Dynamic chromatin architecture identifies new autoimmune-associated enhancers for *IL2* and novel genes regulating CD4+ T cell activation

Matthew C. Pahl*[1,2], Prabhat Sharma*[1,3], Rajan M. Thomas*[1,3], Zachary Thompson[1,3], Zachary Mount[1,3], James Pippin[1,2], Peter A. Morawski[4], Peng Sun[1,3], Chun Su[1,2], Daniel J. Campbell[4,5], Struan F.A. Grant[1,2,6,7,8], Andrew D. Wells[1,3,9,10]

[1]Center for Spatial and Functional Genomics, The Children's Hospital of Philadelphia
[2]Division of Human Genetics, The Children's Hospital of Philadelphia
[3]Department of Pathology, The Children's Hospital of Philadelphia
[4]Benaroya Research Institute at Virginia Mason, Seattle, WA
[5]Department of Immunology, University of Washington School of Medicine, Seattle, WA
[6]Department of Genetics, Perelman School of Medicine, University of Pennsylvania
[7]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania
[8]Division of Endocrinology and Diabetes, The Children's Hospital of Philadelphia
[9]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania
[10]Institute for Immunology, Perelman School of Medicine, University of Pennsylvania

*Co-lead authors

Correspondence:
Andrew D. Wells, PhD
The Children's Hospital of Philadelphia Research Institute
916D Abramson Research Center
3615 Civic Center Boulevard
Philadelphia, PA 19104
Email: wellsa@chop.edu
Phone: 215-590-8710

ABSTRACT

Genome-wide association studies (GWAS) have identified hundreds of genetic signals associated with autoimmune disease. The majority of these signals are located in non-coding regions and likely impact *cis*-regulatory elements (cRE). Because cRE function is dynamic across cell types and states, profiling the epigenetic status of cRE across physiological processes is necessary to characterize the molecular mechanisms by which autoimmune variants contribute to disease risk. We localized risk variants from 15 autoimmune GWAS to cRE active during TCR-CD28 costimulation of naïve human CD4+ T cells. To characterize how dynamic changes in gene expression correlate with cRE activity, we measured transcript levels, chromatin accessibility, and promoter-cRE contacts across three phases of naive CD4+ T cell activation using RNA-seq, ATAC-seq, and HiC. We identified ~1,200 protein-coding genes physically connected to accessible disease-associated variants at 423 GWAS signals, at least one-third of which are dynamically regulated by activation. From these maps, we functionally validated a novel stretch of evolutionarily conserved intergenic enhancers whose activity is required for activation-induced *IL2* gene expression in human and mouse, and is influenced by autoimmune-associated genetic variation. The set of genes implicated by this approach are enriched for genes controlling CD4+ T cell function and genes involved in human inborn errors of immunity, and we pharmacologically validated eight implicated genes as novel regulators of T cell activation. These studies directly show how autoimmune variants and the genes they regulate influence processes involved in CD4+ T cell proliferation and activation.

INTRODUCTION

GWAS has linked hundreds of regions of the human genome to autoimmune disease susceptibility. The majority of GWAS variants are located in non-coding regions of the genome, and likely contribute to disease risk by modulating cis-regulatory element (cRE) activity to influence gene expression[1]. Identifying causal variants and effector genes molecularly responsible for increased disease risk is critical for identifying targets for downstream molecular study and therapeutic intervention. An understanding of how non-coding variants function is often limited by incomplete knowledge of the mechanism of action, *i.e.*, whether a variant is located in a cRE, in which cell types a cRE is active, and which genes are regulated by which cRE.

CD4+ T cells are key regulators of innate and adaptive immune responses that combat infection by orchestrating the activity of other immune cells. In their quiescent, naïve state, 'helper' T cells traffic between the blood and secondary lymphoid tissues as part of an immunosurveillance program maintained by transcription factors such as KLF2, TOB, and FOXO[2–4]. Upon encountering specific antigen, a cascade of signals activated through the TCR and costimulatory receptors results in large-scale changes in gene expression driven by factors like NFKB, NFAT, IRF4, and STATs, leading to proliferation and differentiation into specialized Th1, Th2, and Th17 subsets capable of participating in protective immunity[5,6]. CD4+ T cells are also key players in the induction and pathogenesis of autoimmunity. The cis-regulatory architecture of CD4+ T cells is enriched for autoimmune disease GWAS variants[7–14], and T cells from autoimmune patients harbor distinct epigenetic and transcriptomic signatures linking dysregulated gene expression with disease pathogenesis. Autoimmune variants may contribute to the breakdown of immune self-tolerance by shifting the balance between autoreactive conventional vs. regulatory CD4+ T cells, by altering cytokine production, or by promoting auto-antigen production[15,16].

Chromatin conformation assays allow for identification of putative target genes of autoimmune variant-containing cRE in close spatial proximity to gene promoters. Previous work using promoter-capture HiC and HiC in combination with other epigenetic marks have implicated sets of autoimmune variants and effector genes that may participate in T cell activation[17–19]. In addition to chromatin conformation approaches, multiple complementary approaches have been developed to link disease associated SNPs to their downstream effector genes. Expression quantitative trait mapping (eQTL) and chromatin co-accessibility data can be used to implicate effector genes through statistical association of genotypes with readouts of expression or other molecular markers[20]. However, these approaches do not account for the relevant 3D structure of the genome in the nucleus, and are highly susceptible to *trans*-effects and other confounding factors.

In this study, we measured the impact of TCR-CD28 activation on the autoimmune-associated *cis*-regulatory architecture of CD4+ helper T cells, and by comparing our data to those of several orthogonal GWAS effector gene nomination studies, identify hundreds of effector genes not implicated previously. We find that 3D chromatin-based approaches exhibit 2- to 10-fold higher predictive sensitivity than eQTL and ABC statistical approaches when benchmarked against a ‘gold standard set’ of genes underlying inborn errors in immunity and tolerance. Our maps of autoimmune SNP-gene contacts also predicted a stretch of evolutionarily conserved, intergenic enhancers that we show are required for normal expression of the canonical T cell activation gene *IL2* in both human and mouse, whose activity is influenced by autoimmune risk variants. The set of variant-connected effector genes defined by 3D physical proximity to autoimmune-associated cRE is enriched for genes that regulate T cell activation, as validated pharmacologically in this study and by CRISPR-based screens in orthogonal studies[21–23].

RESULTS

Gene expression dynamics as a function of naïve CD4+ T cell activation

To explore the universe of genes and cREs that are affected by and may contribute to CD4+ activation, we characterized the dynamics of gene expression, chromatin accessibility, and 3D chromatin conformation in human CD4+CD62L^{hi}CD44^{lo} naïve T cells purified directly *ex vivo* and in response to *in vitro* activation through the T cell receptor (TCR) and CD28 for 8 or 24 hours using RNA-seq (N=3 donors), ATAC-seq (N=3 donors), and HiC (N=2 donors, **Figure 1A**). As sequencing depth and restriction enzyme cutting frequency affects the resolution and reliability of HiC experiments[24], we constructed libraries with two four-cutter restriction enzymes and sequenced to a total of approximately 4 billion unique-valid reads per timepoint. We verified the reproducibility of replicate ATAC-seq and RNA-seq libraries with principal component analysis and HiC libraries using distance-controlled, stratum-adjusted correlation coefficient (SCC). Replicate samples were highly correlated and clustered by activation stage (**Figure 1 - Figure Supplement 1A-C**). As expected, quiescent naïve CD4+ T cells expressed high levels of *SELL, TCF7, CCR7,* and *IL7R*, and rapidly upregulated *CD69, CD44, HLADR,* and *IL2RA* upon stimulation (**Figure 1B, Supplementary File 1**). Genome-scale gene set variance analysis based on MsigDB hallmark pathways showed that genes involved in KRAS and Hedgehog signaling are actively enriched in quiescent naive CD4+ T cells, while stimulated cell transcriptomes are enriched for genes involved in cell cycle, metabolism, and TNF-, IL-2/STAT5-, IFNg-, Notch-, and MTORC1-mediated signaling pathways (**Supplementary File 2**).

To define global gene expression dynamics during the course of CD4+ T cell activation, we performed pairwise comparisons and k-means clustering, identifying 4390 differentially expressed genes after 8 hours of stimulation (3289 upregulated and 1101 downregulated) and 3611 differentially expressed genes between 8 hours and 24 hours (3015 upregulated and 596 downregulated, **Figure 1 - Figure Supplement 1D**) that could be further separated into five clusters based on their distinct trajectories (**Figure 1C**, **Figure 1 - Figure Supplement 1E-G**,

Supplementary File 1). Genes upregulated early upon activation (cluster 1; n=1621 genes) are enriched for pathways involved in the unfolded protein response, cytokine signaling, and translation (*e.g.*, *IL2, IFNG, TNF, IL3, IL8, IL2RA, ICOS, CD40LG, FASLG, MYC, FOS, JUNB, REL, NFKB1, STAT5A,* **Supplementary File 3**). Genes downregulated late (cluster 2; n=1600 genes) are moderately enriched for receptor tyrosine kinase signaling, cytokine signaling, and extracellular matrix organization. Genes monotonically increasing (cluster 3; n=2676 genes) are highly enriched for pathways involved in infectious disease and RNA stability, translation and metabolism, and moderately enriched for pathways involved in the unfolded protein response, cellular responses to stress, regulation of apoptosis, and DNA repair (*e.g.*, *TBX21, BHLHE40, IL12RB2, STAT1, CCND2, CDK4, PRMT1, ICAM1, EZH2,* **Supplementary File 3**). Genes downregulated early (cluster 4; n=1628 genes) are enriched for pathways involved in inositol phosphate biosynthesis, neutrophil degranulation, and metabolism of nucleotides (*e.g.*, *KLF2, IL7R, RORA, IL10RA*), and genes upregulated late (cluster 5; n=2154 genes) are highly enriched for pathways involved in cell cycle, DNA unwinding, DNA repair, chromosomal maintenance, beta-oxidation of octanoyl-CoA, and cellular response to stress (*e.g.*, *CDK2, E2F1, CDK1, CCNE1, CCNA2, PCNA, WEE1, CDC6, ORC1, MCM2*). These patterns are consistent with known changes in the cellular processes that operate during T cell activation, confirming that *in vitro* stimulation of CD4+ T cells recapitulates gene expression programs known to be engaged during a T cell immune response.

Dynamic changes in chromosomal architecture and genome accessibility during naïve CD4+ T cell activation

To understand the *cis*-regulatory dynamics underlying the observed activation-induced changes in gene expression, we examined CD4+ T cell nuclear chromosome conformation and chromatin accessibility as a function of stimulation state using HiC and ATAC-seq. The human genome consists of ~3 meters of DNA that is incorporated into chromatin and compacted into

the ~500 cubic micron nucleus of a cell in a hierarchically ordered manner. This degree of compaction results in only ~1% of genomic DNA being accessible to the machinery that regulates gene transcription[25], therefore a map of open chromatin regions (OCR) in a cell represents its potential gene regulatory landscape. Open chromatin mapping of human CD4+ T cells at all states identified a consensus *cis*-regulatory landscape of 181,093 reproducible OCR (**Supplementary File 4**). Of these, 14% (25,291) exhibited differential accessibility following 8 hours of stimulation (FDR<0.05). Most differentially accessible regions (DAR) became more open (18,887), but some DAR (6,629) showed reduced accessibility (**Figure 2A**, **Supplementary File 5**). The change in accessibility over the next 16 hours of stimulation showed the opposite dynamic, with 6,629 regions exhibiting reduced accessibility, and only 4,417 DAR becoming more open (total of 11,046 DAR, **Figure 2A**, **Supplementary File 5**). These OCR represent the set of putative cRE with dynamic activity during T cell activation.

The vast majority of putative cRE are located in intergenic or intronic regions of the genome far from gene promoters, meaning that the specific impact of a given cRE on gene expression cannot be properly interpreted from a one-dimensional map of genomic or epigenomic features. To predict which cRE may regulate which genes in CD4+ T cells across different states of activation, we created three-dimensional maps of cRE-gene proximity in the context of genome structure. The highest order of 3D nuclear genome structure is represented by A/B compartments, which are large chromosomal domains that self-associate into transcriptionally active (A) *vs*. inactive (B) regions[26]. In agreement with prior studies, we find that OCR located in active A compartments exhibit higher average accessibility than those OCR located in less active B compartments (**Figure 2 – Figure Supplement 1A, Supplementary File 6**), a trend observed across all cell states. Likewise, genes located in A compartments show higher average expression than those located in B compartments (**Figure 2 – Figure Supplement 1B**). A quantitative comparison across cell states showed that 94% of the CD4+ T cell genome remained stably compartmentalized into A (42%) and B (52%), indicating that activation does

not cause a major shift in the large-scale organization of the genome within the nucleus (**Figure 2B**).

Within each A or B compartment, the genome is further organized into topologically associating domains (TADs)[27]. These structures are defined by the fact that genomic regions within them have the potential to interact with each other in 3D, but have low potential to interact with regions outside the TAD. The location of TAD boundaries can influence gene expression by limiting the access of cRE to specific, topologically associated genes. While ~80% of TAD boundaries remained stable across all states, 20% of TAD boundaries (8925) changed as a function of T cell activation (**Figure 2C**, **Figure 2 – Figure Supplement 1C-D**, **Supplemental File 7**). TAD boundary dynamics were categorized and 2198 boundaries exhibited a change in strength, 2030 boundaries shifted position, and 4697 boundaries exhibited more complex changes such as loss of a boundary resulting in merger of two neighboring TAD, addition of a boundary splitting one TAD into two, or a combination of any of these changes. Genes nearby dynamic TAD boundaries are enriched for pathways involved in RNA metabolism, cellular response to stress, and the activity of PTEN, p53, JAK/STAT, Runx and Hedgehog (**Supplementary File 6**). We also detected chromatin stripes, which are TAD-like structures that consist of a genomic anchor region that is brought into contact with a larger loop domain via an active extrusion mechanism. Chromatin stripes are contained within and/or overlap TAD regions, and are enriched for active enhancers and super-enhancers[28,29]. We identified 1526 chromatin stripes in quiescent naïve CD4+ T cells, 1676 stripes in 8 hour stimulated cells, and 2028 stripes at 24 hours post-stimulation (**Figure 2 - Figure Supplement 2A, Supplementary File 8**). Consistent with prior studies in other cell types, chromatin stripes were preferentially located in A compartments, and genes and OCR within stripe regions showed increased expression and chromatin accessibility (**Figure 2 - Figure Supplement 2B-D**).

