WashU Epigenome Browser with the following link:
http://epigenomegateway.wustl.edu/browser/?genome=hg19&session=YAIzYBfrl9&statusId=1698051079

### Cell Culture and Primary T cell Isolation

Mouse ES cells (v6.5, Novus Biologicals: NBP1-41162) were cultured in Knockout DMEM (Gibco) + 15% FBS and leukemia inhibitory factor (LIF, Millipore) to 80% confluence. GM12878 (Coriell), Jurkat, and My-La ($CD4^+$) cells (ATCC) were grown in RPMI 1640 (Gibco) with 15% FBS to a concentration of 500,000 to 1 million cells per mL. Normal donor human peripheral blood cells were obtained fresh from AllCells. $CD4^+$ T cells were enriched from peripheral blood using the RosetteSep Human $CD4^+$ T Cell Enrichment Cocktail (StemCell Technology). For $CD4^+$ T helper cell subtypes, Naïve T cells were sorted as $CD4^+CD25^-CD45RA^+$, $T_H17$ cells were sorted as $CD4^+CD25^-CD45RA^-CCR6^+CXCR5^-$, and $T_{reg}$ cells were sorted as $CD4^+CD25^+CD127^{lo}$. For HiChIP experiments, 500,000 - 1 million cells were sorted into RPMI + 10% FCS. For ATAC-seq experiments, 55,000 cells were sorted into RPMI + 10% FCS. Post-sort purities of > 95% were confirmed by flow cytometry for each sample.

Primary human coronary artery smooth muscle cell (HCASMC) line derived from a normal human donor heart was purchased from Cell Applications, Inc. (350-05A) and cultured in smooth muscle growth medium (Lonza, CC-3182) supplemented with hEGF, insulin, hFGF-b, and 5% FBS. Cells were grown according to Lonza's instructions.

### Cell Fixation

Detached cell lines or sorted $CD4^+$ T cells were pelleted and resuspended in fresh 1% formaldehyde (Thermo Fisher) at a volume of 1 mL formaldehyde for 1 million cells. Cells were incubated at room temperature for 10 min with rotation. Glycine was added at a final concentration of 125mM to quench the formaldehyde, and cells were incubated at room temperature for 5 min with rotation. Finally, cells were pelleted and washed with PBS, pelleted again, and stored at –80 °C or immediately taken into the HiChIP protocol.

641 **HiChIP**
642 The HiChIP protocol was performed as previously described[16] with the following
643 modifications. For primary T cells, we performed HiChIP on as many cells as we could
644 obtain from a blood donation - approximately 500,000 - 1 million cells per T cell subtype
645 per replicate. We performed two minutes of sonication, no Protein A bead preclearing,
646 used 4 µg of H3K27ac antibody (Abcam ab4729), and captured the chromatin-antibody
647 complex with 34 µL of Protein A beads (Thermo Fisher). Qubit quantification post ChIP
648 ranged from 5 – 25 ng depending on the cell type and amount of starting material. The
649 amount of Tn5 used and PCR cycles performed were based on the post ChIP Qubit
650 amounts, as previously described[16].
651
652 25m cell line libraries were generated as previously described[16]. For low cell number
653 mouse embryonic stem cell samples, we performed two minutes of sonication and no
654 Protein A bead preclearing. Either 4 µg or 2µg of H3K27ac antibody (Abcam ab4729)
655 was used for ChIP in 500k or 100k/50k cells, respectively, and the chromatin-antibody
656 complex was captured with 34 (500k cells) or 20 µL (100k/50k cells) of Protein A beads.
657 Post-ChIP Qubit quantification for the 25m cell samples was approximately 1.5 µg. For
658 lower cell numbers, quantification was 30, 10, and 5 ng for 500k, 100k, and 50k cells,
659 respectively. The amount of Tn5 used and PCR cycles performed were based on the
660 post ChIP Qubit amounts, as previously described.
661
662 HiChIP samples were size selected by PAGE purification (300-700 bp) for effective
663 paired-end tag mapping, and therefore were removed of all primer contamination which
664 would contribute to recently reported "index switching" on the Illumina HiSeq 4000
665 sequencer[50]. All libraries were sequenced on the Illumina HiSeq 4000 to an average
666 depth of 500-600M total reads.
667
668 **HiChIP Data Processing**
669 HiChIP paired-end reads were aligned to hg19 or mm9 genomes using the HiC-Pro
670 pipeline[51]. Default settings were used to remove duplicate reads, assign reads to MboI
671 restriction fragments, filter for valid interactions, and generate binned interaction
672 matrices. HiC-Pro filtered reads were then processed into a .hic file using the
673 hicpro2juicebox function. The Juicer pipeline HiCCUPS tool was used to identify high
674 confidence loops[4] using the same parameters as for the GM12878 *in situ* Hi-C map:
675 hiccups -m 500 -r 5000,10000 -f 0.1,0.1 -p 4,2 -i 7,5 -d 20000,20000 .hic_input
676 HiCCUPS_output. For T cell Juicer loops, performing the default Juicer calls resulted in
677 a high rate of false positives upon visual inspection of the interaction matrix. We
678 therefore called loops with the same HiCCUPS parameters in two biological replicates
679 for each T cell subtype and then filtered loops for those that were reproducibly called in
680 both replicates. In addition, we removed all loops greater than 1 Mb.
681
682 1D signal enrichment and peak calling were generated from the HiC-Pro filtered
683 contacts file. Intrachromosomal contacts were filtered and both anchors were extended
684 by 75 base pairs. The combined bed file containing both anchors was then used to

generate bigwigs for visualization in the WashU Epigenome Browser or call peaks using MACS2.

Allele-specific HiChIP data processing was achieved using HiC-Pro's allele-specific analysis features[51]. First, HCASMC phasing data[41] was used to mask the hg19 genome and make indexes. HiC-Pro settings were similar as to described above, with the exception that reads were aligned to the masked genome, and then assigned to a specific allele based on phasing data.

**Interaction Matrices and Virtual 4C Visualization**

HiChIP interaction maps were generated with Juicebox using Knight-Ruiz (KR) matrix balancing and visualized using Juicebox software at 500 kb, 25 kb, 10 kb, and 5 kb resolutions as indicated in each analysis[4]. For 1 kb profiles, raw matrix counts were visualized in Java TreeView.

Virtual 4C plots were generated from dumped matrices generated with Juicebox. The Juicebox tools dump command was used to extract the chromosome of interest from the .hic file. The interaction profile of a specific 5 kb or 10 kb bin containing the anchor was then plotted in R. Replicate reproducibility was visualized with the mean profile shown as a line and the shading surrounding the mean representing the standard deviation between replicates. For the HCASMC data we observed low read coverage for allele-specific v4Cs at loci of interest. This is due to the density of SNPs for this genotype, and a low number of reads containing a phased SNP. We therefore could not observe interaction profiles when visualizing separate replicates with the standard deviation. We therefore utilized pseudoreplicates for the HCASMC v4C visualizations[52].

High confidence Juicer loop calls were loaded into the WashU Epigenome Browser along with corresponding ATAC-seq profiles and publically available H3K27ac ChIP-seq data from the Roadmap Epigenome Project. Browser shots from WashU track sessions were then included in virtual 4C and interaction map anecdotes.

**Reproducibility Scatter Plots and Correlations**

Biological and technical reproducibility comparisons between HiChIP experiments were generated by counting reads supporting a set of Juicer loop calls. For biological replicate comparisons, the loop set called from the merged replicates was used. For the comparison in mES cells between low cell number and 25 million cells, the union loop set between the two maps was used. The Pearson correlation between replicates or experiments was calculated from depth-normalized reads using the cor() function in R. Scatter plots were plotted using log-transformed raw reads supporting each loop

**Differential Analysis of HiCCUPS Loop Calls**

Juicer loop calls from the three T cell subtypes were initially combined into a union set of T cell loops. Loop signal was then obtained for the biological replicates of each T cell subtype. Vanilla coverage square root (VCsqrt) normalized signal for the interaction

729 matrix of each biological replicate using the Juicebox tools dump command. Normalized
730 signal was then assigned to the union loop set in each replicate.
731
732 VCsqrt signal per sample was quantile-quantile normalized under the assumption that
733 overall signal was identically distributed across all samples. Following normalization,
734 samples for Naïve, $T_H17$ and $T_{reg}$ cells shared Pearson correlations of 0.938, 0.942 and
735 0.934, respectively. PCA was performed using the prcomp function in R, which
736 demonstrated that the first PC, which exhibited near-identical loadings across the six
737 samples, explained 93% of the variance across the six samples. This was taken to
738 represent the shared signal across cell types. PCs 2-4 explained 2.2%, 2.0% and 1.4%,
739 respectively.
740
741 To study cell type-specific looping, the residual signal per loop was taken after
742 projecting the loop onto the unit vector along the diagonal (equal signal per cell type).
743 Cell type-specific and differential looping analysis were performed with the top and
744 bottom 5% of the distributions of either residual signal or differences between cell type
745 residual signals. Hierarchical clustering was performed using the union of all differential
746 loops in these extremes and using 1 minus the Pearson correlation as the distance
747 metric. QQ plots were generated by permuting residuals from the same cell type or
748 individual and summing them together and using this distribution to calculate p values
749 for the observed sums.
750
751 In parallel, differential loops were called using edgeR for both the mES and T cell
752 datasets. Again, biological replicate loop signal was obtained across a union set of
753 Juicer loops. We then used edgeR to identify loops with significant changes in signal
754 among pair-wise comparisons (FDR $< 0.1$, $\log_2 FC > 1$). Importantly, inspection of
755 differential loops identified from the two methods revealed high concordance.
756
757 Gene density was calculated from Ensembl gene annotations. GC content was
758 calculated per 10 kb bin using the bedtools nuc function and aggregated as needed.
759 Notably, Spearman correlation between gene density of an entire chromosome and the
760 number of differential loop anchors (rho = 0.914) was much higher than the correlations
761 between the variance in cell type signal per loop anchor and number of genes per 10 kb
762 window (rho = 0.322) and between differential loop anchors and gene density per 100
763 kb section (rho = 0.083). Correlations between GC content and number of differential
764 loops were similar at both the chromosome (rho = 0.729) and 100 kb bin (rho = 0.148)
765 levels, but while local GC content is likely to confound relative abundance, it is unclear
766 how chromosome-wide GC content could have the same effect.
767
768 For mES analysis of H3K27ac- and cohesin-mediated HiChIP, we performed edgeR to
769 obtain the biased loops for each factor. To determine functional bias of the top loops,
770 overlap was determined between edgeR differential loop anchors and relevant ChIP-seq
771 peaks. Smc1a ChIP-seq peaks were obtained from a published dataset[53]. CTCF, RNA

Polymerase II, and H3K27ac ChIP-seq peaks were obtained from the mouse ENCODE repository[54].

**Promoter/Enhancer Annotations of HiChIP Loops**

Promoters were defined as 1 kb regions centered at the TSS, and enhancers were identified as chromHMM enhancers not overlapping with promoters in any cell type. We annotated loop anchors as 'other' if the anchor did not contain a promoter or enhancer as defined above.

**RNA Expression Analysis**

Previously generated RNA-seq data[55] from Naïve, $T_H17$ and $T_{reg}$ cells was downloaded as fastqs from ArrayExpress. Illumina adaptors were trimmed using CutAdapt and Ensembl's cDNA transcripts were quantified using kallisto. Sleuth was used to identify transcripts that were differentially expressed across cell types with FDR controlled at 5%. The mean TPM was calculated per cell type, and TSS differential looping quantiles at genes with with nonzero expression were correlated with differential expression quantiles of the same genes. Only 10 kb segments of the genome that contained a single annotated gene were considered to avoid errors in attribution of looping signal per 10 kb bin. For genes with multiple annotated TSSs the 10 kb bin corresponding to the median TSS was used. Significance was assessed by the cor.test function in R.

**Comparing HiChIP to Hi-C, ATAC-seq, ChIP-seq, and Capture-C datasets**

T cell subset ChIP-seq data was obtained from the WashU Roadmap repository. A union set of peaks were called using MACS2 and peaks were quantified using Bedtools intersect. Normalization for ChIP was performed using quantile normalization using "preprocessCore" package in R.

Differential EIS was determined using TMM normalization in the "edgeR" package in R. The significantly differential EIS ($\log_2$ fold-change > 1 and FDR < 10%) were determined for each pairwise comparison. For each differential EIS, the maximum ATAC/ChIP signal peak was assessed in each 10 kb anchor (to bias against low signal peaks) and then the maximum $\log_2$ fold-change was compared to the differential EIS.

Differential 1D HiChIP was determined using TMM normalization in the "edgeR" package in R. The significantly differential 1D HiChIP in EIS 10 kb anchors ($\log_2$ fold-change > 1 and FDR < 10%) were determined for each pairwise comparison. For each differential 1D HiChIP anchor, the maximum ATAC/ChIP signal peak was assessed (to bias against low signal peaks) and then the maximum $\log_2$ fold-change was compared to the differential 1D HiChIP.

To determine if reference cell line Hi-C data with primary T cell H3K27ac ChIP-seq could recapitulate EIS in primary T cells, GM Hi-C loop anchors were binned with increasing T cell subset ChIP signal as well as GM H3K27ac ChIP-seq signal (Encode). Loop overlap was then determined for the different H3K27ac ChIP signal bins with HiChIP loops.

817
818 To compare directly with $CD4^+$ Capture-C data, CHiCAGO loops were called in Naïve,
819 $T_H17$, and $T_{reg}$ datasets. CHiCAGO calls were then combined into a union set, and loop
820 signal was obtained for the biological replicates of each T cell subtype using the
821 Juicebox tools dump command. We then used edgeR to identify loops with significant
822 changes in signal among pair-wise comparisons (FDR < 0.05). Total and differential
823 loops were then overlapped with $CD4^+$ Capture-C CHiCAGO data.
824
825 **Distance-matched eQTL SNP-TSS Comparisons**
826 We obtained three groups of eQTL SNP-TSS pairs within 1 Mb distance for HiChIP EIS
827 comparisons. The treatment group contains CD4 eQTL-TSS targets. We have two
828 distance-matched groups as control. The first control group contains the CD4 eQTL
829 SNP-random TSS pairs such that the distance between eQTL SNP and random TSS
830 differs by at most 5 kb with the treatment group. The second control group contains liver
831 eQTL SNP-TSS targets that are also distance matched with the treatment group. The
832 random eQTL SNP-TSS pairs were generated by individual chromosome, such that
833 number of control pairs and treatment pairs are the same for every chromosome. In
834 total, there were 158,482 distance-matched eQTL-TSS pairs. We compared the 5 kb
835 resolution EIS among the three eQTL SNP-TSS groups for all three T cell subtypes.
836 Results show that in all the cases, the EIS between CD4 eQTL-TSS targets were
837 significantly higher than the two control groups (p-value < $10^{-16}$).
838
839 **Distance-matched Fine-mapped SNP-TSS Comparisons**
840 We obtained a list of putatively causal SNPs from the PICS SNP list[25] (PICS probability
841 > 0.5), as well as fine-mapped SNPs associated with inflammatory bowel disease
842 (IBD)[38] or Type 1 Diabetes (T1D)[39]. Next, we obtained all SNPs in LD with each
843 putatively causal SNP using European linkage disequilibrium blocks determined by all
844 SNPs with an $r^2 \geq 0.8$ with the SNP being considered. For the fine-mapped (T1D/IBD)
845 sets, using SNPs in LD with highly significant GWAS SNPs may mean that there are
846 several SNPs of equal or greater significance in the control set, but we still expect an
847 enrichment relative to the LD block.
848
849 We collected all the synthetic pairs between the putatively causal immune-disease
850 related SNPs (IBD, T1D and PICS) and nearby genes within 300 kb distance. To
851 perform the distance-matched EIS comparisons, for each fine-mapped SNP category,
852 we selected the SNP-TSS control pairs which satisfy two constraints: (1) the selected
853 control SNP is positioned at least 5 kb away from the fine-mapped SNP in the same
854 linkage disequilibrium block; (2) the distance of SNP-TSS control pair differs with fine-
855 mapped SNP and the target gene at most 5 kb.
856
857 **Calculation of Disease-specific GWAS SNP Enrichment in Loop Anchors**
858 We categorized GWAS SNPs into sets relevant to diseases of the immune system and,
859 separately, diseases with no known immune component[25]. For each disease in the
860 immune or non-immune set, we determined the proportion of all GWAS SNPs

associated with that disease which overlap the positions of loop anchors based on a union set of loops identified in Naïve, $T_H17$, and $T_{reg}$ cells. The ratio of the proportion of immune and non-immune overlaps relative to a shuffled control was reported as the enrichment of immune GWAS SNPs.

**SNP-TSS Loop Analyses**

We obtained 7,747 PICS SNPs that are associated with autoimmune disease or non-immune traits[25]. 4,331 (55.9%) were associated with autoimmune disease, and 3,416 (44.1%) were associated with non-immune traits. In addition we obtained a set of SNPs associated with six overlapping autoimmune diseases using the GRASP catalog (genome-wide significance p-value $< 10^{-8}$).

We constructed a synthetic loop set for immune and non-immune SNPs and any TSS within 1 Mb of each SNP. We then assigned VCsqrt signal in each biological replicate of the three T cell subtypes to the synthetic loop set, as described above.

VCsqrt signal per sample was quantile-quantile normalized as above. In this analysis, we did not restrict to HiCCUPS-identified loops but instead examined all possible interactions between a SNP and TSS within 1 Mb. Many of these interactions do not exist, and therefore had little or no matrix-balanced signal supporting them. While we removed all SNP TSS pairs below an average of 1 normalized read per sample from subsequent analyses, in general these false interactions contributed little to the overall differential signal for a trait.

H3K27ac data was downloaded from the WashU Roadmap repository. PICS SNPs were taken from Farh, et al[25]. Rather than require strict membership within H3K27ac peaks, PICS SNPs were labeled as active if they lied within 8 kb of a peak, raising the number of nominally functional SNPs from ~700 to ~3200 per cell type, out of 7735 total candidate SNPs.

Differential looping across cell types was assessed by one-sided t-test per trait and activity partition if there were at least eight PICS SNP-TSS pairs in the partition. $T_H17$ bias was defined as $T_H17$ total loop signal minus Naïve total loop signal; $T_{reg}$ bias was defined as $T_{reg}$ total loop signal minus Naïve total loop signal; and Naïve bias was defined as Naïve total loop signal minus one-half times $T_H17$ and $T_{reg}$ total loop signal. For the main text figure, Naïve bias was assessed only using SNPs that were active in Naïve cells, $T_H17$ bias from SNPs active in $T_H17$ and not Naïve, and $T_{reg}$ bias from SNPs active in $T_{reg}$ and not Naïve. p-values were corrected for multiple hypothesis testing by the Holm method using the p.adjust function in R. Bias assessed from SNPs with the opposing cell type specificities (e.g. Naïve bias using SNPs active in $T_H17$ and $T_{reg}$ but not Naïve) yielded no significant hits after correction.

**Correction for False Positives**

Despite not being originally implemented in enrichment interaction datasets, we have previously demonstrated that KR and VC matrix balancing corrects for false positives caused by high 1D fragment visibility[16]. Loop calls are either from HiCCUPS or Fit Hi-C, which are KR and VC matrix balanced, respectively. For differential analysis of HiCCUPS (Figure 3) and SNP-TSS Fit Hi-C (Figure 4c) loop calls, we are restricted within a set of loops that were obtained from matrix balancing. Therefore, while differential loops can be driven by both changes in looping strength and/or 1D ChIP signal, the final interactions being observed are loops. Additionally, we performed differential analysis of HiCCUPS loop calls using two separate methods – one using VCsqrt normalized reads and another with non-normalized reads (edgeR) and found high agreement. The SNP-TSS synthetic loop analysis in Figure 4b was not restricted to loop calls, however was performed with VCsqrt matrix balanced reads to avoid false positives. Those SNP-TSS synthetic loops were then subset by overlap with Fit Hi-C for further differential analyses in Figure 4c. Finally, virtual 4C analysis was performed on non-normalized reads to highlight EIS contributions of both 1D and 3D signal changes, however HiCCUPS loop calls are included in the relevant anecdotes.

**SNP and TSS Connectivity Analysis**

Among all immune and non-immune SNPs, 2,562 (33.1%) were located within a 25 kb region centered around TSS, 2,618 (33.8%) were located in a gene body, and the remaining 2,567 (33.1%) were located in intergenic regions.