Within active topological structures, transcriptional enhancers can regulate the expression of distant genes by looping to physically interact with gene promoters[30,31]. To identify potential

regulatory cRE-gene interactions, we identified high confidence loop contacts across all cell states using Fit-HiC (merged 1kb, 2kb, and 4kb resolutions, **Figure 2D**). This approach detected 933,755 loop contacts in quiescent naïve CD4+ T cells, 900,267 loop contacts in 8-hour stimulated cells, and 551,802 contacts in 24-hour stimulated cells (2,099,627 total unique loops). Approximately 23% of these loops involved a gene promoter at one end and an OCR at the other, and these promoter-interacting OCR were enriched for enhancer signatures based on flanking histone marks from CD4+ T cells in the epigenome roadmap database[32] (**Figure 2 - Figure Supplement 1E**). T cell activation resulted in significant reorganization of the open chromatin-promoter interactome, as 907 promoter-OCR exhibited increased contact and 1333 showed decreased contact following 8 hours of stimulation (**Figure 2E**). Continued stimulation over the next 16 hours was associated with an increase in the contact frequency of 41 promoter-OCR pairs, while only 4 pairs exhibited decreased contact (**Figure 2E**). Activation-induced changes in chromatin architecture and gene expression were highly correlated, as genomic regions exhibiting increased promoter connectivity became more accessible at early stages of stimulation, which was associated with increased gene transcription from connected promoters (**Figure 2F, Supplementary File 9**). The accessibility of promoter-connected OCR and the expression of their associated genes decreased globally from 8 to 24 hours of stimulation (**Figure 2F**). We compared these loop calls to a prior chromatin capture analysis in CD4+ T cells by Burren et al.[18] and found that roughly 40% of stable loops in both stimulated and unstimulated cells were identical in both studies, despite differing in approach (HiC vs. PCHiC), analysis (HiC vs. CHiCAGO), sample (naïve CD4+ vs. total CD4+), timepoint (8 vs. 4 hour), and donor individuals (**Figure 2 - Figure Supplement 1F**). As expected, unstimulated samples were more similar than activated samples.

We next focused on the 5 sets of genes with the dynamic expression patterns defined in **Figure 1D**, and identified 57,609 OCR that contact dynamic gene promoters in at least one stage. Most dynamic genes contacted between 1 and ~35 OCR, with a median of 10 OCR per

gene, but a handful of dynamic genes were observed in contact with over 100 distinct open chromatin regions (**Figure 2 - Figure Supplement 1G**). Similarly, most OCR were connected to a single dynamic gene, but many contacted more than one gene (median 2 genes per OCR), suggesting that most dynamic genes have a complex regulatory architecture. Increased gene expression upon activation correlated with an increase in the accessibility and promoter contact frequency of distant cRE (**Figure 2 - Figure Supplement 1H**), as exemplified by *GEM* and *IRF4* (**Figure 2 - Figure Supplement 3A**,**B**). Conversely, the 3D regulatory architecture of genes like KLF2 and DPEP2, which were downregulated following stimulation, exhibited decreased contact and accessibility (**Figure 2 - Figure Supplement 3C**,**D**).

Transcription factor footprints enriched in dynamic open chromatin identify regulators of T cell activation

To explore what factors drive dynamic changes in the regulatory architecture of the CD4+ T cell genome during activation, we conducted a quantitative footprinting analysis at 1,173,159 transcription factor (TF) motifs located in the consensus CD4+ T cell open chromatin landscape using the average accessibility of the region surrounding each motif as a measure of regulatory activity (**Supplementary File 10**). Activation of naïve CD4+ T cells resulted in increased chromatin accessibility around bZIP motifs at tens of thousands of genomic regions by 8 hours post-stimulation (**Figure 3A** and **C**), which was associated with increased expression of Fos and Jun family members constituting the AP-1 complex, as well as the bZIP factors BATF, BACH1, and BACH2 (**Figure 3A**). The activity of NFE2 and SMAD was increased without increased expression (**Figure 3A** and **C**), likely due to post-translational regulation of these factors by phosphorylation[33]. Conversely, the motifs for a number of TF exhibited significantly reduced accessibility early after stimulation, including those for EGR2 and FOXP3 that are known to negatively regulate T cell activation[34,35] (**Figure 3A**). By 24 hours post-activation, bZIP activity remained largely unchanged compared to 8 hours (**Figure 3B**), but a number of factors showed

decreased activity. These include several members of the Sp family, the Myc cofactor MAZ that also cooperates with CTCF to regulate chromatin looping[36], KLF2, which controls genes involved in naïve CD4+ T cell quiescence and homing[2,37], NRF1, a factor implicated in age-associated T cell hypofunction[38], and EGR2 and 3, which are known to oppose T cell activation and promote tolerance[34,39,40] (**Figure 3B** and **C**).

To explore how transcription factor activity may operate *via* the CD4+ T cell open chromatin landscape to regulate distinct programs of dynamic gene expression during TCR/CD28-induced activation, we focused on TF motifs enriched among those OCR specifically contacting promoters of dynamic genes identified by our clustering analysis (**Figure 3 – Figure Supplement 1A**). The set of OCR contacting dynamic gene promoters were enriched for the motifs of 89 expressed (TPM>5) transcription factors, as compared to motifs present in the total open chromatin landscape (**Figure 3D**). The majority of this TF activity was enriched in OCR connected to genes highly upregulated at 24 hours post-activation (clusters 3 and 5, **Figure 3D**), with the exception of CREB3, ELF1, ESRRA, GABPA, RELA, XPB1, ZNF384, and the transcriptional repressor IKZF1 known as a strong negative regulator of T cell activation and differentiation[41–44]. Conversely, motifs for IKZF1, ZNF384, GABPA, ESRRA, and ELF1 were highly enriched in the set of OCR contacting genes down-regulated early after activation (cluster 4, **Figure 3D**). Motifs for KLF2 and the metabolic gene regulator SREBF1 were likewise enriched in OCR connected to down-regulated genes. Open chromatin regions interacting with genes in cluster 2 are negatively enriched for this set of TF except for CTCF (**Figure 3D**).

Finally, we integrated TF footprint, promoter connectome, and gene co-expression data to construct TF-gene regulatory networks likely operating at each timepoint. The connections between regulatory nodes are based on physical promoter-TF footprint interactions with confidence weighted by gene co-expression (**Figure 3E**, **Supplementary File 11**). Highly co-expressed genes at the core of the unstimulated CD4+ T cell regulatory network encode transcription factors such as KLF2, ETS1, IKZF1, and TCF7[45–47] that are known to be involved

in T cell gene silencing, quiescence, homeostasis, and homing. Genes connected to the top factors in this network were enriched for pathways involved in immune signaling, DNA replication and repair, protein secretion, and programmed cell death (**Figure 3 – Figure Supplement 1B**). Costimulation through the TCR and CD28 induced a set of core network genes active at both time points with known roles in T cell activation and differentiation (*NFKB1, JUNB, MYC, IRF4, STAT5, STAT1, LEF1, ATF4*). Also part of this set is *PLAGL2*, an oncogene in the Wnt pathway that regulates hypoxia-induced genes[48] with no prior defined role in T cell activation. Additional nodes specifically implicated at 8 hours post-activation are HIF1A, the major sensor of cellular hypoxia[49], and XBP1, a major transcriptional mediator of the unfolded ER protein response with defined roles in T cell activation, differentiation, and exhaustion[50,51]. Factors specifically implicated at 24 hours post-activation include E2F1, a transcriptional regulator of both cell cycle and apoptosis in T cells[52,53], BHLHE40, a factor known to control multiple aspects of T cell metabolism and differentiation[54], and the Myc cofactor MAZ that has not been previously studied in the context of T cell function. Genes connected to factors in the activated T cell networks were enriched for pathways involved in cytokine signaling, the interferon response, transcription, cell cycle, DNA replication and repair, and programmed cell death (**Figure 3 – Figure Supplement 1B**). Together, these data indicate that concurrent but separable stimulation-induced gene programs are the result of the activity of distinct sets of DNA binding factors mobilized by antigen and costimulatory receptor signaling in naïve CD4+ T cells.

Identification of autoimmune variants associated with CD4+ T cell cRE and predicted effector genes

Following our established variant-to-gene (V2G) mapping approach to implicate functional SNPs and their effector genes using the combination of GWAS and chromatin conformation capture data[55–59], we intersected promoter-interacting OCR with autoimmune SNPs from the

95% credible set derived from 15 autoimmune diseases (**Figure 4A**, **Supplementary Files 12** and **13**). Constraining the GWAS SNPs in this way reduced the credible set size from an average of 14 variants per sentinel to 3 variants per sentinel. To determine whether open chromatin in physical contact with dynamically regulated genes in CD4+ T cells is enriched for autoimmune disease heritability, we performed a partitioned LD score regression analysis. This landscape was enriched for variants associated with susceptibility to inflammatory bowel disease (IBD), ulcerative colitis (UC), type I diabetes (T1D), lupus (SLE), celiac disease (CEL), allergy (ALG), eczema (ECZ), and rheumatoid arthritis (RA), but not for variants associated with psoriasis (PSO) or juvenile idiopathic arthritis (JIA) (**Figure 4B**, **Supplementary File 14**). The OCR connected to genes upregulated early and/or progressively upon activation (clusters 1 and 3) were most strongly enriched for ALG, CEL, IBD/UC, RA and T1D heritability, while SLE and ECZ heritability was most enriched in OCR connected to genes upregulated later post-activation (clusters 3 and 5, **Figure 4B**). SLE was also the only disease (besides PSO and JIA) that was not enriched in open chromatin connected to down-regulated genes (cluster 4, **Figure 4B**).

The promoter-connected open chromatin landscape for all CD4+ T cell states in this study contains 2606 putatively causal variants linked to ~half of the sentinel signals (423) for the 15 diseases analyzed, and are in contact with a total of 1836 genes (**Supplementary File 13**). A total of 1151 autoimmune variants localized to the promoters of 400 genes (-1500/+500bp from the TSS, **Supplementary File 15**). These variants were on average 103 kb from the TSS of their connected gene (**Figure 4C**), and each variant contacted an average of 5 genes (**Figure 4D**). The majority of linked SNPs interact with genes in addition to the nearest gene, and ~half of linked SNPs 'skip' the nearest gene to target only distant genes (**Figure 4E**). Approximately 60% of connected genes were implicated across all timepoints (**Figure 4F**, **Supplementary File 13**), while ~40% (753) are dynamically regulated (clusters 1-5) in response to TCR/CD28 co-stimulation. Examples of SNP-genes pairs that exhibit dynamic accessibility, chromosome contact, and expression in response to T cell activation are the *SIK1*, *PARK7*, *DUSP5*,

CLEC2D, *TRIP10*, *GPR108* and *IL2* loci (**Figure 4G**, **Supplementary File 13**). The *TRIP10* and *GPR108* promoters were each captured in contact with a high confidence variant rs1077667 (PP>0.99), which is located in an intron of *TNFSF14* and is associated with multiple sclerosis (**Figure 4H**). The accessibility of this SNP and it's contact with *TRIP10* and *GPR108* increased following activation (**Figure 4H**). Conversely, the allergy-associated SNP rs7380290 is accessible and contacts the *SIK1* promoter in resting cells, but shows reduced accessibility and promoter connectivity upon activation (**Figure 4I**). *TRIP10* encodes a cytoskeletal binding protein involved in endocytosis that suppresses glucose uptake in response to insulin signaling[60], *GPR108* encodes an orphan G-protein coupled receptor, and *SIK1* encodes a salt-inducible kinase with roles in cancer, epilepsy, and myeloid signaling[61]. None of these genes have been previously studied in the context of T cell activation.

Comparative predictive power against orthogonal variant-to-gene mapping approaches

In an attempt to gauge the predictive power of our approach relative to other V2G approaches, we compared our chromatin capture-based autoimmune GWAS effector gene predictions to the predictions of four other chromosome capture-based studies in human CD4+ T cells[17–19,62], four single-cell eQTL studies in human CD4+ T cells[13,62–64], and a set of 449 genes that when mutated in humans cause inborn errors of immunity[65]. As part of the comparison, we harmonized the chromatin-based datasets using our chromatin loop calling approach and integrated with the same 95% credible set of autoimmune variants to create comparable physical contact maps. A total of 6936 unique genes were implicated by all studies. Our HiC-ATAC-seq approach in naïve and activated human CD4+ implicated 1947 effector genes, 400 of which through proxies in open promoters and 1836 through contacts with distal variants, of which 752 were dynamically regulated throughout activation. The HiC dataset from Gate et al. implicated 110 genes, and the capture-HiC datasets from Burren et al., Yang et al., and Javierre et al. implicated 1408, 2572, and 2368 genes, respectively (**Supplementary File**

16). A total of 369 autoimmune GWAS effector genes were predicted in common by our HiC approach and the pcHiC-based approaches. The eQTL studies by Soskic et al., Gate et al., Ye et al., and Schmiedel et al. (DICE) implicated 171, 15, 15, and 221 genes, and the co-accessibility data from Gate et al. implicated 45 genes. Concordance between our gene predictions and the pcHiC predictions ranged from 15-24% (**Figure 5A**), while concordance between our predictions and the eQTL predictions was low (0.6%~5%, **Figure 5A**), and concordance among eQTL studies was also low (1.6%-13%, **Figure 5A**). However, when we co-localized the Soskic variant-gene pairs with autoimmune GWAS SNPs, 41 of these eQTL were also predicted by our study, representing a concordance 10-fold higher than obtained by a random sampling of genes within 500 kb of autoimmune variants (**Figure 5B**). A total of 1519 of the genes implicated in our study were not predicted by eQTL approaches, while 644 genes from our CD4 T cell V2G data were not implicated by the other chromatin-based datasets, and altogether 562 putative autoimmune effector genes were uniquely predicted by our study. Potential sources of variation between the results of the studies are outlined in the Discussion. A total of 17 genes were implicated by all CD4+ T cell-based approaches that nominated at least 150 genes (APIP, BCL2L11, ERAP1, ERAP2, CD5, FIGNL1, IL18R1, METTL21B, MRPL51, PPIF, PTGER4, PXK, RMI2, RNASET2, SERPINB1, TAPBPL, VAMP1).