To capture the confident SNP-TSS connections for SNP-TSS pair identification, we overlapped the anchors of significant loops that were identified by Fit-Hi-C[56] with the SNP/TSS locations. We identified 14,738 SNP-TSS pairs that were supported by Fit-Hi-C loops, and there were 3,046 unique SNPS connected with at least one gene. Among those 3,046 SNPs, there were 2,181 (71.6%) and 865 (28.4%) SNPs annotated with immune and non-immune SNPs, respectively. As expected, the immune disease SNPs are more likely to connect with genes (Fisher's exact test, p-value = $4.8 \times 10^{-85}$).

**Phasing of HCASMC Samples**

We used BEAGLE 4.1 to impute and phase recalibrated variants using 1000 Genome phase 3 version 5a as a reference panel. The Beagle phasing algorithm was set on the following criteria. At each iteration that the algorithm performs, phased input data are used to build a local model of a haplotype-cluster. After the local haplotype-cluster model is created, for each individual phased haplotypes are sampled using the induced diploid HMM conditional on the individual's genotypes. The sampled haplotypes are then used as the input to feed in the next iteration of the algorithm. In the final iteration, instead of sampling haplotypes, BEAGLE uses the Viterbi algorithm to select the conditional on the diploid HMM and the individual's genotype data and to obtain the haplotypes for each individual that possess the greatest probability, and these most-likely haplotypes are the final output of the BEAGLE phasing algorithm.

**Allelic Mapping Bias Simulation**

948 We constructed a personal genome by editing the reference genome (hg19) according
949 to SNP information. SNPs labeled as “1|0” or “1|1” in the vcf file were replaced with the
950 alternative allele for genome 1. SNPs labeled as “0|1” or “1|1” in the vcf file were
951 replaced with the alternative allele for genome 2. Loop anchors were extended 100 bp
952 in both directions and sequences were extracted using samtools for each
953 genome. ~20X reads were simulated for each region, using wgsim with parameters “ -e
954 0.01 -d 100 -s 20 -1 75 -2 75 -S -1 -h -R 0.1”. Simulated reads were mapped to the
955 “N” masked genome. Mapping parameters were the same used by HiC-Pro (“--very-
956 sensitive -L 30 --score-min L, -0.6, -0.2 --end-to-end --reorder --phred33-quals). Allelic
957 specific reads were separated according to the SNP information and then counted for
958 each loop anchor using bedtools.
959
960 **ATAC-seq**
961 Cells were isolated and subjected to ATAC-seq as previously described[16]. Briefly,
962 55,000 cells were pelleted, resuspended in 50 µL lysis buffer (10mM Tris-HCl, pH 7.4,
963 3mM MgCl2, 10mM NaCl, 0.1% NP-40 (Igepal CA-630), and immediately centrifuged at
964 500 rcf for 10 min at $4^0$C. The nuclei pellets were resuspended in 50 µL transposition
965 buffer (25 µl 2X TD buffer, 22.5 µL $dH_20$, 2.5 µL Illumina Tn5 transposase), and
966 incubated at $37^0$C for 30 min. Transposed DNA was purified with MinElute PCR
967 Purification Kit (Qiagen), and eluted in 10 µL EB buffer.
968
969 **ATAC-seq Data Processing**
970 Adapter sequence trimming using SeqPurge and mapping to hg19 using Bowtie2 were
971 performed. These reads were then filtered for mitochondrial reads, low quality, and PCR
972 duplicates. The filtered reads for each sample were merged and peak calling was
973 performed by MACS2. Each individual sample reads in peaks were quantified using
974 Bedtools intersect with the MACS2 narrow peaks. Peak counts were then combined into
975 a matrix NxM where N represents called peaks and M represents the samples and each
976 value Di,j represents the peak intensity for the respective peak i in sample j. This matrix
977 was then normalized using the “CQN” package in R to minimize bias in GC content and
978 length.
979
980 **CRISPRi Validation of HiChIP Targets**
981 For virus production, $5 \times 10^6$ of HEK293T cells were plated per 10 cm plate. The
982 following day, plasmid encoding lentivirus was co-transfected with pMD2.G and psPAX2
983 into the cells using Lipofectamine 3000 (Thermo Fisher, L3000) according to the
984 manufacturer’s instructions. Supernatant containing viral particles was collected 48
985 hours after transfection and filtered. For lentivirus encoding individual sgRNAs, virus
986 was concentrated 10-fold using Lenti-X concentrator (Clontech, 631232) and stored at -
987 80°C.
988
989 In order to generate a My-La cell line expressing CRISPRi, $2 \times 10^6$ of My-La cells were
990 plated per T75 flask. A dCas9-BFP-KRAB-2A-Blast construct was generated by
991 inserting a 2A-Blast cassette into dCas9-BFP-KRAB (Addgene 46911). 24 hours after

992 plating, lentivirus harboring the dCas9-KRAB construct was added with polybrene (4 µg
993 / mL). Media was changed 24 hours after infection, and then again 48 hours after
994 infection with Blasticidin (Thermo Fisher, A1113903) at 4 µg / mL. Blasticidin resistant
995 cells were selected for eight days with changing media every other day.
996
997 Three different U6 were used for transcription of three different sgRNAs targeting the
998 candidate enhancers, as previously described[57]. For *MYC* locus CRISPRi experiments,
999 each enhancer was targeted by one guide, and therefore all *MYC* GM or My-La
1000 enhancers together were targeted in one experiment. For the PICS SNP CRISPRi
1001 experiments, three guides were targeted to a single SNP-containing enhancer. One of
1002 three sgRNAs was cloned into a lentiviral vector with a human (pMJ117, Addgene
1003 85997), mouse (pMJ179, Addgene 85996) or bovine (pMJ114, Addgene 85995) U6
1004 promoter. These U6-sgRNA constructs were then combined into lentivirus with a
1005 Puromycin-2A-mCherry vector, which was modified from Addgene 46914. My-La-
1006 CRISPRi cells were infected with lentivirus harboring 3 sgRNAs and selected by
1007 Puromycin (Thermo Fisher, A11138) at a final concentration of 1 µg / mL. Previously
1008 reported sgRNAs targeting VPS54 or SEC24C were used for validating CRISPRi
1009 functionality in My-La cell line[58].
1010
1011 For readout of CRISPRi validation, we performed qRT-PCR and cell growth assays on
1012 three biological and two technical replicates. For qRT-PCR, RNA was Trizol extracted
1013 (Thermo Fisher, 15596026) and purified using the Zymo RNA Clean and Concentrator
1014 kit (Zymo Research, R1016). qRT-PCR was performed with Brilliant qRT-PCR
1015 mastermix (Agilent, 600825). Ct values were measured by using Lightcycler 480
1016 (Roche) and relative expression level was calculated by ddCt method compared to a
1017 *GAPDH* control. Primer sequences are listed in Supplementary Table 6. For cell growth,
1018 we used the CellTiter-Glo kit (Promega, G7572) according to the manufacturers
1019 instructions. Statistics for both RNA and cell growth changes were calculated using a
1020 Student's t test against the non-targeting control.
1021
1022 **CRISPRa Validation of HiChIP Targets**
1023 Jurkat cells were transduced with a lentiviral dCas9-VP64-2A-GFP expression vector
1024 (Addgene 61422). Single GFP+ cells were sorted by FACS into the wells of a 96-well
1025 plate, and a clone with bright uniform GFP expression were selected for use in future
1026 experiments.
1027
1028 sgRNAs were cloned in arrayed format for CD69 HiChIP peaks falling outside the range
1029 of the tiling CRISPRa screen[26]. sgRNAs were chosen based on high predicted on-target
1030 activity[59] and low predicted off-target activity[60]. sgRNAs were cloned into the lentiviral
1031 expression vector "pCRISPRia-v2" (Addgene 84832) as described in Horlbeck et al[61].
1032 Lentivirus was produced by transfecting HEK293T with standard packaging vectors
1033 using *Trans*IT-LTI Transfection Reagent (Mirus, MIR 2306). Media was changed 24
1034 hours post-transfection. Viral supernatant was harvested at 48 and 72 hr following
1035 transfection and immediately used for infection of Jurkat-dCas9-VP64 cells.

1036
1037 Jurkat-dCas9-VP64 cells were infected with lentiviral sgRNAs by resuspending cells in a
1038 1:1 mix of fresh media and lentiviral supernatant at a final concentration of $0.25 \times 10^6$
1039 cells/mL with 5 µg / mL polybrene. Cells were spinfected for 1 hour at 1000 rcf, 32 °C.
1040 The next day, half of the media was removed and replaced with fresh lentiviral
1041 supernatant, and the spinfection was repeated. The next day, the cells were
1042 resuspended in fresh media with 1.5 µg / mL puromycin and cultured for 2 days to
1043 remove uninfected cells. For readout of CRISPRa validation, we performed qRT-PCR
1044 and FACS on two biological and two technical replicates. RNA extraction and qRT-PCR
1045 was performed as described above. Expression of CD69 on infected cells (GFP+BFP+)
1046 was analyzed by flow cytometry with an Attune NxT flow cytometer (Life Technologies).
1047 Statistics for both RNA and protein level changes were calculated with a one-way
1048 ANOVA followed by a Dunnet's multiple comparisons test against the non-targeting
1049 control.
1050
1051 **SUPPLEMENTARY FIGURE LEGENDS**
1052
1053 **Supplementary Figure 1. H3K27ac HiChIP enriches E-P-associated chromatin**
1054 **contacts. (A)** Schematic of chromatin contacts captured in H3K27ac HiChIP. **(B)** Loop
1055 call overlap for cohesin HiChIP and H3K27ac HiChIP in mES cells. **(C)** Contact distance
1056 distribution for loops that are biased in cohesin versus H3K27ac HiChIP. **(D)** Proportion
1057 of cohesin and H3K27ac biased HiChIP loops that have cohesin, CTCF, H3K27ac, and
1058 RNA polymerase II binding in at least one loop anchor. **(E)** Virtual 4C interaction profile
1059 of an H3K27ac-biased loop focused at the *MALAT1* promoter. **(F)** Virtual 4C interaction
1060 profile of a cohesin-biased loop with associated low transcriptional activity.
1061
1062 **Supplementary Figure 2. H3K27ac HiChIP achieves high chromatin loop signal-**
1063 **over-background at low cell inputs. (A)** KR balanced interaction matrices focused
1064 around the *Etv5* locus in mES cells with decreasing cellular starting material. **(B)** Read
1065 support reproducibility of loops in H3K27ac HiChIP libraries from 25 million cells
1066 compared to HiChIP in lower cell input libraries. **(C)** Aggregate peak analysis of loops in
1067 mES H3K27ac HiChIP libraries.
1068
1069 **Supplementary Figure 3. H3K27ac HiChIP generates reproducible chromatin loop**
1070 **signals at low cell inputs. (A)** Comparison of KR balanced interaction maps in
1071 H3K27ac HiChIP biological replicates. **(B)** Read support reproducibility of loops
1072 between H3K27ac HiChIP biological replicates. **(C)** HiCCUPS loop call overlap between
1073 H3K27ac HiChIP libraries from 25 million and 50 thousand mES cells. **(D)** Preseq library
1074 complexity estimation of H3K27ac HiChIP libraries from 25 million and 50 thousand
1075 mES cells.
1076
1077 **Supplementary Figure 4. H3K27ac HiChIP biological replicates from primary**
1078 **sorted T cells are highly reproducible. (A)** FACS strategy for Naïve, $T_H17$, and $T_{reg}$
1079 cells starting from human peripheral blood. Post-sort validation was used to ensure