To measure the predictive power of the sets of genes implicated by each approach above, a precision-recall analysis was performed against the ‘gold standard’ human inborn errors in immunity gene set, based on the hypothesis that genes under the control of autoimmune-associated variants would cause monogenic immune disease phenotypes if subjected to deleterious coding mutations. The set of genes from our study with disease-associated SNPs located in their open promoters overlapped with only 5% of the genes from the gold standard set (**Supplementary File 16**, **Figure 5C**). However, leveraging our HiC data to include distal autoimmune variants captured interacting with open gene promoters increased the sensitivity of the approach 3-fold to 16% (71 overlapping genes, **Figure 5C**), confirming the relevance of

long-range promoter-cRE/SNP contacts. This level of sensitivity is comparable to that of other chromosome capture-based datasets for predicting HIEI genes in CD4+ T cells (**Figure 5C**). Conversely, the eQTL-based approaches were 3- to ~12-fold less sensitive than the chromatin-constrained approaches in predicting HIEI genes, with most recalling 0-2.5% and the DICE eQTL dataset recalling 6% (**Figure 5C**). Restricting our 3D chromatin V2G to only those genes dynamically regulated during T cell activation reduced predictive power from 16% to 10%, but the fraction of these genes in the HIEI gold standard set (precision) increased from ~4% to 6.2% (**Figure 5C**), indicating that focusing the V2G set on those SNP-gene pairs that are dynamically regulated by T cell activation reduces potential false positive predictions. The precision of our dataset was superior to that of the other chromosome capture-based datasets except for the Burren pcHiC dataset (**Figure 5C**). The eQTL approaches were comparable to the 3D chromatin-based datasets in precision (~3-6%, **Figure 5C**).

We also analyzed the relative predictive power of a recently proposed 'activity-by-contact' (ABC) model[66] that uses a multi-tissue averaged HiC dataset instead of cell type-matched 3D chromatin-based data to link variants to genes. This model constrains the input data by removing bidirectional, antisense, and small RNAs, and eliminate genes ubiquitously expressed across tissues based on the assumption that these genes are driven primarily by elements adjacent to their promoters. Also, while most human genes have multiple alternative transcriptional start sites (median = 3), ABC only annotates only one promoter per gene. To apply the activity-by-contact gene nomination model to our CD4+ T cell chromatin-based V2G data, we used our ATAC-seq data and publicly available CD4+ T cell H3K27ac ChIP-seq data as input, and integrated this with GWAS and the averaged HiC dataset from the original ABC study[66,67]. The ABC model nominated 650 genes compared to 1836 genes using our cell type-matched HiC data and analysis pipeline. Only 357 genes were nominated by both approaches, while 1479 genes nominated by our approach were not implicated by the ABC model, and 293 genes not implicated by our approach were newly implicated by ABC. To measure the predictive

power of the ABC approach, we re-ran the precision-recall re-analysis with all datasets subjected to the ABC gene-promoter filter (**Fig. 5D**). We found that applying the restricted ABC model promoter annotation to all datasets did not have a large effect on recall, however, the precision of several of the datasets were affected. For example, using the restricted promoter/gene set reduced the precision of our V2G approach and artificially inflated the precision of the 'nearest gene to SNP' metric. The precision-recall analysis also shows that the ABC score-based approach is only half as powerful at predicting HIEI genes as the chromatin-based V2G approaches (**Fig. 5D**). This indicates that informing GWAS data with cell type- and state-specific 3D chromatin-based data brings more target gene predictive power than application of the multi-tissue-averaged HiC used by the ABC model. Together, these analyses indicate that chromosome capture-based V2G is a more sensitive approach for identifying 'true' effector genes than eQTL or ABC approaches, but comes with additional predicted genes that represent either false positives or true effector genes for which monogenic LOF/GOF mutations have not yet been characterized in humans.

Functional validation of novel autoimmune V2G-predicted enhancers at the *IL2* locus

Our 3D chromatin-based analysis specifically predicts dynamic, disease-associated regulatory elements in intergenic space at the *IL2* locus. The *IL2* gene encodes a cytokine with crucial, pleotropic roles in the immune system, and dysregulation of IL-2 and IL-2 receptor signaling leads to immunodeficiency and autoimmune disorders in mice and humans[68–71]. Activation-induced transcription of *IL2* involves an upstream regulatory region (URR) ~375 bp from the TSS that has served as a paradigm of tissue-specific, inducible gene transcription for nearly four decades[72–74]. However, the presence of evolutionarily conserved non-coding sequences (CNS) in the ~150 kb of intergenic space 46, 51, 80, 83, 85, 96, 122, and 128 kb upstream of the TSS suggest that additional regulatory elements may have evolved to control *IL2*[75] (**Figure 6A**). The -51 kb CNS contains a SNP linked to T1D, IBD, PSO, CEL and allergy

(rs72669153), while the -85 kb CNS contains a SNP linked to RA (rs6232750) and the -128 kb CNS contains two SNPs linked to T1D, JIA, and SLE (rs1512973 and rs12504008, **Figure 6A**). In TCR/CD28-stimulated naïve CD4+ T cells, these CNS are remodeled to become highly accessible (**Figure 6A**), and they loop to interact physically with the *IL2* URR (**Figure 6B**) at both time points. ChIP-seq analyses in human T cells[76] show that the URR and all distal CNS except -85 are occupied by TF such as Jun/Fos (AP-1), NFAT, and NFkB that are known regulators of *IL2* (**Figure 6 – Figure Supplement 1**), and the -85, -122, and -128 CNS are occupied by additional TF not previously thought to be direct regulators of *IL2*, such as MYC, BCL6, and STAT5 (**Figure 6 – Figure Supplement 1**). Recombinant reporter assays in primary activated human CD4+ T cells showed that the -46, -51, -83, and -128 CNS/OCR can enhance transcription from the URR (**Figure 6C**). To determine whether the native elements contribute to the expression of *IL2*, we targeted each CNS/OCR individually in primary human CD4+ T cells or Jurkat T cells using CRISPR/CAS9 (**Figure 6D**) and measured IL-2 secretion following TCR/CD28 stimulation (**Figure 6E**). Deletion of the -46, -51, -83, -85, -122, and -128 kb elements in primary human CD4+ T cells each resulted in a ~50% reduction in IL-2 production, while deletion of the -80 kb element had little effect. A very similar pattern of impact was observed when these elements were deleted individually in Jurkat T cells (**Figure 6E**). The URR has a stronger contribution to IL-2 production than any individual intergenic element, as deletion of the URR almost completely abrogated activation-induced IL-2 production by both primary CD4+ or Jurkat T cells (**Figure 6E**). To determine whether these intergenic enhancers exist in synergistic epistasis[77] necessary for *IL2* transactivation, we generated Jurkat T cell clones in which the stretch of all 7 elements located 46 kb to 128 kb upstream of *IL2* was deleted using CRISPR/CAS9 (**Figure 6F**). Despite the URR and 46 kb of upstream sequence being intact in these clones, loss of the 81.3 kb stretch of intergenic enhancers renders these cells incapable of expressing *IL2* at both the mRNA and protein level in response to stimulation (**Figure 6G**). These results show that the URR is not sufficient for activation-induced expression of *IL2*, and

that *IL2* has a previously unappreciated, complex, and autoimmune disease-associated regulatory architecture that was accurately predicted by our 3D epigenomic V2G approach. Importantly, we find that the distal *IL2* cRE are highly accessible in quiescent memory T cell subsets (Th1, Th2, Th17, Th1-17, Th22) isolated directly *ex vivo* from human blood, whereas naïve CD4+ T cells and non-IL-2-producing Treg showed little accessibility at these elements (**Figure 6 – Figure Supplement 1D**). This suggests that stable remodeling of distal *IL2* cRE can persist *in vivo* after TCR signals cease, and that this epigenetic imprinting contributes to the immediate, activation-induced production of IL-2 exhibited by memory, but not naïve or regulatory, CD4+ T cells[78–80].

Distal *IL2* enhancers are evolutionarily conserved and impact *in vivo* T cell-mediated immunity in mice

The intergenic *IL2* enhancers defined above are conserved at the sequence level between human and mouse (**Figure 7A**). To test whether enhancer function is likewise evolutionarily conserved, we used zygotic CRISPR/CAS9 targeting to generate mice with a 583 bp deletion of the ortholog of the -128 kb human enhancer CNS/OCR situated 83 kb upstream from mouse *Il2* (*Il2*-83-cRE-ko, C57BL6 background). Deletion of this genomic region did not discernably affect T lymphocyte development in *Il2*-83-cRE-ko mice. However, peripheral $CD4^+$ T cells from *Il2*-83-cRE-ko mice produced substantially less IL-2 at both the protein and mRNA level, and exhibited reduced proliferation, in response to *in vitro* stimulation (**Figure 7B**), confirming that the enhancer function of the orthologous -128 and -83 kb elements is conserved between human and mouse. The *in vitro* induction of Foxp3+ regulatory T cells (Treg) from conventional CD4+ T cell precursors in response to antigenic stimulation in the present of TGF-beta, a differentiation process highly dependent upon IL-2, was reduced 4-fold in *Il2*-83-cRE-ko mice compared to wild-type mice (**Figure 7C**), but could be rescued by addition of exogenous IL-2 to the culture (**Figure 7C**).

To test whether the distal -83 kb *Il2* enhancer contributes to physiologic immune responses *in vivo*, we immunized wild-type and *Il2*-83-cRE-ko mice with the model antigen chicken ovalbumin. *Il2*-83-cRE-ko animals showed a nominal increase in the differentiation of CD4+ T cells into CXCR5+PD-1^{hi} follicular helper T cells (Tfh) in the spleen compared to wild-type animals (**Figure 7D**), a process known to be antagonized by IL-2, and generated significantly elevated levels of ovalbumin-specific IgG antibody following immunization (**Figure 7E**). To determine whether the 83 kb *Il2* enhancer contributes to auto-inflammatory disease susceptibility *in vivo*, we used an adoptive transfer model of IBD in which the development of colitis is determined by the balance between conventional and regulatory CD4+ T cell activities[81,82]. Upon transfer into lymphocyte-deficient Rag1-ko recipients, wild-type CD4+ T cells respond to microbiota in the gut by inducing inflammatory colitis that leads to weight loss (**Figure 7F**, WT Tconv), while co-transfer of wild-type Treg effectively controls inflammation (**Figure 7F**, WT Tconv + WT Treg). Transfer of *Il2*-83-cRE-ko Tconv led to significantly less colitis (**Figure 7F**, *Il2*-83-cRE-ko Tconv), and was associated with a nominal decrease in *Il2*-83-cRE-ko Tconv in the spleen, and a significant decrease in the accumulation of *Il2*-83-cRE-ko Tconv in the mesenteric lymph nodes (MLN) that drain the intestines (**Figure 7F** inset). Treg require IL-2 from Tconv for their function and homeostasis[83], and despite their reduced numbers, *Il2*-83-cRE-ko Tconv were able to support the homeostasis and function of co-transferred wild-type Treg (**Figure 7F**, *Il2*-83-cRE-ko Tconv + WT Treg). That the -83 kb ortholog of the human -128 kb *IL2* enhancer contributes to physiologic immune responses and inflammatory disease susceptibility *in vivo.*

Impact of autoimmune risk-associated genetic variation on *cis*-regulatory element activity

The chromatin conformation approach used here employs GWAS variants as ‘signposts’ to identify disease-relevant regulatory elements and connect them to their target genes, but does

not *per se* determine the effect of disease-associated genetic variation on enhancer activity or gene expression. We experimentally measured variant effects at the *IL2* locus, where the -128 enhancer defined above contains two SNPs linked to T1D, JIA, and SLE (rs1512973 and rs12504008). Using a recombinant reporter assay in primary activated CD4+ T cells, we confirmed that disease-associated genetic variation influences intergenic *IL2* enhancer activity at the -128 kb element, in that the risk allele contributes significantly less transcriptional activity than the reference allele (**Figure 8A**).

Autoimmune variants are likely to influence disease risk by altering the activity of *cis*-regulatory elements in T cells. At genome scale, we identified over 1000 cRE likely impacted by autoimmune disease-associated genetic variation, in that they contain autoimmune risk SNPs that are predicted to decrease or increase transcription factor binding affinity. Overall, 1370 autoimmune risk variants in open chromatin were predicted to influence the activity of 495 DNA binding factors (**Supplementary File 17**), including PLAG1, PRDM1, BACH2, MYC, TBX21, BHLHE40, LEF1, TCF7, BCL6, IRF, p53, STAT (**Figure 8 – Figure Supplement 1A**), and hundreds of NFkB, EGR, KLF, FKH/FOX, and FOS-JUN sites (**Figure 8B**). For example, the T1D SNP rs3024505 (PP = 0.200) connected to the promoters of *IL9* and *FAIM3* (**Figure 8 – Figure Supplement 1B**) and the celiac SNP rs13010713 (PP = 0.154) contacting the *ITGA4* promoter (**Figure 8 – Figure Supplement 1C**) are predicted to disrupt binding sites for *MZF1* (**Figure 8 – Figure Supplement 1D**) and SOX4 (**Figure 8 – Figure Supplement 1E**). Similarly, the MS SNP rs1077667 contacting the promoters of *GPR108* and *TRIP10* (**Figure 4H**) is predicted to reduce affinity for TP53, TP63, and OCT2/POU2F2 (**Figure 8 – Figure Supplement 1F**).

Functional validation of chromosome capture-based V2G-implicated effector genes

To determine whether genes identified *via* their physical interaction with autoimmune variants in CD4+ T cell chromatin contact maps tend to be directly involved in T cell activation

and function, we compared the set of autoimmune genes implicated by chromatin contacts in this study to sets of genes identified in CRISPR-based screens that control aspects of CD4+ T cell activation like proliferation and expression of the inflammatory genes *IFNG, CTLA4, IL2, IL2RA,* and *TNF*[21–23]. The set of all V2G-implicated genes was highly enriched for genes shown to regulate IL-2, IL-2 receptor, CTLA-4, and proliferation (**Figure 8C, Figure 8 – Figure Supplement 2A, Supplementary File 18**). For example, 202 genes shown to regulate IL-2 production and 166 genes shown to regulate proliferation were also implicated in our autoimmune V2G set (**Figure 8 – Figure Supplement 2A, Supplementary File 16**). Genes implicated by V2G in activated CD4+ T cells were moderately enriched for genes known to control the production of IFN-G, but at the individual disease level, only genes connected to CRO- and ATD-associated variants were enriched for IFNG regulatory genes (**Figure 8C, Figure 8 – Figure Supplement 2A**). Genes contacting CRO, PSO, RA, SLE, T1D and VIT variants were moderately enriched for TNF regulatory genes, but the set of all V2G genes was not enriched for TNF genes. We also queried the orthologs of our V2G-implicated genes from the international mouse phenotype consortium (IMPC) database and identified 97 genes that when knocked out give an immune phenotype and 126 V2G genes that result in a hematopoietic phenotype (**Figure 8D**, **Figure 8 – Figure Supplement 2B**, **Supplementary File 18**). This frequency of observed immune/hematopoietic phenotypes represents a significant (adjP<0.05) ~30% enrichment over expected. This gene set was also enriched for mortality, homeostasis/metabolism, growth/body size, skeleton, and embryonic phenotypes (**Figure 8D**, **Supplementary File 18**).

An important application of this V2G approach is the identification of novel regulators of T cell activation and their potential as drug targets, as nearly 20% of implicated genes have at least one chemical modulator currently available (**Figure 8E**, **Supplementary File 19**). As shown above, this approach validates genes that are well-studied regulators of T cell function, however, a significant portion of implicated genes are not well-studied and are not currently

known to regulate T cell activation (**Figure 8F**). We observed a trend that genes expressed more highly in immune tissues (GTEX) have in general been better investigated, and identified several less-studied genes that could be novel targets, including the kinases GRK6, PTK6, SIK1, and MAP3K11, the G protein-coupled receptors OXER1, GPR183, GPR18, and KISS1R, the acetylcholine receptor CHRNB1, and the *de novo* purine pathway enzyme GART. To determine whether these V2G-implicated genes are novel regulators of T cell activation, we used commercially available pharmacologic modulators in dose-response assays of activation-induced T cell proliferation. Stimulation of T cells in the presence of ligands for CHRNB1, KISS1R and OXER1 did not significantly affect T cell proliferation, however, small molecules targeting GRK6, PTK6, MAP3K11, GPR183, GART, and SIK1 inhibited T cell activation in the nanomolar to micromolar range (**Figure 8G**). Together, these data show that maps of dynamic, 3D variant-gene chromatin contacts in stimulated CD4+ T cells are able to identify genes with *bona fide* roles in T cell activation.

DISCUSSION

By measuring dynamic changes in chromosome folding, chromatin accessibility, and gene expression in naïve CD4+ T cells as a function of TCR-CD28 costimulation, we identified the putative *cis*-regulatory landscape of autoimmune disease-associated genetic variation and physically connected these elements to their putative effector genes. This and prior chromosome capture-based studies show that most cRE and their variants interact with, and therefore have the potential to regulate, more than one gene (median 5 in this study), supporting a scenario in which multiple effector genes are operative at a GWAS locus. We validated a stretch of novel distal elements predicted by our V2G approach as *bona fide* enhancers for the canonical immune gene *IL2*, and showed that autoimmune-associated genetic variation at one of these elements influences it's activity. We also conducted pharmacologic targeting experiments and compared our results with CRISPR-based studies to validate sets of *bona fide*

effector genes with a confluence of multiple orthogonal lines of evidence supporting their role in CD4+ T cell activation.

We also observe that the vast majority (87-95%) of GWAS effector genes predicted by chromosome capture-based approaches are not the genes nearest to the sentinel SNPs queried in this study, while roughly one-third of eGenes predicted by eQTL approaches are the nearest to a sentinel. This and the low concordance between 3D chromatin vs. eQTL eGene predictions is consistent with the view that eQTL-based and GWAS-based approaches inherently implicate different types of genes. While eQTLs cluster strongly near transcription start sites of genes with simple regulatory structures and low enrichment for functional annotations, GWAS variants are generally far from the TSS of genes with complex regulatory landscapes[84]. Moreover, eGene discovery by eQTL studies using scRNA-seq approaches are significantly limited by the low gene detection power inherent to these methods, particularly in rare cell types.

Our systematic comparison of 3D chromatin-based target gene nomination studies in human CD4+ T cells revealed significant variability between datasets, with the highest concordance exhibited by the Javierre and Burren datasets (37%) and the next highest exhibited by our V2G and the Javierre dataset (24%). Although we harmonized the comparisons as much as possible by subjecting each dataset to the same HiC loop calling and GWAS integration steps, there are several potential sources for the observed discrepancy between the studies. The modes of stimulation are largely comparable, but timepoints and donors varied, and ours was the only study that sorted naïve CD4+ T cells prior to stimulation. The higher concordance among promoter-capture datasets compared to our HiC dataset is likely due to their lower resolution compared to our HiC and their greater sequencing depth focused at promoters compared to HiC. The lower resolution of HindIII-based capture-HiC results in loops called to the wrong promoters[24], and will miss distal SNP interactions at any promoters excluded from the capture set. While HiC is unbiased in this regard, high resolution HiC will fail to call some SNP-promoter

loops because the sequencing depth is spread across the whole genome instead of focused at promoters. However, despite variation between studies, the results show the clear value of tissue-matched chromatin conformation maps compared to tissue-averaged HiC for understanding the complex genetic and epigenetic mechanisms that regulate gene expression and for predicting autoimmune effector genes.

Why do the V2G approaches not capture a larger proportion of genes from the human inborn errors in immunity 'truth set'? Low recall/sensitivity could result from the fact that a substantial portion of the mutations in the full HIEI gene set result in immunodeficiency but not autoimmune phenotypes. However, restricting the HIEI truth set to only those disease gene mutations that result in an autoimmune phenotype (143 genes) only increased recall from 16% to 17.5%, suggesting that this is not a major factor reducing sensitivity. Also, our method uses GWAS signals as an input, and unlike GWAS for height, BMI, blood traits, etc., most autoimmune GWAS signals are likely not saturated, which limits discoverability. Another reason for reduced sensitivity is the likelihood that many of the genes in the HIEI truth set operate in cells other than CD4+ T cells, and consistent with this, when the Javierre pcHiC V2G is extended to all immune cell types tested, recall increases from 17% to 27%. This observation emphasizes how limited inclusion cell types and states in a study can significantly limit the power to detect autoimmune effector genes. Why are so few of our 3D chromatin-based V2G genes present in the HIEI truth set? Low precision could be due to false positives in the V2G gene set; *i.e*., contacts between disease-associated cRE and genes that are not in fact involved in disease susceptibility. This does not argue *per se* against the biological relevance of the cRE or the gene, only that the linkage to disease susceptibility is not relevant. Alternatively, the absence of chromatin-GWAS-implicated genes in the truth set could be a false negative from the point of view of the truth set; *i.e*., monogenic, disease-causing mutations in these genes may exist in the human population, but have not yet been discovered. For example, in the year 2000 approximately 100 monogenic mutations causing human inborn errors of immunity were

known, but this increased by ~11 disease mutations per year until 300 human inborn errors of immunity had been identified by 2017. The advent of next-generation exome/genome sequencing and the COVID-19 pandemic resulted in an increase in the rate of HIEI discovery between 2018-2021 to ~40 per year to the current recognized HIEI of ~450. If we assume a conservative continued rate of HIEI discovery of 25 disease genes per year, the next 10 years will show us an additional ~250 disease genes that are not currently contained in the truth set.

Many of the genes identified in this 3D epigenomic V2G screen have known roles in T cell activation and function. An example is *IL2*, and we used the resulting maps to identify and validate a stretch of previously unknown distal enhancers whose activity is required for *IL2* expression and is influenced by autoimmune genetic variation. Another example is the phosphatase DUSP5 that regulates MAPK signaling during T cell activation[85,86]. However, roles for many of the genes implicated here in T cell activation are not known. For example, one of the top implicated genes, PARK7, is a deglycase studied in the context of Parkinson's disease, but has a recently been shown to modulate regulatory T cell function[87]. The orphan G protein-coupled receptor GPR108 is another top gene uniquely implicated by our chromatin-based V2G that has not been studied in T cells, but was identified in a recent CRISPR screen for genes affecting IL-2 levels[23]. Also co-implicated by our study and recent CRISPR screens are the cannabidiol receptor GPR18[88] and the purine biosynthetic enzyme GART[89]. The GPR18 agonist arachidonyl glycine inhibited CD4+ T cell activation above 10 uM, while the GPR18 antagonist O-1918 slightly enhanced T cell activation. This GPR was implicated by both chromatin- and eQTL-based approaches. Antagonism of GART, an enzyme we previous identified as a V2G effector gene in COVID19 severity[58], with the FDA-approved drug lometrexol inhibited T cell activation in the 10 nM range. Antagonism of GRK6, a member of the G-coupled receptor kinase family associated with insulin secretion and T2D susceptibility[90], and PTK6, an oncogenic kinase studied in the context of cancer[91], led to inhibition of T cell activation in the nM to uM range. These targets were implicated by the other chromatin-based approaches but not

by eQTL. Inhibition of MAP3K11, a kinase that facilitates signaling through JNK1 and IkappaB kinase in cancer[92,93] inhibited stimulation-induced CD4+ T cell proliferation in the 100 nM range, as did 25-hydroxycholesterol, a ligand of the G protein-coupled receptor GPR183 that was implicated by both chromatin- and eQTL-based approaches. SIK1 is a member of the salt-inducible kinase family uniquely implicated in our study that negatively regulates TORC activity[61], and a small molecule SIK1 inhibitor potently antagonized stimulation-induced CD4+ T cell activation in the pM range.

Our integration of high-resolution Hi-C, ATAC-seq, RNA-seq and GWAS data in a single immune cell type across multiple activation states identified hundreds of autoimmune variant-gene pairs at ~half of all GWAS loci studied, and application of this technique to additional immune cell types will likely identify effector genes at many of the remaining loci. This study highlights the value of chromosome conformation data as a powerful biological constraint for focusing variant-to-gene mapping efforts[94], and shows that dynamic changes in the spatial conformation of the genome that accompany cell state transitions alter gene expression by exposing promoters to a varying array of *cis*-regulatory elements, transcription factors, and genetic variants.

MATERIALS AND METHODS

T cell isolation and *in vitro* stimulation

Human primary CD4+ T cells were purified from the apheresis products obtained from healthy, screened human donors through University of Pennsylvania Human Immunology Core (HIC). Naïve CD4+ T cells were purified using EasySepTM human naïve CD4+ T cell isolation kit II (STEM cells Technologies, cat#17555) by immunomagnetic negative selection as per manufacturer's protocol. Isolated untouched, highly purified (93-98%) naïve human CD4 T cells were activated using anti-CD3+anti-CD28 Dynabeads (1:1, Thermofisher scientific, cat # 11161D) for 8-24 hours. Cells were then used to prepare sequencing libraries. The human

leukemic T cell line Jurkat was obtained from ATCC, cloned by limiting dilution, and clones with high activation-induced secretion of IL-2 were selected for further study. Single-cell mouse lymphocyte suspensions were prepared from spleen and lymph nodes isolated from 6-week-old female wild-type or *Il2*-83cRE-ko mice (C57BL6 background). Mouse CD4+ T cells were purified by negative selection using a CD4+ T cell isolation kit (Miltenyi Biotec, cat. # 130-104-454). For CD8 depletion, CD8+ T cells were positively stained using CD8a (Ly-2) microbeads (Miltenyi Biotec, cat. # 130-117-044) and the negative flow-through fraction was collected. Tregs were purified from the mouse lymphocytes using a CD4+CD25+ regulatory T cell isolation kit (Miltenyi Biotec, cat. # 130-091-041). The purity of the isolated cells was checked by flow cytometry, showing approximately 95% cell purity. For iTreg induction, CD4+ CD25- T cells (1x 10^6) were activated in a 24-well plate pre-coated with anti-CD3 (1 µg/mL) and the cells were cultured for 72 hours in RPMI culture medium with soluble anti-CD3 (0.5 µg/mL), IL-2 (25 units/mL), TGF-beta (3 ng/mL), anti-IFN-gamma (5 µg/mL), and anti-IL-4 (5 µg/mL). iTreg induction cultures were also set up without adding IL-2. iTreg cultures were harvested after 72 hours, and cells were stained for Foxp3 expression.