purity of T cell subtypes. Number represents percent of cells within that gate. **(B)** KR balanced interaction map of T cell subtype biological replicates. **(C)** Read support reproducibility of loops between H3K27ac HiChIP biological replicates in Naïve, $T_H17$, and $T_{reg}$ cells. **(D)** Aggregate peak analysis of loops in Naïve, $T_H17$, and $T_{reg}$ H3K27ac HiChIP libraries.

**Supplementary Figure 5. Validation of HiChIP-identified *CD69* distal enhancers with CRISPR activation.** CRISPRa validation in Jurkat cells of *CD69* distal enhancers. *CD69* protein levels are shown for individual sgRNAs tiling H3K27ac HiChIP-identified distal *CD69* enhancers relative to the *KLRF2* promoter as a locus negative control and a non-targeting negative control.

**Supplementary Figure 6. Global enhancer connectome characterization in T cell differentiation. (A)** ChromHMM classification of union T cell loop anchors. **(B)** Contact distance distribution of union T cell loops. **(C)** Agreement in residual signal observed between sample signals per loop after removing imputed shared signal, calibrated by a null distribution of random pairings of loops. QQ plot shows modest enrichment above random pairings. PCA on residual signal clusters samples first by Naive vs Memory cell types (PC1), and then by donor identity (PC2, 3). **(D)** Overlap of differential interactions between Naïve, $T_H17$, and $T_{reg}$ subtypes. Biased interactions were obtained by performing pairwise comparisons between T cell types and analyzing the top 5% enriched and top 5% depleted EIS in each T cell subtype. **(E)** ChromHMM annotation of total loops, differential loops, and shared loops in all three T cell subtypes. O corresponds to other loop anchors not classified as enhancer or promoter. **(F)** Number of connections for different classes of loop elements. **(G)** Quantification of promoters skipped before an enhancer reaches its loop target.

**Supplementary Figure 7. Positioning of differential HiChIP contacts in gene dense chromosomes. (A)** Distribution of T cell subtype differential HiChIP contacts across different chromosomes compared to the distribution of all loops. **(B)** Correlation of differential loop density with gene density and GC content.

**Supplementary Figure 8. Characterization of conformational landscapes surrounding key T cell regulatory factors. (A – C)** Virtual 4C interaction profiles anchored at the promoters of canonical Naïve, $T_H17$, and $T_{reg}$ regulatory factors.

**Supplementary Figure 9. Chromosome conformation dynamics of canonical T cell regulatory factors. (A – C)** Delta interaction maps focused around known Naïve, $T_H17$, and $T_{reg}$ regulatory factors.

**Supplementary Figure 10. Contribution of H3K27ac ChIP and chromosome conformation to HiChIP EIS. (A)** (Left) Proportion of H3K27ac ChIP peaks within EIS differential loop anchors that are cell-type specific ($\log_2$ fold change > 1) or shared across all three subtypes. (Right) Global correlation of EIS and H3K27ac ChIP fold-

1124 change in different T cell subset pairwise comparisons. **(B)** Same as (A), but using
1125 HiChIP 1D differential signal at EIS biased loop anchors. **(C)** Overlap of H3K27ac
1126 HiChIP and bins of Hi-C loops with increasing T cell subset and GM H3K27ac ChIP-seq
1127 signal. **(D)** Overlap of $CD4^+$ Capture Hi-C[14] with total and differential T cell subset
1128 HiChIP loops. **(E)** $T_{reg}$-specific loops at the *LRRC32* promoter not observed in other
1129 H3K27ac HiChIP T cell subsets nor in $CD4^+$ Capture Hi-C data[14].
1130
1131 **Supplementary Figure 11. Enrichment of autoimmune SNPs in T cell HiChIP loop**
1132 **anchors. (A)** Enrichment of specific PICS autoimmune disease and non-immune SNPs
1133 in anchors of loops called by Juicer and Fit-Hi-C compared to a background shuffled
1134 loop set. **(B)** Enrichment of all PICS autoimmune disease and non-immune SNPs in T
1135 cell subset biased loop anchors and all anchors compared to a background shuffled
1136 loop set.
1137
1138 **Supplementary Figure 12. T cell subtype HiChIP specificity of autoimmune SNPs.**
1139 **(A)** H3K27ac HiChIP signal bias in T cell subsets for PICS SNP-TSS pairs grouped by
1140 each SNP's presence in cell type-specific or shared H3K27ac ChIP peaks up to 8 kb
1141 away. **(B)** H3K27ac HiChIP signal bias in T cell subsets for PICS SNP-TSS pairs
1142 grouped by each SNP's presence in cell type-specific or shared H3K27ac ChIP peaks
1143 up to 2.5 kb away. **(C)** H3K27ac HiChIP signal bias in GM, K562, and My-La cell lines
1144 for PICS SNP-TSS pairs grouped by each SNP's presence in T cell subset-specific or
1145 shared H3K27ac ChIP peaks up to 8 kb away. **(D)** H3K27ac HiChIP signal bias in T cell
1146 subsets for GRASP SNP-TSS pairs grouped by each SNP's presence in cell type-
1147 specific or shared H3K27ac ChIP peaks up to 2.5 kb away. **(E)** Average number of
1148 HiChIP gene targets for non-genic autoimmune disease and non-immune SNPs. **(F)**
1149 Quantification of promoters skipped before a SNP reaches its gene target. **(G)**
1150 Quantification of SNP HiChIP gene targets in autoimmune disease.
1151
1152 **Supplementary Figure 13. Validation of HiChIP signal at SNP-eQTL contacts. (A)**
1153 Validation of HiChIP signal at SNP-TSS pairs using interaction profiles of eQTL SNPs to
1154 ensure they contact their associated target gene promoter. **(B)** Interaction profiles of
1155 CRISPRi-validated loci in My-La.
1156
1157 **Supplementary Figure 14. H3K27ac HiChIP fine-mapping of GWAS variants in**
1158 **haplotype blocks. (A)** Global validation of HiChIP signal at putatively causal SNPs
1159 versus corresponding SNPs in LD ($r^2 \geq 0.8$) for Naïve and $T_{reg}$ cell subtypes. SNP-TSS
1160 pairs were generated from published fine-mapped datasets, compared to a distance-
1161 matched SNP-TSS pair set in the same LD block. **(B)** Interaction profile of the *SATB1*
1162 promoter, and a 1 kb resolution visualization of the SNP-containing enhancer of interest.
1163 Changes in bias between 5 kb and 1 kb resolution reflects EIS focused 1 kb around the
1164 SATB1 TSS and specific SNPs within the enhancer. LD SNPs ($r^2 \geq 0.8$) correspond to
1165 GRASP SNPs (genome-wide significance p-value $< 10^{-8}$). The highlighted SNP is a
1166 PICS closest to focal EIS to *SATB1.*
1167

**Supplementary Figure 15. Chromatin interaction landscape of the 9p21.3 cardiovascular disease risk locus.** HCASMC v4C interaction profiles focused around the promoters of *CDKN2A*, *CDKN2B*, and *ANRIL* within the 9p21.3 locus.

## Supplementary Table Legends

**Supplementary Table 1. HiChIP data processing metrics.** HiC-Pro mapping, Hi-C filtering, duplicate removal, and interaction length statistics for all HiChIP libraries.

**Supplementary Table 2. HiCCUPS high confidence loop calls.** High confidence loop calls for all HiChIP libraries.

**Supplementary Table 3. HiCCUPS differential EIS in T cell subtypes by edgeR.** High confidence loop calls in T cell subsets with edgeR significance and fold-change for pair-wise comparisons.

**Supplementary Table 4. HiCCUPS differential EIS in T cell subtypes.** Top 5% of EIS ranked by cell-type bias for each pair-wise comparison between different T cell subtypes.

**Supplementary Table 5. HiChIP gene targets of autoimmune and CAD SNPs.** SNP type (within gene, at TSS, or intergenic), associated disease, and corresponding loop annotated with the HiChIP gene target. Included are edgeR significance and fold-change for cell-type comparisons of interest.

**Supplementary Table 6. sgRNA and primer oligonucleotide sequences.** sgRNA and primer sequences used throughout the study.

## REFERENCES

1. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
2. van Arensbergen, J., van Steensel, B. & Bussemaker, H. J. In search of the determinants of enhancer-promoter interaction specificity. *Trends Cell Biol.* **24,** 695–702 (2014).
3. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326,** 289–293 (2009).
4. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159,** 1665–1680 (2014).

1206 5. Fulco, C. P. *et al.* Systematic mapping of functional enhancer-promoter connections with
1207 CRISPR interference. *Science* aag2445 (2016). doi:10.1126/science.aag2445
1208 6. Altshuler, D., Daly, M. J. & Lander, E. S. Genetic mapping in human disease. *Science* **322,**
1209 881–888 (2008).
1210 7. Kumar, V. *et al.* Human Disease-Associated Genetic Variation Impacts Large Intergenic
1211 Non-Coding RNA Expression. *PLOS Genet.* **9,** e1003201 (2013).
1212 8. Edwards, S. L., Beesley, J., French, J. D. & Dunning, A. M. Beyond GWASs: Illuminating the
1213 Dark Road from Association to Function. *Am. J. Hum. Genet.* **93,** 779–797 (2013).
1214 9. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.
1215 *Nucleic Acids Res.* **42,** D1001–D1006 (2014).
1216 10. Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional
1217 connections with IRX3. *Nature* **507,** 371–375 (2014).
1218 11. Claussnitzer, M. *et al.* FTO Obesity Variant Circuitry and Adipocyte Browning in
1219 Humans. *N. Engl. J. Med.* **373,** 895–907 (2015).
1220 12. Grubert, F. *et al.* Genetic Control of Chromatin States in Humans Involves Local and
1221 Distal Chromosomal Interactions. *Cell* **162,** 1051–1065 (2015).
1222 13. Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-
1223 resolution capture Hi-C. *Nat. Genet.* **47,** 598–606 (2015).
1224 14. Javierre, B. M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-
1225 coding Disease Variants to Target Gene Promoters. *Cell* **167,** 1369–1384.e19 (2016).
1226 15. Sanjana, N. E. *et al.* High-resolution interrogation of functional elements in the
1227 noncoding genome. *Science* **353,** 1545–1549 (2016).

1228 16. Mumbach, M. R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed
1229 genome architecture. *Nat. Methods* **advance online publication,** (2016).
1230 17. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of
1231 transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39,** 311–318
1232 (2007).
1233 18. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and
1234 predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 21931–21936 (2010).
1235 19. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental
1236 enhancers in humans. *Nature* **470,** 279–283 (2011).
1237 20. Ferraro, A. *et al.* Interindividual variation in human T regulatory cells. *Proc. Natl.*
1238 *Acad. Sci. U. S. A.* **111,** E1111-1120 (2014).
1239 21. Arvey, A. *et al.* Genetic and epigenetic variation in the lineage specification of
1240 regulatory T cells. *eLife* **4,** e07571 (2015).
1241 22. Bettelli, E. *et al.* Reciprocal developmental pathways for the generation of
1242 pathogenic effector TH17 and regulatory T cells. *Nature* **441,** 235–238 (2006).
1243 23. Acosta-Rodriguez, E. V. *et al.* Surface phenotype and antigenic specificity of human
1244 interleukin 17-producing T helper memory cells. *Nat. Immunol.* **8,** 639–646 (2007).
1245 24. Morita, R. *et al.* Human blood CXCR5(+)CD4(+) T cells are counterparts of T
1246 follicular cells and contain specific subsets that differentially support antibody secretion.
1247 *Immunity* **34,** 108–121 (2011).
1248 25. Farh, K. K.-H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune
1249 disease variants. *Nature* **518,** 337–343 (2015).

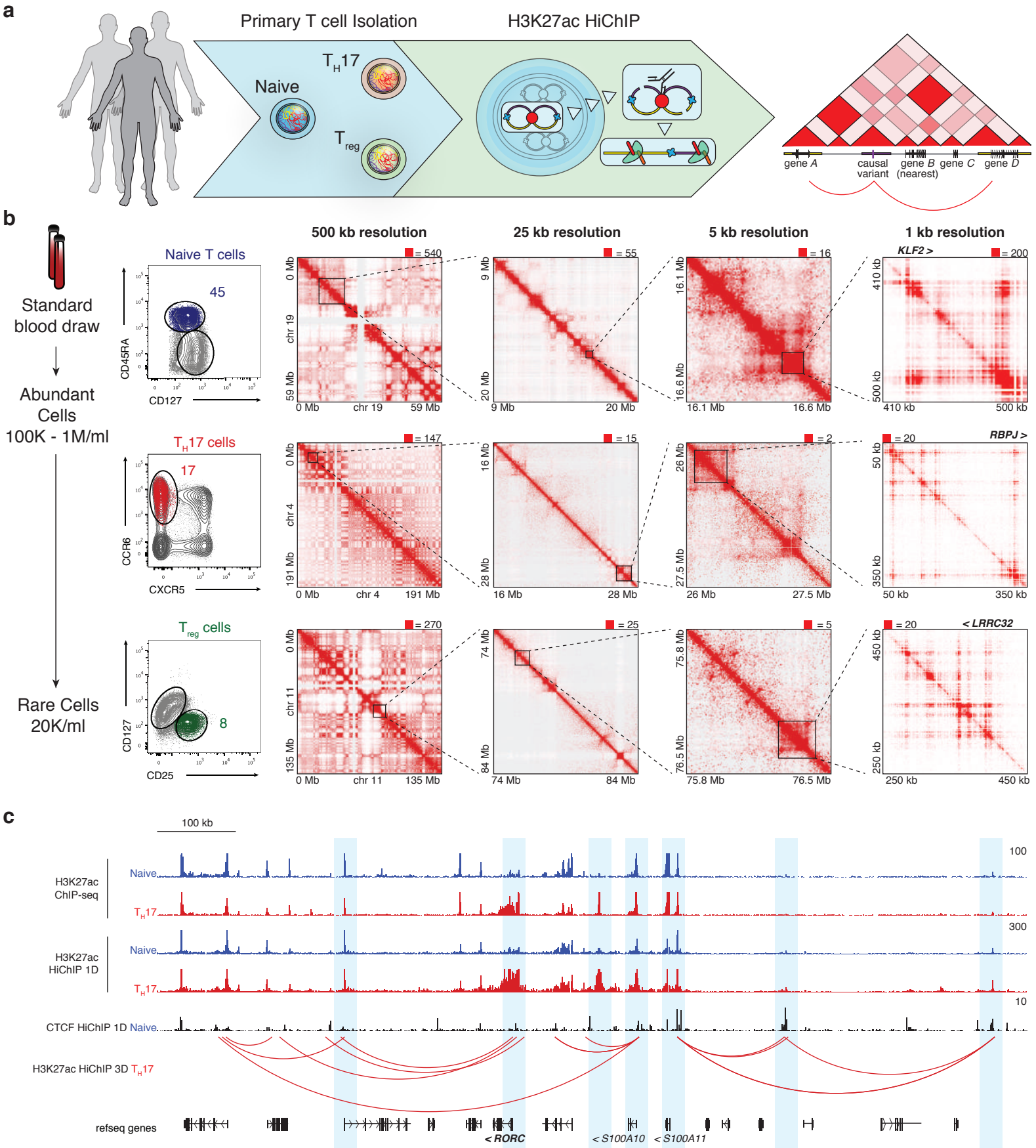
1250 26. Simeonov, D. R. *et al.* Discovery of an autoimmunity-associated IL2RA enhancer by
1251 unbiased targeting of transcriptional activation. *bioRxiv* 091843 (2016).
1252 doi:10.1101/091843
1253 27. Engreitz, J. M. *et al.* Local regulation of gene expression by lncRNA promoters,
1254 transcription and splicing. *Nature* **539,** 452–455 (2016).
1255 28. Dao, L. T. M. *et al.* Genome-wide characterization of mammalian promoters with
1256 distal enhancer functions. *Nat. Genet.* **49,** 1073–1081 (2017).
1257 29. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and
1258 characterization. *Nat. Methods* **9,** 215–216 (2012).
1259 30. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain
1260 formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U. S. A.* **112,**
1261 E6456-6465 (2015).
1262 31. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a
1263 topological basis for transcription regulation. *Cell* **148,** 84–98 (2012).
1264 32. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition
1265 of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-
1266 binding proteins and nucleosome position. *Nat. Methods* **10,** 1213–1218 (2013).
1267 33. Samstein, R. M. *et al.* Foxp3 exploits a pre-existent enhancer landscape for
1268 regulatory T cell lineage specification. *Cell* **151,** 153–166 (2012).
1269 34. Ye, C. J. *et al.* Intersection of population variation and autoimmunity genetics in
1270 human T cell activation. *Science* **345,** 1254665 (2014).
1271 35. Petukhova, L. *et al.* Genome-wide association study in alopecia areata implicates
1272 both innate and adaptive immunity. *Nature* **466,** 113–117 (2010).