<u>*Il2*-83-cRE ko Mice</u>

The CRISPR/CAS method was employed to delete the intergenic -83 CNS sequence between *Il2* and *Il21*. CRISPR guide RNAs were designed (two guide RNAs upstream of -83CNS) and one guide RNA (downstream of -83CNS) using guide RNA design tools (http://crispr.mit.edu.guides). The pX335 plasmid (Addgene #42335) was used to generate a DNA template for sgRNA *in vitro* transcription. To *in vitro* transcribe the sgRNA, a T7 sequence was incorporated at the 5' end of the guide RNA sequence, and a part of the sgRNA scaffold sequence from px335 was added at the 3' end of the guide RNA. Oligonucleotide sequences containing the guide RNA (underlined) along with these added sequences were synthesized by IDT. The oligonucleotide sequences were as follows:

T7_guideRNA#1_scaffold#1:

TTAATACGACTCACTATAGGTTTTCCACGGATCTGCTCGGGTTTTAGAGCTAGAAATAGC

T7_guideRNA#1_scaffold#2:

TTAATACGACTCACTATAGGTGCTTTCTAGGTGAAGCCCCGTTTTAGAGCTAGAAATAGC

T7_guideRNA#1_scaffold#3:

TTAATACGACTCACTATAGGTCATTTGAGCCTAACTACTCGTTTTAGAGCTAGAAATAGC

Along with the above-cited sequences, a reverse primer sequence (AGCACCGACTCGGTGCCACT) from PX335 was used to amplify a PCR-generated template for sgRNA *in vitro* transcription. The PCR product (~ 117 bp) was verified on a 2% agarose gel and then gel-purified using the QIAQuick gel extraction kit. The T7 sgRNA PCR product (500 ng) was used as a template for *in vitro* transcription with the T7 High Yield RNA synthesis kit (NEB, cat # E2040S). The reaction was incubated at 37°C for 4 hours, and then the sgRNA was purified using the MEGAclear kit (Life Technologies, cat # AM1908) according to the kit protocol. The sgRNA was eluted with elution buffer preheated to 95°C. The eluted sgRNA was centrifuged at 13,000 rpm for 20 minutes at 4°C, and the suspension was transferred into a new RNase-free tube. The sgRNA quality was checked by a bioanalyzer using an RNA nano Chip. The sgRNA was diluted to 500 ng/µl in injection buffer (sterile 10 mM Tris/0.1 mM EDTA, pH 7.5) and stored in a -80°C freezer until use. An injection mixture (final volume 30 µl) was prepared in injection buffer by mixing 500 ng/µl of each of the sgRNAs (left and right) with Cas9 mRNA (1 µg/µl, TriLink, cat # L-6125). Fertilized eggs collected from B6/129 mice were microinjected at the CHOP transgenic core and transferred into pseudo-pregnant B6 females. The pups were genotyped using primers flanking the targeted sequence (Forward primer: TTAGGACCTCACCCATCACAA and reverse primer: CATGCCCAGCTACTCTGACAT). The PCR product was cloned into the Promega PGEM T Easy TA cloning vector, and plasmid DNA was Sanger sequenced to determine the size of the deletion. The targeting resulted in mice with a 500 bp and 583 bp deletion at the targeted -83 CNS site and these mutant mice showed same

phenotype. The *Il2*-83-cRE mutant heterozygous male mice were backcrossed with B6 females for 10 successive generations.

ELISA

Cell culture supernatants collected at various time intervals *from in vitro* stimulated T cells of mice or humans were analyzed for IL-2 by ELISA using kits purchased from ThermoFisher Scientific (mouse IL-2 ELISA kit: cat # 88-7024-88) and human IL-2 ELISA kit, cat # 88-7025-76). IL-2 ELISA was performed following the protocol provided by the vendor.

In vivo ovalbumin immunization

Eight week-old female WT and Il2-CNS-83 KO mice were injected intraperitoneally with 50 µg of ovalbumin (Sigma cat # A5503-5G) mixed with IFA (Sigma cat # F5506). Blood was collected from the immunized mice after 10 days, and serum was separated. The level of ovalbumin-specific IgG in the serum was determined by ELISA using an ovalbumin-specific IgG ELISA kit purchased from MyBiosource (cat # MBS763696). The immunized mice were sacrificed on day 10 of immunization, and spleen and lymph nodes were collected. Splenocytes and lymph node cells were stained with CD4 BV785, CD25 BV650, CD44 Percp Cy5.5, CD62L APC eFL780, CXCR5 BV421, PD-1 APC, Bcl-6 PE antibodies. TFH frequency was determined by flow cytometry analysis.

In vivo inflammatory colitis model

Conventional CD4+CD25-ve T cells and CD4+CD25+ Tregs were purified from spleens and lymph nodes of male wild-type or Il2-CNS-83 ko mice (6-8 weeks of age). T cells were transferred into Rag1-ko male mice (5 per group) by retro-orbital injection of 1 million wild-type or Il2-CNS-83 ko Tconv cells alone or along with (0.25×10^6) wild-type Tregs. Experimental Rag1-ko recipients were weighed 3 times per week and scored for IBD-induced clinical symptoms. Mice were sacrificed 72 days post-transfer, and spleen, mesenteric lymph nodes, and colon were collected. Single-cell suspensions were prepared, cells were stained for CD4,

CD8, CD25, CD44, and Foxp3, and analyzed by flow cytometry. The absolute T cell count was estimated using cell count and cell frequency derived from the flow cytometry analysis.

RNA-seq library generation and sequencing

RNA was isolated from ~ 1 million of each cell stage using Trizol Re- agent (Invitrogen), purified using the Directzol RNA Miniprep Kit (Zymo Research), and depleted of contaminating genomic DNA using DNAse I. Purified RNA was checked for quality on a Bioanlayzer 2100 using the Nano RNA Chip and samples with RIN>7 were used for RNA-seq library preparation. RNA samples were depleted of rRNA using QIAseq Fastselect RNA removal kit (Qiagen). Samples were then processed for the preparation of libraries using the SMARTer Stranded Total RNA Sample Prep Kit (Takara Bio USA) according to the manufacturer's instructions. Briefly, the purified first-strand cDNA is amplified into RNA-seq libraries using SeqAmp DNA Polymerase and the Forward and the Reverse PCR Primers from the Illumina Indexing Primer Set HT for Illumina. Quality and quantity of the libraries was assessed using the Agilent 2100 Bioanalyzer system and Qubit fluorometer (Life Technologies). Sequencing was performed on the NovaSeq 6000 platform at the CHOP Center for Spatial and Functional Genomics.

ATAC-seq library generation and sequencing

A total of 50,000 to 100,000 sorted cells were centrifuged at 550 g for 5 min at 4 °C. The cell pellet was washed with cold PBS and resuspended in 50 µL cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% NP-40/IGEPAL CA-630) and immediately centrifuged at 550 g for 10 min at 4 °C. Nuclei were resuspended in the Nextera transposition reaction mix (25 µL 2x TD Buffer, 2.5 µL Nextera Tn5 transposase (Illumina Cat #FC-121-1030), and 22.5 µL nuclease free H2O) on ice, then incubated for 45 min at 37 °C. The tagmented DNA was then purified using the Qiagen MinElute kit eluted with 10.5 µL Elution Buffer (EB). Ten microliters of purified tagmented DNA was PCR amplified using Nextera primers for 12 cycles to generate each library. PCR reaction was subsequently cleaned up using 1.5x AMPureXP beads

(Agencourt), and concentrations were measured by Qubit. Libraries were paired-end sequenced on the Illumina HiSeq 4000 platform (100 bp read length).

Hi-C library preparation

Hi-C library preparation on FACS-sorted CD4+ T cells was performed using the Arima-HiC kit (Arima Genomics Inc), according to the manufacturer's protocols. Briefly, cells were crosslinked using formaldehyde. Crosslinked cells were then subject to the Arima-HiC protocol, which utilizes multiple restriction enzymes to digest chromatin. Arima-HiC sequencing libraries were prepared by first shearing purified proximally-ligated DNA and then size-selecting 200-600 bp DNA fragments using AmpureXP beads (Beckman Coulter). The size-selected fragments were then enriched using Enrichment Beads (provided in the Arima-HiC kit), and then converted into Illumina-compatible sequencing libraries with the Swift Accel-NGS 2SPlus DNA Library Kit (Swift, 21024) and Swift 2S Indexing Kit (Swift, 26148). The purified, PCR-amplified DNA underwent standard QC (qPCR, Bioanalyzer, and KAPA Library Quantification [Roche, KK4824]) and was sequenced with unique single indexes on the Illumina NovaSeq 6000 Sequencing System using 200 bp reads.

ATAC-seq data analysis

ATAC-seq peaks from libraries unstimulated and stimulated Naïve CD4+ T cells were called using the ENCODE ATAC-seq pipeline (https://www.encodeproject.org/atac-seq/). Briefly, pair-end reads from three biological replicates for each cell type were aligned to hg19 genome using bowtie2[95], and duplicate reads were removed from the alignment. Narrow peaks were called independently for each replicate using macs2[96] (-p 0.01 --nomodel --shift -75 --extsize 150 -B --SPMR --keep-dup all --call-summits) and ENCODE blacklist regions (ENCSR636HFF) were removed from peaks in individual replicates. Reproducible peaks, peaks called in at least two replicates, were used to generate a consensus peakset. Signal peaks were normalized using csaw[97] in 10kb bins background regions and low abundance peaks (CPM>1) were excluded from the analysis. Tests for differential accessibility was conducted with the glmQLFit approach

implemented in edgeR[98] using the normalization factors calculated by csaw. OCRs with FDR < 0.05 and abs($\log_2$FC > 1) between stages were considered differentially accessible.

Hi-C data analysis

Paired-end reads from two replicates were pre-processed using the HICUP pipeline v0.7.4[99], with bowtie as aligner and hg19 as the reference genome. The alignment .bam file were parsed to .pairs format using pairtools v0.3.0 (https://github.com/open2c/pairtools) and pairix v0.3.7 (https://github.com/4dn-dcic/pairix), and eventually converted to pre-binned Hi-C matrix in .cool format by cooler v0.8.10[100] with multiple resolutions (500bp, 1kbp, 2kbp, 2.5kbp, 4kbp, 5kbp, 10kbp, 25kbp, 40kbp, 50kbp, 100kbp, 250kbp, 500kbp, 1Mbp and 2.5Mbp) and normalized with ICE method[101]. Replicate similarity was determined by HiCRep v1.12.2[100] at 10K resolution. For each sample, eigenvectors were determined from an ICE balanced Hi-C matrix with 40kb resolution using cooltools v0.3.2[102] and first principal components were used to determine A/B compartments with GC% of genome region as reference track to determine the sign. Differential TAD comparison was performed using TADcompare with the default settings for each chromosome (v1.4.0)[103]. Finally, for each cell type, significant intra-chromosomal interaction loops were determined under multiple resolutions (1kb, 2kb and 4kb) using the Hi-C loop caller Fit-Hi-C2 v2.0.7[104] (FDR<1e-6) on merged replicates matrix. The consensus chromatin loops within resolution were identified by combining all three stages. These sets of loops were used as consensus for quantitative differential analysis explained below. The final consensus interaction loops for visualization were collected by merging loops from all the resolutions with preference to keep the highest resolution. Quantitative loop differential analysis across cell types was performed on fast lasso normalized interaction frequency (IF) implemented in multiCompareHiC v1.8.0[105] for each chromosome at resolution 1kb, 2kb and 4kb independently. The contacts with zero interaction frequency (IF) among more than 80% of the samples and average IF less than 5 were excluded from differential analysis. The QLF test based on a generalized linear model was performed in cell type-pairwise comparisons, and p-values were

corrected with FDR. The final differential loops were identified by overlapping differential IF contacts with consensus interaction loops.

Bulk RNA-seq data analysis

Bulk RNA-seq libraries were sequenced on an Illumina Novaseq 6000 instrument. The pair-end fastq files were mapped to the genome assembly hg19 by STAR (v2.6.0c) independently for each replicate. The GencodeV19 annotation was used for gene feature annotation and the raw read count for gene feature was calculated by htseq-count (v0.6.1)[106] with parameter settings -f bam -r pos -s yes -t exon -m union. The gene features localized on chrM or annotated as rRNAs, small coding RNA, or pseudo genes were removed from the final sample-by-gene read count matrix. Gene set variation analysis was performed using GSVA 1.42.0[107] with the MSigDB hallmark geneset[108], with resulting scores analyzed using limma (limma_3.50.3)[109]. Low abundance peaks (CPM>1) were excluded from the analysis. Testing for differential expression was conducted with the glmQLFit approach implemented in edgeR[98]. Genes with FDR<0.05 and abs($\log_2$FC>1) between stages were considered differentially expressed. Differential genes were then clustered using k-means clustering. The number of clusters was determined using the elbow method on the weighted sum of squares, where was set to k=5. Score for how similar each gene followed the clusters expression pattern was determined by calculating pearson correlation coefficients between each gene in the cluster and the cluster centroid.

Transcription factor footprinting and motif analysis

Transcription factor footprints were called using Regulatory Analysis Toolbox HINT-ATAC (v0.13.0) with pooled ATAC-seq data for each stage and consensus peak calls[110]. The rgt-hint footprinting was run with parameters –atac-seq, --paired-end, and organism=hg19 set. The output footprint coordinates were subsequently matched using rgt-motifanalysis matching with parameters --organism hg19 and –pseudocount 0.8 set. The JASPAR2020 position weight matrix database was used to match footprints[111]. Differential analysis of TF binding across conditions was performed using rgt-hint differential with parameters –organism hg19, --bc, --nc

24 using the motif matched transcription factor footprints. An activity score is then calculated based on the accessibility surrounding the footprint.