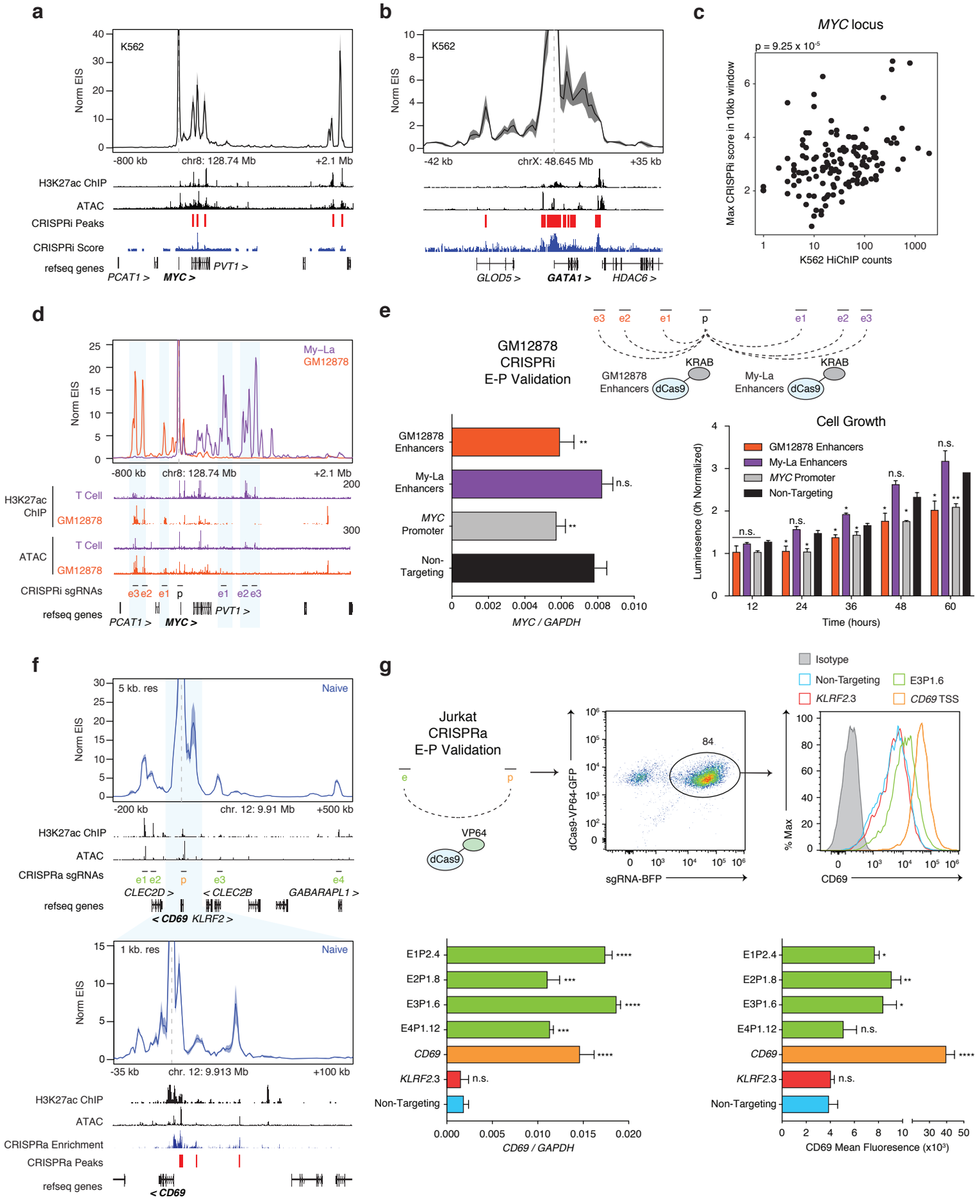
1273 36. Raj, T. *et al.* Polarization of the effects of autoimmune and neurodegenerative risk
1274 alleles in leukocytes. *Science* **344,** 519–523 (2014).
1275 37. Schiering, C. *et al.* The alarmin IL-33 promotes regulatory T-cell function in the
1276 intestine. *Nature* **513,** 564–568 (2014).
1277 38. Huang, H. *et al.* Fine-mapping inflammatory bowel disease loci to single-variant
1278 resolution. *Nature* (2017). doi:10.1038/nature22969
1279 39. Onengut-Gumuscu, S. *et al.* Fine mapping of type 1 diabetes susceptibility loci and
1280 evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat. Genet.*
1281 **47,** 381–386 (2015).
1282 40. O'Shea, J. J., Lahesmaa, R., Vahedi, G., Laurence, A. & Kanno, Y. Genomic views of
1283 STAT function in CD4+ T helper cell differentiation. *Nat. Rev. Immunol.* **11,** 239–250
1284 (2011).
1285 41. Miller, C. L. *et al.* Integrative functional genomics identifies regulatory mechanisms
1286 at coronary artery disease loci. *Nat. Commun.* **7,** 12092 (2016).
1287 42. Nurnberg, S. T. *et al.* Coronary Artery Disease Associated Transcription Factor
1288 TCF21 Regulates Smooth Muscle Precursor Cells That Contribute to the Fibrous Cap.
1289 *PLoS Genet.* **11,** e1005155 (2015).
1290 43. McPherson, R. *et al.* A common allele on chromosome 9 associated with coronary
1291 heart disease. *Science* **316,** 1488–1491 (2007).
1292 44. Helgadottir, A. *et al.* A common variant on chromosome 9p21 affects the risk of
1293 myocardial infarction. *Science* **316,** 1491–1493 (2007).
1294 45. Clarke, R. *et al.* Genetic variants associated with Lp(a) lipoprotein level and coronary
1295 disease. *N. Engl. J. Med.* **361,** 2518–2528 (2009).

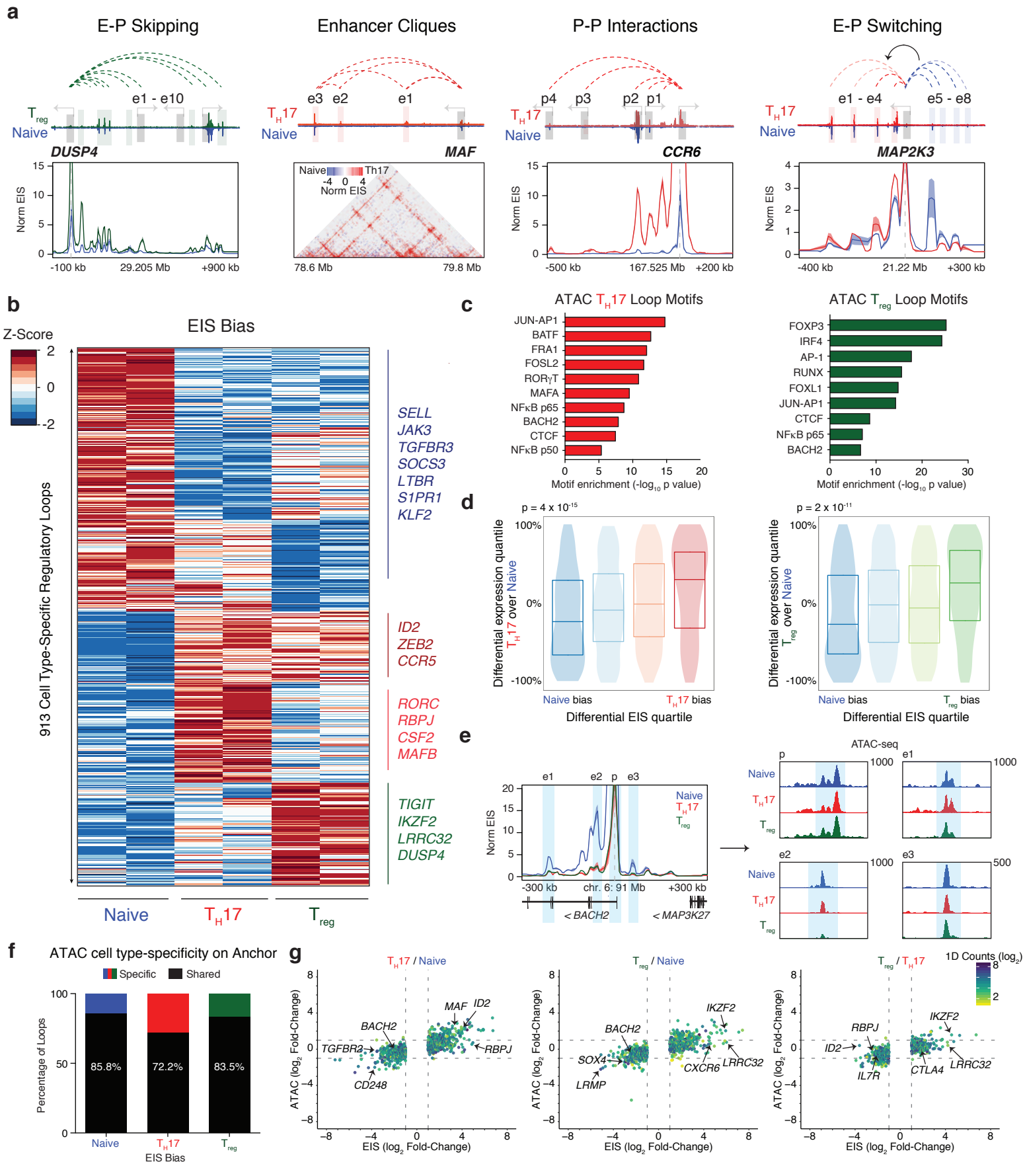
1296 46. CARDIoGRAMplusC4D Consortium *et al.* Large-scale association analysis identifies
1297 new risk loci for coronary artery disease. *Nat. Genet.* **45,** 25–33 (2013).
1298 47. Leung, D. *et al.* Integrative analysis of haplotype-resolved epigenomes across human
1299 tissues. *Nature* **518,** 350–354 (2015).
1300 48. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell
1301 differentiation. *Nature* **518,** 331–336 (2015).
1302 49. Franzén, O. *et al.* Cardiometabolic risk loci share downstream cis- and trans-gene
1303 regulation across tissues and diseases. *Science* **353,** 827–830 (2016).
1304 50. Index Switching Causes "Spreading-Of-Signal" Among Multiplexed Samples In
1305 Illumina HiSeq 4000 DNA Sequencing | bioRxiv. Available at:
1306 http://biorxiv.org/content/early/2017/04/09/125724. (Accessed: 19th May 2017)
1307 51. Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data
1308 processing. *Genome Biol.* **16,** 259 (2015).
1309 52. Gjoneska, E. *et al.* Conserved epigenomic signals in mice and humans reveal immune
1310 basis of Alzheimer's disease. *Nature* **518,** 365–369 (2015).
1311 53. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin
1312 architecture. *Nature* **467,** 430–435 (2010).
1313 54. Yue, F. *et al.* A comparative encyclopedia of DNA elements in the mouse genome.
1314 *Nature* **515,** 355–364 (2014).
1315 55. Bonnal, R. J. P. *et al.* De novo transcriptome profiling of highly purified human
1316 lymphocytes primary cells. *Sci. Data* **2,** 150051 (2015).

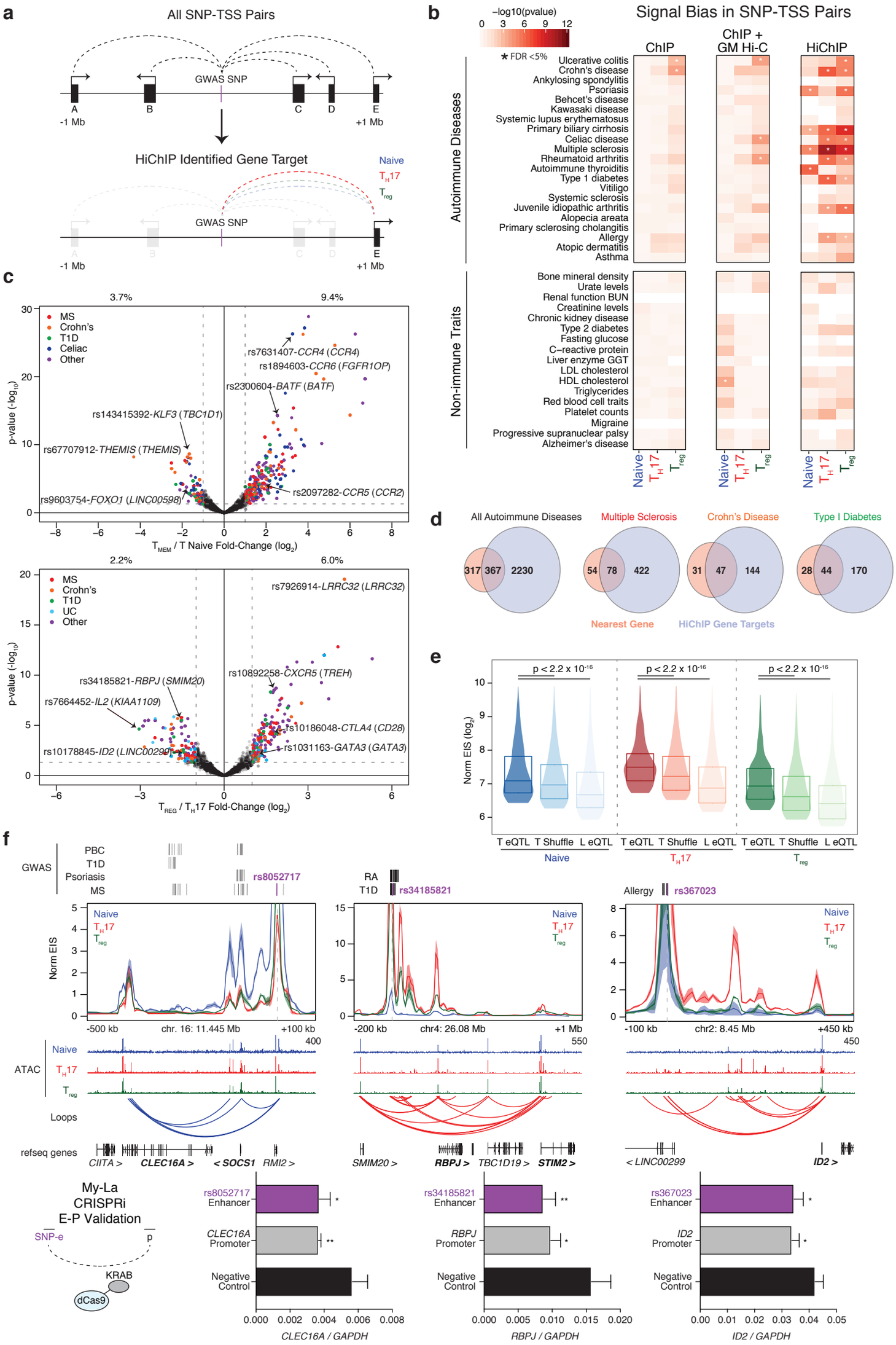
1317 56. Ay, F., Bailey, T. L. & Noble, W. S. Statistical confidence estimation for Hi-C data
1318 reveals regulatory chromatin contacts. *Genome Res.* gr.160374.113 (2014).
1319 doi:10.1101/gr.160374.113
1320 57. Adamson, B. *et al.* A Multiplexed Single-Cell CRISPR Screening Platform Enables
1321 Systematic Dissection of the Unfolded Protein Response. *Cell* **167,** 1867–1882.e21
1322 (2016).
1323 58. Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and
1324 Activation. *Cell* **159,** 647–661 (2014).
1325 59. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-
1326 target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34,** 184–191 (2016).
1327 60. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat.*
1328 *Biotechnol.* **31,** 827–832 (2013).
1329 61. Horlbeck, M. A. *et al.* Nucleosomes impede Cas9 access to DNA in vivo and in vitro.
1330 *eLife* **5,** (2016).

1331
1332

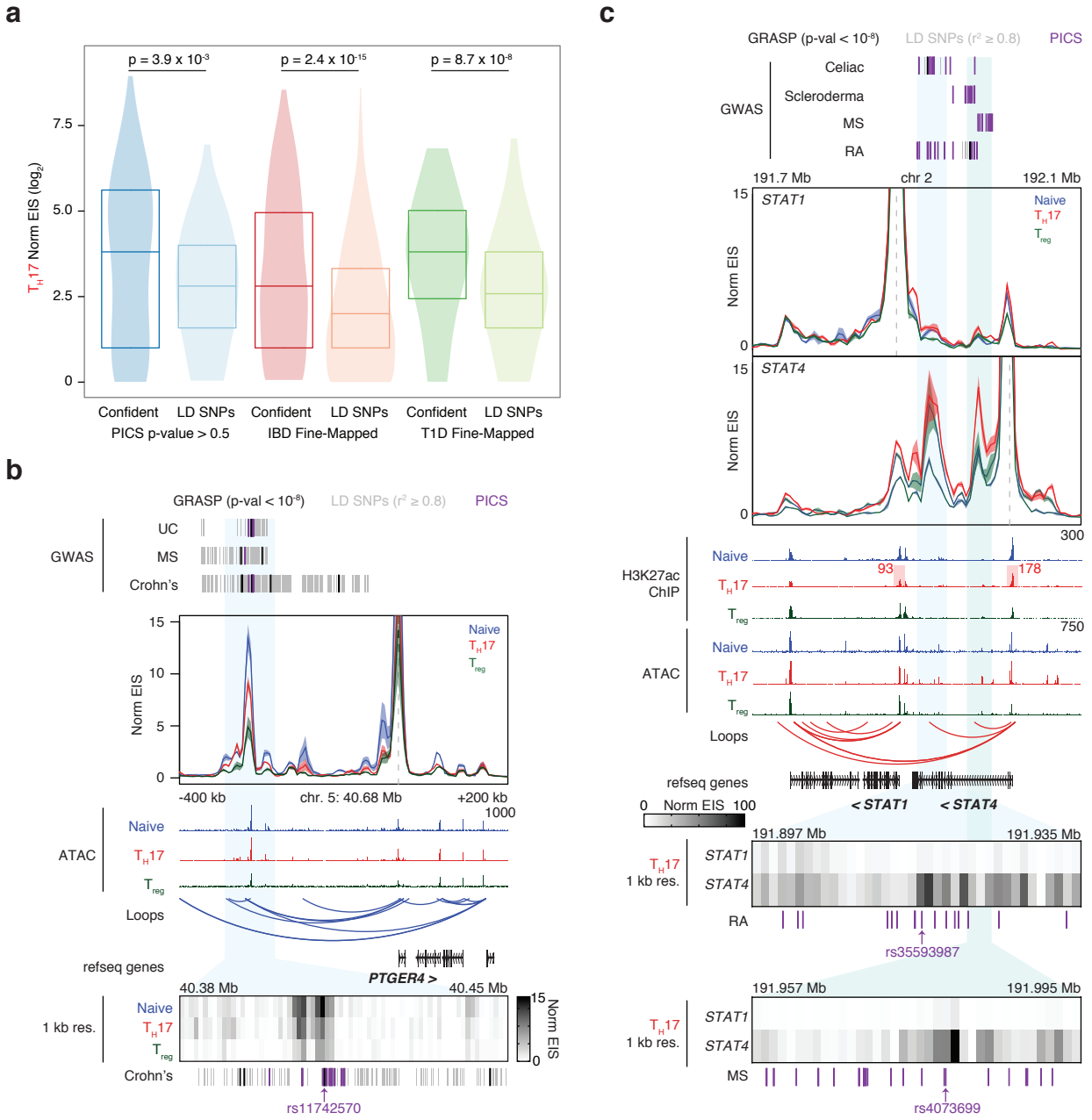
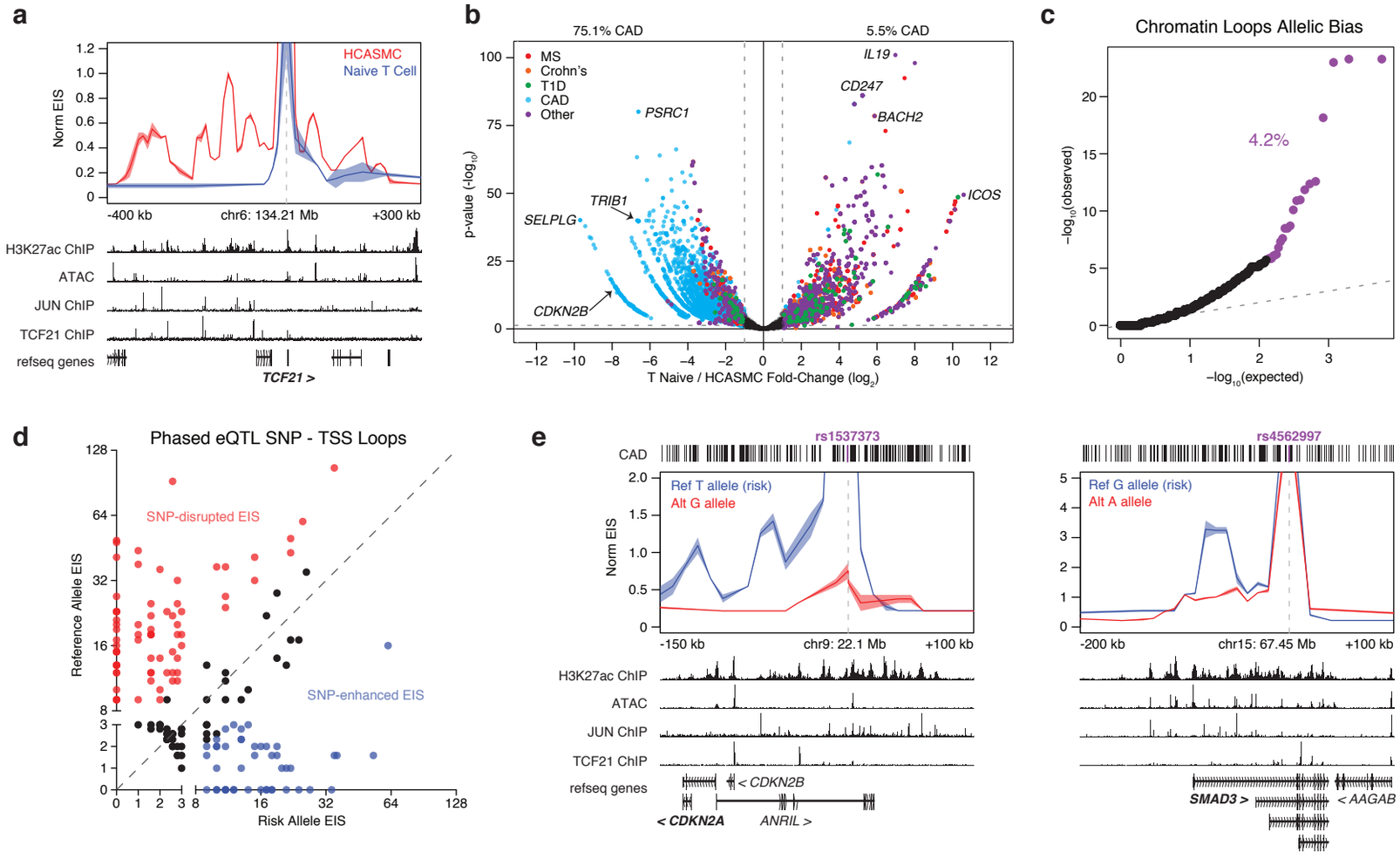