Partitioned heritability LD score regression enrichment analysis

Partitioned heritability LD score regression[112] (v1.0.0) was used to identify heritability enrichment with GWAS summary statistics and open chromatin regions annotated to genes. The baseline analysis was performed using LDSCORE data (https://data.broadinstitute.org/alkesgroup/LDSCORE) with LD scores, regression weights, and allele frequencies from the 1000G phase 1 data. The summary statistics were obtained from studies as described in **Supplementary File 14** and harmonized with the munge_sumstats.py script. Annotations for Partitioned LD score regression were generated using the coordinates of open chromatin regions that contact gene promoters through Hi-C loops for each cell type. Finally, partitioned LD scores were compared to baseline LD scores to measure enrichment fold change and enrichment p-values, which were adjusted with FDR across all comparisons.

Variant-to-gene mapping using HiC-derived promoter contacts

95% credible sets were determined as previously described[113]. Briefly, P values from GWAS summary statistics were converted to Bayes factors. Posterior probabilities were then calculated for each variant. The variants were ordered from highest to lowest posterior probabilities added to the credible set until the cumulative sum of posterior probabilities reached 95%. Variants in the 95% credible set were then intersected with the CD4+ T cell promoter interacting region OCRs from the three timepoints using the R package GenomicRanges (v1.46.1)[114].

Genomic reference and visualizations

All analyses were performed using the hg19 reference genome using gencodeV19 as the gene reference. Genomic tracks were visualized with pyGenomeTracks v3.5[115]. HiC matrices depict the log1p(balanced count) from the cooler count matrix. ATAC-seq tracks were generated from bigwig files that were normalized using deeptools[116].

ABC model predictions

We used the ABC model (PMID: 31784727) using the our Tcell ATAC-seq data (unstimulated or 24 hours stimulated) generated from Naive CD4+ T cells merged acrossed replicates. For H3K27ac we retrieved paired fastq files from H3K27ac MINT-ChIP data from ENCODE for resting (unstimated) and activated (36hr stimulated) CD4 T cells derived from thymus. The deduplicated bam files were used as signal files for the ABC model, the consensus peak set were used as the input set of enhancers. The accession numbers for unstimulated cells are: ENCFF732NOS, ENCFF658TLX, ENCFF605OGN, ENCFF012XYW, ENCFF407QZP, ENCFF556QWM, ENCFF888IHV, ENCFF260FUF, ENCFF550EVW, ENCFF747HGZ, ENCFF067XWQ, ENCFF939YBH, ENCFF749LGW, ENCFF788LWV, ENCFF610FRB, ENCFF538GQX, ENCFF355PCA, ENCFF040OMT, ENCFF878SHZ, ENCFF210WNQ, ENCFF849ZJH, ENCFF925ARM, ENCFF112YNU, ENCFF808JFV, ENCFF448KHE, ENCFF591BFV, ENCFF366UTF, ENCFF172EWR. The accession numbers for stimulated cells are: ENCFF442DFF, ENCFF870ZYC, ENCFF189UPS, ENCFF725KSA, ENCFF618YAV, ENCFF765AJK, ENCFF854QUS, ENCFF180TKV, ENCFF761RCK, ENCFF449EAE, ENCFF915VPK, ENCFF575XFH, ENCFF227AGX, ENCFF810XFH, ENCFF367OMH, ENCFF114BBZ, ENCFF855XPP, ENCFF132AKN, ENCFF663MHJ, ENCFF244TPB, ENCFF230RIJ, ENCFF648FPC, ENCFF797CIK, ENCFF096JPW, ENCFF704KMT, ENCFF385POD, ENCFF395RIG, ENCFF043RRJ, ENCFF027RNR, ENCFF333RGD, ENCFF048PMV, ENCFF887XSA, ENCFF188CNA, ENCFF614IZY, ENCFF984TTZ, ENCFF977RDC, ENCFF393EWO, ENCFF781VQT, ENCFF638UCZ, ENCFF873QZH, ENCFF553JPM, ENCFF589HRL, ENCFF721ZCN, ENCFF496HYV, ENCFF455WBD, ENCFF679UTJ, ENCFF200MYN, ENCFF459ARH.

The retrieved files were trimmed/preprocessed using fastp with the default settings. Following this, matched R1 and R2 files were concatenated to a single R1/R2 file pair. The reads were aligned to hg19 and duplicates were removed as described for the ATAC-seq data analysis. The deduplicated bam files were used as input to the ABC model. We ran the ABC model with

the default settings for hg19 and provided reference files for the blacklist, promoter annotation, chromosome sizes reference, and set of ubiquitiously expressed genes. The HiC data used was the KR normalized averaged HiC dataset that was derived from a set of cell lines (GM12878, NHEK, HMEC, RPE1, THP1, IMR90, HU- VEC, HCT116, K562, KBM7; average_hic.v2.191020.tar.gz) at 5kb resolution then scaled by powerlaw distribution as described in the original manuscript. The default threshold of 0.2 was used to link enhancers to genes. The resulting elements were intersected with the credible set list used in this study for precision-recall analyses.

Precision-recall analyses against HIEI genes

To benchmark our autoimmune effector gene predictions against prior eQTL and chromatin capture studies, we curated a list of 449 expert curated genes from Human Errors in Inborn Immunity 2022, where mutations have been reported to cause immune phenotypes in humans[65]. Most of these genes were identified through rare loss of function mutations in the coding region. In addition to this we also took a subset of 145 genes reported to specifically result in autoimmune phenotypes. These two sets of genes were treated as “gold standard” sets of genes to compare different approaches to predict GWAS effector genes. We curated the results of several other chromatin conformation or eQTL based methods of assigning variants to genes to compare with our study. For the chromatin confirmation studies, we obtained loop calls from the associated publications and intersected with the list variants in the 95% credible set from 15 GWAS used to predict effector genes using GenomicRanges (v1.46.1)[114]. For eQTL studies, we obtained the summary stats files and identified eGenes linked to any member of the credible set reported as an eQTL in the transcriptome wide association analysis (TWAS; p-value < 1e-4). Precision was calculated as the ratio of the number predicted genes in the truth set to the total number of predicted genes, while recall was calculated as the ratio of the number predicted genes in the truth set to the total number of genes in truth set.

Colocalized eQTL comparisons

In addition to comparisons with the eGenes identified by TWAS in the Soskic et al. study, we also compared overlap of gene nominations in our study with the eQTLs that colocalized with an autoimmune GWAS[13]. Enrichment of sentinel-gene assignments was conducted similarly as described previously[56]. Briefly, a null distribution was constructed by randomly selecting genes within 1 Mb of the sentinel compared to the set of colocalized cis-eQTL[13] found in through 10,000 iterations. The observed overlap reports the set of gene identified by both our HiC based approach with the set of colocalized eQTLs. We report the empirical P value of the observed value relative to null distribution.

International mouse phenotyping consortium comparisons

The set of HiC implicated genes were compared to the mouse international phenotyping consortium set of genes with reported phenotypes[117]. We converted the list of V2G implicated genes to mouse homologs using homologene. We tested for enrichment for each phenotype using a one-sided proportion test implicated in R prop.test with type set to “upper”.

Identification of pharmacological agents

We queried the Drug-Gene Interaction Database with the set of V2G implicated genes for chemicals using rDGIdb (v1.20.0)[118]. To identify the number of papers for each gene with at least one drug annotated to target it, we queried pubmed titles and abstracts using the R package RISmed (v2.3.0) with each gene’s name and either “autoimmune” and the list of autoimmune diseases (**Supplementary File 17**). A score to approximate expression specificity was computed using the sum GTEX median expression values (v8) for whole blood or spleen divided by other tissues[119].

Lentiviral-based CRISPR/CAS9 targeting in Jurkat cells

LentiCRISPRv2-mCherry vectors encoding gRNA-CAS9 and the fluorescent reporter mCherry were used for Jurkat targeting. CRISPR guide RNAs (sgRNA) targeting human IL-2-21 intergenic -46, -51, -80, -83, -85, -122, -128 CNS regions were designed using http://crispr.tefor.net and cloned into lentiCRISPRv2-mCherry. Empty vector without gRNA

insert was used as a control. Below is the list of CRISPR gRNA for causing deletion of human IL2-21 intergenic regions:

CNS region	CRISPR gRNA	
-46 CNS	gRNA #1	AGGATGCCTACCTCCAAATG
	gRNA # 2	AGGTGACAACATTTAGTCAG
-51 CNS	gRNA #1	GGCAACGAAATTCACTGTGA
	gRNA # 2	ATTCTAACAGGAATCATTCG
-80 CNS	gRNA # 1	GTTCTACCTATGCCGCATTG
	gRNA # 2	GAGATTTACTCAGTCCAATG
-83 CNS	gRNA # 1	GTGACAAGCATGACTCTACA
	gRNA # 2	GTGATGGTGAATTAAGCTGA
-85 CNS	gRNA #1	AGGGTTTTCTAGTTACGAGA
	gRNA # 2	ATGGTTAGTTAGCTCCCAAG
-96 CNS	gRNA#1	TGGGAAAAACATCTTACCTG
	gRNA # 2	TGGCCCATGAACCATCAAAG
-122 CNS	gRNA# 1	GTTATTAATCTAAGCGGAGA
	gRNA# 2	GGAAGTTAGGCAGTCAATCG
-128 CNS	gRNA#1	CTTCAATCATTGCATTCCAC
	gRNA# 2	TGACACCACCCCTGCTTGAG

HEK 293T cells were grown in RPMI1640 complete medium (RPMI1640 + 1X P/S, 1x L-Glu, 10% FBS), 37C, 7%CO2. 293T cells were transfected with 10 ug of lenti-CRISPR-V2-CRE construct along with packaging plasmid 6 ug of PsPAX2 (Addgene, Cat #12260) and 3.5 ug of PmD2.G (Addgene, Cat #12259) using Lipofectamine 2000 transfection reagent (Invitrogen cat # 11668019). After 6 hours, the transfection medium was replaced with complete culture medium. Transfected cells were incubated at 37C for 48-72 hours in a cell culture incubator. and then the Lentiviral supernatants were harvested and spun at 300g for 5 minutes to remove

cellular debris. Lentiviral supernatants were concentrated using Lenti-X™ Concentrator (Takara Bio, Cat # 631232) and then centrifuged at 1500g for 30 minutes at 4°C and supernatant was discarded. The lentiviral pellet was resuspended at a ratio of 1:20 of the original volume using RPMI media and concentrated virus supernatant aliquots were prepared and stored until use at -80°C. To achieve high transduction efficiency, the viral supernatant was titrated in Jurkat cells through transduction using various dilutions of the viral supernatants and transduction efficiency was determined by mcherry expression analyzed through flow cytometry. Jurkat cells were seeded in a 24 well plate at 0.5 x106/well in culture media, viral supernatant with 8 ug/mL of polybrene was added to each well. Spinfection was performed for 90 min. at 2500 rpm, and transduced cells were equilibrated at 37C for ~6 hrs, followed by incubation at 37C 5%CO2 for ~72 hours culture. Transduced Jurkat Cells were then harvested and stimulated using PMA (15 ng/mL) + Ionomycin (1 uM) + human anti-CD28 (2 ug/ml), BioXcell cat # BE0248 in 96 well culture plates (TPP, cat # 92097) in triplicates. Cell culture supernatants were collected at the end of culture and analyzed for IL-2 by ELISA using a kit (cat # 88-7025-76) purchased from Thermofisher Scientific.

gRNA-CAS9 RNP-based targeting in primary human CD4+ T cells

Primary Human CD4 T cells derived from 5 normal healthy donors were obtained from Human Immunology core (University of Pennsylvania). Alt-R S.p. HiFi Cas9 Nuclease V3 cat # 1081061 CAS9 and following list of Alt-R® CRISPR-Cas9 sgRNA targeting IL-2-21 CRE were purchased from Integrated DNA Technologies, USA.

Name	20nt gRNA Seq
46CNS-1	GACGTATATAGTCATCTGAT
46CNS-2	TCTGTGGAGCTGCTGCGTTA
46CNS-3	AGGATGCCTACCTCCAAATG
51CNS-1	ATTCTAACAGGAATCATTCG

51CNS-2	GAGTAAAAAGACGTGTTACC
51CNS-3	GGCAACGAAATTCACTGTGA
80CNS-1	AATGCGGCATAGGTAGAACT
80CNS-2	GAGATTTACTCAGTCCAATG
80CNS-3	TGGTTGTCACAGTAACTAGG
83CNS-1	GTGACAAGCATGACTCTACA
83CNS-2	GTGATGGTGAATTAAGCTGA
83CNS-3	TGGGCTCTGACTCACTTAGA
85CNS-1	GCTAACATTGACTTCTCTAC
85CNS-2	AATCTATGCAAGGGGTGAAT
85CNS-3	GCATCATGATGAAGCTTATC
122CNS-1	GTTATTAATCTAAGCGGAGA
122CNS-2	GGAAGTTAGGCAGTCAATCG
122CNS-3	CACTTTGTGCTCGGATGCTC
128CNS-1	CTTCAATCATTGCATTCCAC
128CNS-2	TCAAGCAGGGGTGGTGTCAA
128CNS-3	TGGTGATTCATCTTTAGCAT
IL-2URR-1	TCCATTCAGTCAGTCTTTGG
IL-2URR-2	TAGTGTCCCAGGTGATTTAG
IL-2URR-3	TAGAGCTATCACCTAAGTGT
B2m-1	GAGTAGCGCGAGCACAGCTA
B2m-2	CACGCGTTTAATATAAGTGG

Primary Human CD4 T (5-10e6) were incubated with sgRNA and CAS9 protein complex and electroporation was done using P3 Primary Cell 4D-Nucleofector™ X Kit L (Lonza, cat # V4XP-

3024) and Lonza 4D nucleofector system. B2M gene was CRISPR targeted as a positive control. As per manufacturer's protocol cells were electroporated pulse code FI-115 in 100ul cuvette format. After nucleofection, cells were allowed to rest in the complete media for ~2 days. Cells were then harvested, washed with PBS and aliquots of cells were used for further experimentation such as flow staining and cell activation. Primary human CD4 T Cells (0.1e6/well) were seeded in triplicates (for each experimental condition) in 96 well plate format in RPMI complete medium and stimulated using ImmunoCult™ Human CD3/CD28 T Cell Activator (STEM cells Technologies, cat # 10971). Cell culture supernatants were collected at 4h of stimulation and stored at -80C until assayed. IL-2 ELISA was performed using Thermofisher Scientific IL-2 Human Uncoated ELISA Kit with Plates, cat # 88-7025-76.