a
Primary T cell Isolation
H3K27ac HiChIP
Naive
$T_H17$
$T_{reg}$
gene A
causal variant
gene B (nearest)
gene C
gene D
b
Standard blood draw
Abundant Cells 100K - 1M/ml
Rare Cells 20K/ml
Naive T cells
45
CD45RA
CD127
$T_H17$ cells
17
CCR6
CXCR5
$T_{reg}$ cells
8
CD127
CD25
500 kb resolution
25 kb resolution
5 kb resolution
1 kb resolution
= 540
= 55
= 16
KLF2 >
= 200
= 147
= 15
= 2
RBPJ >
= 20
= 270
= 25
= 5
< LRRC32
= 20
chr 19
chr 4
chr 11
0 Mb
59 Mb
9 Mb
20 Mb
16.1 Mb
16.6 Mb
410 kb
500 kb
191 Mb
16 Mb
28 Mb
26 Mb
27.5 Mb
50 kb
350 kb
135 Mb
74 Mb
84 Mb
75.8 Mb
76.5 Mb
450 kb
250 kb
c
100 kb
H3K27ac ChIP-seq
Naive
$T_H17$
100
H3K27ac HiChIP 1D
300
CTCF HiChIP 1D Naive
10
H3K27ac HiChIP 3D $T_H17$
refseq genes
< RORC
< S100A10
< S100A11


a
K562
Norm EIS
-800 kb
chr8: 128.74 Mb
+2.1 Mb
H3K27ac ChIP
ATAC
CRISPRi Peaks
CRISPRi Score
refseq genes
PCAT1 >
MYC >
PVT1 >
b
K562
Norm EIS
-42 kb
chrX: 48.645 Mb
+35 kb
GLOD5 >
GATA1 >
HDAC6 >
c
MYC locus
p = 9.25 x 10-5
Max CRISPRi score in 10kb window
K562 HiChIP counts
d
My–La
GM12878
Norm EIS
-800 kb
chr8: 128.74 Mb
+2.1 Mb
H3K27ac ChIP
T Cell
GM12878
ATAC
CRISPRi sgRNAs
e3 e2 e1 p e1 e2 e3
refseq genes
PCAT1 >
MYC >
PVT1 >
e
GM12878 CRISPRi E-P Validation
KRAB
GM12878 Enhancers
My-La Enhancers
dCas9
GM12878 Enhancers
My-La Enhancers
MYC Promoter
Non-Targeting
MYC / GAPDH
Cell Growth
GM12878 Enhancers
My-La Enhancers
MYC Promoter
Non-Targeting
Luminesence (0h Normalized)
Time (hours)
n.s.
f
5 kb. res
Naive
Norm EIS
-200 kb
chr. 12: 9.91 Mb
+500 kb
H3K27ac ChIP
ATAC
CRISPRa sgRNAs
e1 e2 p e3 e4
CLEC2D >
< CLEC2B
GABARAPL1 >
refseq genes
< CD69
KLRF2 >
1 kb. res
Naive
Norm EIS
-35 kb
chr. 12: 9.913 Mb
+100 kb
H3K27ac ChIP
ATAC
CRISPRa Enrichment
CRISPRa Peaks
refseq genes
< CD69
g
Jurkat CRISPRa E-P Validation
e
p
VP64
dCas9
dCas9-VP64-GFP
sgRNA-BFP
84
Isotype
Non-Targeting
E3P1.6
KLRF2.3
CD69 TSS
% Max
CD69
E1P2.4
E2P1.8
E3P1.6
E4P1.12
CD69
KLRF2.3
Non-Targeting
CD69 / GAPDH
CD69 Mean Fluoresence (x10³)


a
E-P Skipping
Enhancer Cliques
P-P Interactions
E-P Switching
e1 - e10
e3 e2 e1
p4 p3 p2 p1
e1 - e4
e5 - e8
Treg
TH17
Naive
DUSP4
MAF
CCR6
MAP2K3
Norm EIS
-100 kb
29.205 Mb
+900 kb
78.6 Mb
79.8 Mb
-500 kb
167.525 Mb
+200 kb
-400 kb
21.22 Mb
+300 kb
b
EIS Bias
Z-Score
913 Cell Type-Specific Regulatory Loops
SELL
JAK3
TGFBR3
SOCS3
LTBR
S1PR1
KLF2
ID2
ZEB2
CCR5
RORC
RBPJ
CSF2
MAFB
TIGIT
IKZF2
LRRC32
DUSP4
c
ATAC TH17 Loop Motifs
JUN-AP1
BATF
FRA1
FOSL2
RORγT
MAFA
NFκB p65
BACH2
CTCF
NFκB p50
Motif enrichment (-log10 p value)
ATAC Treg Loop Motifs
FOXP3
IRF4
AP-1
RUNX
FOXL1
JUN-AP1
CTCF
NFκB p65
BACH2
d
p = 4 x 10-15
p = 2 x 10-11
Differential expression quantile
TH17 over Naive
Treg over Naive
Naive bias
TH17 bias
Treg bias
Differential EIS quartile
e
ATAC-seq
e1 e2 p e3
Naive
TH17
Treg
-300 kb
chr. 6: 91 Mb
+300 kb
< BACH2
< MAP3K27
f
ATAC cell type-specificity on Anchor
Specific
Shared
Percentage of Loops
85.8%
72.2%
83.5%
EIS Bias
g
TH17 / Naive
Treg / Naive
Treg / TH17
1D Counts (log2)
ATAC (log2 Fold-Change)
EIS (log2 Fold-Change)
BACH2
TGFBR2
CD248
MAF
ID2
RBPJ
SOX4
LRMP
IKZF2
CXCR6
LRRC32
IL7R
CTLA4


a
All SNP-TSS Pairs
GWAS SNP
A
B
C
D
E
-1 Mb
+1 Mb
HiChIP Identified Gene Target
Naive
TH17
Treg
b
Signal Bias in SNP-TSS Pairs
−log10(pvalue)
0 3 6 9 12
* FDR <5%
ChIP
ChIP + GM Hi-C
HiChIP
Autoimmune Diseases
Ulcerative colitis
Crohn's disease
Ankylosing spondylitis
Psoriasis
Behcet's disease
Kawasaki disease
Systemic lupus erythematosus
Primary biliary cirrhosis
Celiac disease
Multiple sclerosis
Rheumatoid arthritis
Autoimmune thyroiditis
Type 1 diabetes
Vitiligo
Systemic sclerosis
Juvenile idiopathic arthritis
Alopecia areata
Primary sclerosing cholangitis
Allergy
Atopic dermatitis
Asthma
Non-immune Traits
Bone mineral density
Urate levels
Renal function BUN
Creatinine levels
Chronic kidney disease
Type 2 diabetes
Fasting glucose
C−reactive protein
Liver enzyme GGT
LDL cholesterol
HDL cholesterol
Triglycerides
Red blood cell traits
Platelet counts
Migraine
Progressive supranuclear palsy
Alzheimer's disease
Naive
TH17
Treg
c
3.7%
9.4%
MS
Crohn's
T1D
Celiac
Other
rs7631407-CCR4 (CCR4)
rs1894603-CCR6 (FGFR1OP)
rs2300604-BATF (BATF)
rs143415392-KLF3 (TBC1D1)
rs67707912-THEMIS (THEMIS)
rs9603754-FOXO1 (LINC00598)
rs2097282-CCR5 (CCR2)
p-value (-log10)
TMEM / T Naive Fold-Change (log2)
2.2%
6.0%
MS
Crohn's
T1D
UC
Other
rs7926914-LRRC32 (LRRC32)
rs34185821-RBPJ (SMIM20)
rs10892258-CXCR5 (TREH)
rs7664452-IL2 (KIAA1109)
rs10186048-CTLA4 (CD28)
rs1031163-GATA3 (GATA3)
rs10178845-ID2 (LINC00299)
TREG / TH17 Fold-Change (log2)
d
All Autoimmune Diseases
Multiple Sclerosis
Crohn's Disease
Type I Diabetes
317 367 2230
54 78 422
31 47 144
28 44 170
Nearest Gene
HiChIP Gene Targets
e
p < 2.2 x 10-16
Norm EIS (log2)
T eQTL T Shuffle L eQTL
Naive
TH17
Treg
f
GWAS
PBC
T1D
Psoriasis
MS
rs8052717
RA
T1D
rs34185821
Allergy
rs367023
Naive
TH17
Treg
Norm EIS
-500 kb
chr. 16: 11.445 Mb
+100 kb
-200 kb
chr4: 26.08 Mb
+1 Mb
-100 kb
chr2: 8.45 Mb
+450 kb
ATAC
Loops
refseq genes
CIITA >
CLEC16A >
< SOCS1
RMI2 >
SMIM20 >
RBPJ >
TBC1D19 >
STIM2 >
< LINC00299
ID2 >
My-La CRISPRi E-P Validation
SNP-e
p
KRAB
dCas9
rs8052717 Enhancer
CLEC16A Promoter
Negative Control
CLEC16A / GAPDH
rs34185821 Enhancer
RBPJ Promoter
Negative Control
RBPJ / GAPDH
rs367023 Enhancer
ID2 Promoter
Negative Control
ID2 / GAPDH

# Mumbach et al. (CHANG), Figure 5



**Mumbach et al. (CHANG), Figure 6**