Recombinant reporter assays

The *IL2* URR was cloned into Xho I and Hind III sites of PGL4 Luc vector (Promega) and then ~500 bp individual sequence representing *IL2-IL21* intergenic CNS at -46, -51, -80, -83, -85, -128 were cloned upstream to IL-2 URR at the Xho I site of the pGL4 vector. Primary human CD4 T cells obtained from 5 normal healthy donors were transfected with PGL4-cRE-URR constructs. Briefly, primary human CD4 T cells were activated with anti-CD3+ anti-CD28 Dynabeads (1:1) (Thermofisher scientific, cat # 11161D) + IL-2 overnight and then 1 million aliquots of cells (triplicates) were electroporated using Nucleofector 2b, human T cell Nuclefactor kit (Lonza VPA # 1002, program # T-020) with 2ug of PGL4 firefly vector constructs along with 0.2 ug of PGL4 Renilla vector; cells were allowed to rest ON in RPMI + IL-2 and then re-stimulated with plate-bound anti-CD3+ anti-CD28 (2 ug/ml each) for 5 hrs. Dual luciferase assay was performed with cell lysate prepared using Promega dual luciferase assay kit. Cell lysate was prepared in PLB as per manufacturer's protocols and then firefly and renilla luciferase activities were analyzed by spectramax ID5 (Molecular Devices). Firefly luciferase activity was normalized against the internal control renilla luciferase activity.

Pharmacologic validation of novel T cell activation regulatory genes

The drugs lometrexol (GART antagonist), 25-hydroxycholesterol (GPR183 ligand), epinephrine & norepinephrine (CHRNB agonists), and arachidonoyl glycine (GPR18 agonist) were purchased from Sigma. The GPR18 antagonist O-1918 was purchased from ChemCruz. CEP1347 (MAP3K11 antagonist), kisspeptin 234 (KISSIR ligand) and the PTK6 antagonist tilfrinib were purchased from Tocris. The GRK6 antagonist GRK-IN-1 was purchased from DC Chemicals. The SIK1 antagonist HG-91-9-1 was purchased from Selleckchem. The OXER1 agonist 5-Oxo-ete was purchased from Cayman chemicals. The drugs were dissolved in DMSO or ethanol as suggested by the vendor and working stocking concentrations were prepared in RPMI1640 medium or PBS. The drug effect on T cell proliferation was assayed using murine lymphocytes cultured under TCR and CD28 activation conditions. Spleen and lymph nodes were collected from 6-8 weeks old female C57BL/7 mice and single cell lymphocyte cell suspensions were prepared in RPMI 1640 complete medium. 20 million lymphocytes were labeled with CFSE and re-suspended in RPMI 1640 medium. Cells (0.5 million labeled cells/well) were loaded into 48 well culture plate and activated with mouse anti-CD3 and anti-CD28 agonistic antibodies (1μg mL each). Drugs at the indicated concentrations were added in the culture medium and both untreated and drug treated cell cultures were incubated at 37°C for 72 hours in a cell culture incubator. Cells were harvested after 3 days of culture, washed with PBS and then stained with live-dead aqua dye. After washing with FACS buffer (PBS containing 2% FBS), cells were stained with fluorochrome conjugated antibodies CD4-APC, CD8-PB, CD44-Percp Cy5.5 and CD25-BV650. Stained cells were analyzed on a Beckman Coulter Cytoflex S flow cytometer. The division profile of CD4+ CFSE+ T cells were gated on live populations. The flow data was analyzed using Flowjo10 software and the number of divided CD4+ T cells were determined as described previously[120].

FIGURE LEGENDS

Figure 1: Defined gene expression dynamics throughout activation of naïve CD4+ T cells. (**A**) Diagram of study design: RNA-seq, ATAC-seq, and HiC libraries were prepared from sorted CD4+CD62L^{hi}CD44^{lo} naïve T cells (donor N=3) prior to or following 8hr or 24hr stimulation with anti-CD3/28 beads. cREs are classified to genes allowing for interrogation of the pattern of expression, accessibility, and chromatin structure changes of autoimmune-associated variants. (**B**) Heatmap showing normalized expression of known markers of T cell activation. (**C**) K-means clustering of differentially expressed genes into five groups using the elbow and within-cluster-sum of squares methods to select the number of clusters. The centroid of the cluster is depicted as a black line, and members of the cluster are depicted as colored lines. Color indicates the Pearson's correlation coefficient between the gene and the cluster centroid, with red indicating higher correlation, and blue lower correlation.

Figure 1 – Figure Supplement 1: Library sequencing reproducibility and expression clustering. (A) PCA of RNA-seq libraries. (B) PCA of ATAC-seq libraries. (C) Stratified correlation (SCC) of HiC libraries from two donors by stage of activation. (D) Volcano plot of RNA-seq data for CD4+ T cells unstimulated vs. 8 hr stimulation (top) and 8 hr vs. 24 hr stimulation (number of DEG indicated). (E) Elbow plot of the within group sum of squares used to determine optimal cluster number. (F) Genes per expression cluster. (G) Centroids +/- standard deviation for each cluster.

Figure 2: Chromatin accessibility and architecture changes across three stages of T cell activation. (**A**) Volcano plots depicting the $\log_2$ change in accessibility in all reproducible OCRs (present in 2 of the 3 replicates) compared to the $-\log_{10}$(FDR), significant points are indicated in red (FDR < 0.05; abs($\log_2$FC) > 1). (**B**) Genomic regions were classified by membership in A vs. B compartments at each time point. The pie chart depicts regions with differential A/B compartment assignment in response to activation. (**C**) TAD structure was determined at each time point, and regions exhibiting a shift, split/merge, change in strength, or more complex rearrangements are depicted in the pie chart. (**D**) Loop calls identified from HiC data for each

timepoint called at three resolutions (1kb, 2kb, 4kb bins). (**E**) Differential loop-calls called across all resolutions. (**F**) Density plots of cRE accessibility and gene expression separated by whether contact frequency increased, decreased, or remained stable from the transition from unstimulated to 8 hour activation and 8 hour to 24 hour activation.

Figure 2 – Figure Supplement 1: Characterization of CD4+ T cell epigenomic data. (A) Accessibility of OCR and (B) expression of genes located in A vs. B compartments (log2 fkpm). Differential accessibility (C) and expression (D) of cRE-gene pairs in regions exhibiting differential TAD structure. (E) Enrichment of promoter interacting region OCR for annotated regulatory elements in CD4+ T cells from the epigenome roadmap project (E038_15). (F) Overlap of cRE-promoter loops defined in this study to loops defined by Burren et al. in a prior promoter capture Hi-C study. (G) Distribution of OCR contacted per gene (top, median = 8 OCRs per gene). Distribution of genes contacted per OCR (bottom, median = 3). (H) Differential promoter-OCR connectivity by HiC as a function of differential OCR accessibility (central bar = median, boxplot edges = 25-75 quantiles, whiskers = 1.5x inter-quantile range).

Figure 2 – Figure Supplement 2: Characterization of chromatin stripes during CD4+ T cell activation. (A) Stripes called per timepoint. (B) Proportion of SNPs annotated in A vs B compartments. Accessibility of OCR (C) and expression of genes (D) located in anchor (small more interactive region of stripe), stripe excluding the anchor (stripe), or outside stripes.

Figure 2 – Figure Supplement 3: Large-scale epigenetic changes at dynamically expressed genes during CD4+ T cell activation. HiC contact frequency matrix (heatmap), chromatin accessibility (grey), and loop calls (red) for (A) GEM, (B) IRF4, (C) KLF2, and (D) DPEP2. The IRF4 matrix is truncated at by end of the chromosome annotation, and inset shows the expression (TPM) at each time point for each gene.

Figure 3: Transcription factors prediction potential regulators of chromatin status and expression changes. Footprints were annotated to each TF motif by sequence matching to the

JASPAR motif database, and the average accessibility of the region surrounding each motif was used as a measure of activity at each time point. Scatterplots depict the change in accessibility for each TF motif (activity score) and $\log_2$FC of TF gene expression between unstimulated and 8 hour activation (**A**) or 8 hour and 24 hour activation (**B**). TF were classified as differentially expressed (orange), differentially active (brown), both (red), or neither (grey). Dot size indicates the number of predicted footprints occupied by each motif. (**C**) Average accessibility (normalized by depth and motif number) in a 200bp window surrounding motif footprints for FOSL2::JUNB, Smad2::Smad3, MAZ, and KLF2 from three timepoints (unstim: green; 8hr: blue; 24hr: red). (**D**) Z-score of TF motif enrichment for cRE connected to the five clusters compared to all OCR. Lighter color indicates higher specificity of enrichment for that TF cluster. (**E**) Connections between different TFs based on physical interactions and predicted regulation-based co-expression determined by GENIE3 for the three timepoints. Node color indicates expression (TPM; darker = higher expression, lighter = lower expression), edge color reflects confidence in the interaction called by GENIE3 (darker higher confidence).

Figure 3 – Figure Supplement 1: Enrichment of transcription family members and gene regulatory network construction. (A) -log10(P) enrichment of transcription factors annotated by family for each expression cluster. (B) Top five pathways enriched for a subset of TF-regulated gene expression.

Figure 4: Variant-to-gene mapping of autoimmune-associated loci implicates genetic variants in the control of T cell activation. (**A**) Identification of genes in contact with open chromatin regions harboring autoimmune-associated SNPs. Sentinel SNPs from 15 autoimmune GWAS (**Table S11**) were used to identify the 95% credible set of proxy SNPs for each lead SNP. SNP locations were integrated with ATAC-seq and HiC data to identify the 95% credible set of accessible SNPs in physical contact with a gene promoter. Genes are further refined based on expression dynamics over the time course of T cell activation. (**B**) Partitioned LD score regression for autoimmune GWAS studies using the OCR in contact with the genes in

the five clusters defined by RNA-seq. Circle size = enrichment, color = significance, **FDR<0.01, ***FDR<0.001. (**C**) Log distribution of the 1D distance between each proxy SNP and its interacting gene based on 3D chromatin conformation (median=103 kbp). (**D**) Distribution of the number of genes contacted by each accessible variant (median=5). (**E**) Fraction of open disease-associated variants that interact with no gene promoters (grey), only with the nearest gene promoter (purple), with the nearest gene and a distant gene(s) (orange), and with only a distant gene(s) (blue). (**F**) Number of genes identified by variant-to-gene mapping at each time point. Black = shared in all stages, blue = shared in two stages, red = specific to one stage. (**G**) Set of implicated proxy-genes pairs that are both differentially expressed, display differential accessibility, and chromatin contact across T cell activation. (**H**) Example MS variant rs1077667 (PP=1.0) exhibits increased accessibility and contact with the promoters of the *GPR108* and *TRIP10* at 8 hours post activation, which are increased in expression at this time point. (**I**) Example allergy variant rs7380290 interacts with the *SIK1* gene that is upregulated after activation.

Figure 5: **Comparative predictive power of orthogonal V2G approaches**. (**A**) Percent overlap in predicted genes among chromatin-based and eQTL-based V2G approaches. (**B**) Physical variant-gene associations (HiC) are enriched for statistical variant-gene associations (eQTL) in activated CD4+ T cells. The histogram depicts the null distribution of shared variant-gene pairs expected at random (~5) while the red line indicates the observed number of variant-gene pairs (41) shared with the 127 eQTL identified by Soskic et al. in a similar CD4+ T cell activation system. (**C**) Expression of the shared genes. (**D**) Precision-Recall analysis of V2G gene predictions against the set of monogenic human inborn errors in immunity.

Figure 6: Functional validation of autoimmune V2G-implicated cRE at the *IL2* locus. (**A**) The combination of evolutionary conservation (blue), open chromatin (red), and autoimmune disease-associated SNPs at the *IL2* locus identify putative cRE in quiescent vs. 8-hour activated naïve CD4+ T cells. (**B**) Activation-dependent TAD/sub-TAD structure (heatmaps), chromatin

remodeling (grey) and promoter-OCR interactions (red) at the *IL2* locus. (**C**) Recombinant reporter assay showing transcriptional activity of the *IL2* URR (+35 to -500 from the TSS) in activated primary CD4+ T cells (N= 7 donors) compared to a basal promoter (pGL4, left panel). The right panel depicts transcriptional activity of the CNS regions indicated in **A** cloned upstream of the URR. All regions except the -80 CNS show statistically significant activity relative to the URR alone (N=7, P<0.05, line = median, box = 95/5% range). (**D**) Scheme of CRISPR/CAS9-based deletion of individual *IL2* CNS using flanking gRNAs. (**E**) Activation-induced secretion of IL-2 protein by CRISPR-targeted primary CD4+ T cells (N=5 donors, left panel) or Jurkat cells (N=3 replicates, right panel) relative to untargeted control (CAS9, no gRNA) cells (****P<0.0001, ***P<0.001, **P<0.01, *P<0.05, line = median, box = 95/5% range). In primary cells, *B2M* gRNAs served as an irrelevant targeted control. (**F**) Scheme of CRISPR/CAS9-based deletion of the 81.3 kb region containing all distal *IL2* cRE using flanking gRNAs in Jurkat cells. (**G**) Activation-induced IL-2 protein (left panel) and mRNA (right panel) by control (black) vs. 81.3 kb deleted (red) Jurkat cells (N=3 separate clones).

Figure 6 – Figure Supplement 1: TF occupancy and stability of distal IL2 cRE in CD4+ T cell subsets. Evolutionary conservation (blue), open chromatin (red), and autoimmune disease-associated SNPs at the IL2 locus in quiescent (A) and 8-hour activated (B) naïve CD4+ T cells. (C) Curated TF occupancy at the IL2 URR and distal cRE from the ReMap atlas of regulatory regions.

Figure 7: Distal *IL2* enhancers are evolutionarily conserved and impact *in vivo* T cell-mediated immunity in mice. (**A**) Map of the human and mouse *IL2* locus depicting mammalian conservation (dark blue), ATAC-seq from activated human and mouse CD4+ T cells (red), and orthologous non-coding sequences conserved between mouse and human (light blue). (**B**) IL-2 protein secretion (top left panel) and *Il2* mRNA expression (top right panel) by CD4+ T cells from wild-type or *Il2*-83 cRE ko mice in response to stimulation with plate-bound anti-CD3 and anti-CD28 Ab *in vitro*. Bottom panels show soluble anti-CD3-induced *in vitro* clonal expansion of

wild-type or *Il2*-83 cRE ko CD4+ (left) and CD8+ (right) T cells measured by dye dilution[120]. (**C**) Foxp3 expression by murine CD4+CD25- Tconv stimulated with anti-CD3 Ab and TGF-beta in the absence (top panels) vs. presence (bottom panels) of exogenous IL-2. Wild-type or *Il2*-83 cRE ko mice (N=6) were immunized intraperitoneally with chicken ovalbumin in incomplete Freund's adjuvant. The frequency of CD4+PD1+CXCR5^{hi} follicular helper T cells in the spleens of 3 animals was measured by flow cytometry at day 5 post-immunization (**D**), and levels of ovalbumin-specific IgG in the serum of 3 animals were measured at day 10 post-immunization (**E**). (**F**) In vivo IBD model. CD4+CD25- Tconv from wild-type (pink, red) or *Il2*-83 cRE ko (purple, blue) mice were transferred alone (purple, pink) or together with wild-type CD4+CD25+ Treg (red, blue) into RAG1-ko mice (N=5 per group). Animal weight was monitored for 40 days, and the number (in millions) of activated CD4+ Tconv in the spleens and mesenteric lymph nodes was measured by flow cytometry (inset). Statistical significance (*p<0.05) was measured by ANOVA or t-test.

Figure 8: Functional validation of autoimmune V2G-implicated genes. (**A**) Recombinant reporter assay in primary activated CD4+ T cells (N= 7 donors) showing transcriptional activity of the reference vs. risk alleles of the *IL2* -128 cRE relative to the URR alone (P<0.0001). (**B**) Prominent TF motifs predicted to be disrupted (blue) or stabilized (red) by promoter-connected autoimmune SNPs. (**C**) 3D chromatin-based V2G genes are enriched for CRISPR-implicated genes that regulate CD4+ T cell activation. Observed enrichment of genes regulating multiple aspects of CD4+ T cell activation (IL-2, IL-2 receptor, CTLA-4, IFNg, TNFa or proliferation) from CRISPR screens by Freimer, Schmidt, and Shifrut among sets of 3D chromatin-implicated genes among individual diseases (green, FDR<0.05) or all diseases (purple, FDR<0.05). (**D**) Enrichment for 3D chromatin-based autoimmune V2G genes among genes with germline knock-out mouse immune (red) and other (black) phenotypes (adjP<0.05, IMPC database). (**E**) Autoimmune V2G-implicated genes with at least one pharmacologic modulator (rDGIdb). (**F**) Comparison of the number of manuscripts retrieved from PubMed related to autoimmune

disease for each V2G gene with pharmaceutical agents available (x-axis) with an immune-specific expression score computed using the sum GTEX median expression values (v8) for whole blood or spleen divided by other tissues (y-axis). Genes highlighted in red were selected for functional validation in **G**. (**G**) Dose-dependent impact of the indicated pharmacologic agents targeting the V2G-implicated genes *KISS1R, CHRNB, OXER1, GPR18, GRK6, PTK6, MAP3K11, GPR183, GART,* and *SIK1* on proliferation of anti-CD3/28 activated murine (left panel) or human (right panel) CD4+ T cells *in vitro* (N=4).

Figure 8 – Figure Supplement 1: Predicted impact of autoimmune disease-associated genetic variation at V2G-implicated loci. (A) Top 50 TF motifs impacted by autoimmune SNPs (see Table S15). HiC contact frequency matrix (heatmap), chromatin accessibility (grey), and loop calls (blue-IL19/ITGA4, red-FAIM3) at the T1D-rs3024505 locus (B) and the celiac- rs13010713 locus (C). Dashed box indicates the SNP affected by allelic variation. Genetic variation at rs3024505 is predicted to disrupt a MZF1 motif (D), and variation at rs13010713 is predicted to disrupt a SOX4 binding motif (E). (F) Variation at MS SNP rs1077667 is predicted to disrupt TP53, TP63, and POU2F2 (OCT2) binding sites.

Figure 8 – Figure Supplement 2: Orthogonal validation of 3D chromatin-based V2G genes. Expression (scaled TPM) of genes implicated by chromatin contacts in (A) prior CRISPRi/a screens and (C) genes with known immune phenotypes in the international mouse phenotyping consortium. Red = increased, blue = decreased.

Supplemental Files

Supplementary File 1: Expression and clustering of differentially expressed genes.

Supplementary File 2: Pathway enrichment of differentially expressed RNA-seq genes

Supplementary File 3: Pathway enrichment of differentially expressed RNA-seq genes by cluster

Supplementary File 4: Accessible chromatin regions

Supplementary File 5: Differential accessible open chromatin regions

Supplementary File 6: A/B compartment calls

Supplementary File 7: Differential TAD boundaries

Supplementary File 8: Stripe calls

Supplementary File 9: Differential contact frequency

Supplementary File 10: Summary of TF footprinting

Supplementary File 11: TF target gene pathway enrichment

Supplementary File 12: List of all GWAS studies included

Supplementary File 13: Variant to gene mapping across all timepoints

Supplementary File 14: Partitioned LD score regression output

Supplementary File 15: Autoimmune variants in open gene promoters

Supplementary File 16: Comparison of V2G approaches

Supplementary File 17: Motifs predicted to disrupt transcription factor binding sites.

Supplementary File 18: V2G genes implicated by orthogonal data

Supplementary File 19: V2G target gene drug repurposing results

REFERENCES

1. Zhang, F. & Lupski, J. R. Non-coding genetic variants in human disease. *Hum Mol Genet* **24**, R102-10 (2015).
2. Weinreich, M. A. *et al.* KLF2 Transcription-Factor Deficiency in T Cells Results in Unrestrained Cytokine Production and Upregulation of Bystander Chemokine Receptors. *Immunity* **31**, 122–130 (2009).
3. Tzachanis, D. *et al.* Tob is a negative regulator of activation that is expressed in anergic and quiescent T cells. *Nat Immunol* **2**, 1174–82 (2001).
4. Yusuf, I. & Fruman, D. A. Regulation of quiescence in lymphocytes. *Trends Immunol* **24**, 380–386 (2003).
5. Zhu, J., Yamane, H. & Paul, W. E. Differentiation of Effector CD4 T Cell Populations. *Annu Rev Immunol* **28**, 445–489 (2010).
6. Saravia, J., Chapman, N. M. & Chi, H. Helper T cell differentiation. *Cell Mol Immunol* **16**, 634–643 (2019).
7. Farh, K. K.-H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337 343 (2014).
8. Chen, L. *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* **167**, 1398 1414.e24 (2016).
9. Soskic, B. *et al.* Chromatin activity at GWAS loci identifies T cell states driving complex immune diseases. *Nat Genet* 1–8 (2019) doi:10.1038/s41588-019-0493-9.
10. Robertson, C. C. *et al.* Fine-mapping, trans-ancestral and genomic analyses identify causal variants, cells, genes and drug targets for type 1 diabetes. *Nat Genet* **53**, 962–971 (2021).
11. Lu, X. *et al.* Global discovery of lupus genetic risk variant allelic enhancer activity. *Nat Commun* **12**, 1611 (2021).
12. Chandra, V. *et al.* Promoter-interacting expression quantitative trait loci are enriched for functional genetic variants. *Nat Genet* **53**, 110–119 (2021).
13. Soskic, B. *et al.* Immune disease risk variants regulate gene expression dynamics during CD4+ T cell activation. *Nat Genet* 1–10 (2022) doi:10.1038/s41588-022-01066-3.
14. Mouri, K. *et al.* Prioritization of autoimmune disease-associated genetic variants that perturb regulatory element activity in T cells. *Nat Genet* 1–10 (2022) doi:10.1038/s41588-022-01056-5.
15. Rioux, J. D. & Abbas, A. K. Paths to understanding the genetic basis of autoimmune disease. *Nature* **435**, 584–589 (2005).
16. Buckner, J. H. Mechanisms of impaired regulation by CD4(+)CD25(+)FOXP3(+) regulatory T cells in human autoimmune diseases. *Nat Rev Immunol* **10**, 849 859 (2010).
17. Javierre, B. M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* **167**, 1369 1384.e19 (2016).
18. Burren, O. S. *et al.* Chromosome contacts in activated T cells identify autoimmune disease candidate genes. *Genome Biol* **18**, 165 (2017).
19. Yang, J. *et al.* Analysis of chromatin organization and gene expression in T cells identifies functional genes for rheumatoid arthritis. *Nat Commun* **11**, 4402 (2020).
20. Bykova, M., Hou, Y., Eng, C. & Cheng, F. Quantitative trait locus (xQTL) approaches identify risk genes and drug targets from human non-coding genomes. *Hum Mol Genet* **31**, R105–R113 (2022).
21. Shifrut, E. *et al.* Genome-wide CRISPR Screens in Primary Human T Cells Reveal Key Regulators of Immune Function. *Cell* **175**, 1958-1971.e15 (2018).
22. Schmidt, R. *et al.* CRISPR activation and interference screens decode stimulation responses in primary human T cells. *Science* **375**, eabj4008 (2022).

23. Freimer, J. W. *et al.* Systematic discovery and perturbation of regulatory genes in human T cells reveals the architecture of immune networks. *Nat Genet* **54**, 1133–1144 (2022).
24. Su, C., Pahl, M. C., Grant, S. F. A. & Wells, A. D. Restriction enzyme selection dictates detection range sensitivity in chromatin conformation capture-based variant-to-gene mapping approaches. *Hum Genet* (2021) doi:10.1007/s00439-021-02326-8.
25. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75 (2012).
26. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289–293 (2009).
27. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376 380 (2012).
28. Mirny, L. A., Imakaev, M. & Abdennur, N. Two major mechanisms of chromosome organization. *Curr Opin Cell Biol* **58**, 142–152 (2019).
29. Yoon, S., Chandra, A. & Vahedi, G. Stripenn detects architectural stripes from chromatin conformation data using computer vision. *Nat Commun* **13**, 1602 (2022).
30. Kim, T.-K. & Shiekhattar, R. Architectural and Functional Commonalities between Enhancers and Promoters. *Cell* **162**, 948 959 (2015).
31. Hamamoto, K. & Fukaya, T. Molecular architecture of enhancer–promoter interaction. *Curr Opin Cell Biol* **74**, 62–70 (2022).
32. Consortium, R. E. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317 330 (2015).
33. Yoon, J.-H. *et al.* Phosphorylation status determines the opposing functions of Smad2/Smad3 as STAT3 cofactors in TH17 differentiation. *Nat Commun* **6**, 7600 (2015).
34. Safford, M. *et al.* Egr-2 and Egr-3 are negative regulators of T cell activation. *Nat Immunol* **6**, 472–480 (2005).
35. Schubert, L. A., Jeffery, E., Zhang, Y., Ramsdell, F. & Ziegler, S. F. Scurfin (FOXP3) Acts as a Repressor of Transcription and Regulates T Cell Activation. *J Biol Chem* **276**, 37672–37679 (2001).
36. Xiao, T., Li, X. & Felsenfeld, G. The Myc-associated zinc finger protein (MAZ) works together with CTCF to control cohesin positioning and genome organization. *Proc National Acad Sci* **118**, e2023127118 (2021).
37. Sebzda, E., Zou, Z., Lee, J. S., Wang, T. & Kahn, M. L. Transcription factor KLF2 regulates the migration of naive T cells by restricting chemokine receptor expression patterns. *Nat Immunol* **9**, 292–300 (2008).
38. Moskowitz, D. M. *et al.* Epigenomics of human CD8 T cell differentiation and aging. *Sci Immunol* **2**, eaag0192 (2017).
39. Harris, J. E. *et al.* Early Growth Response Gene-2, a Zinc-Finger Transcription Factor, Is Required for Full Induction of Clonal Anergy in CD4+ T Cells. *J Immunol* **173**, 7331–7338 (2004).
40. Williams, J. B. *et al.* The EGR2 targets LAG-3 and 4-1BB describe and regulate dysfunctional antigen-specific CD8+ T cells in the tumor microenvironment. *J Exp Med* **214**, 381–400 (2017).
41. Avitahl, N. *et al.* Ikaros Sets Thresholds for T Cell Activation and Regulates Chromosome Propagation. *Immunity* **10**, 333–343 (1999).
42. Bandyopadhyay, S. *et al.* Interleukin 2 gene transcription is regulated by Ikaros-induced changes in histone acetylation in anergic T cells. *Blood* **109**, 2878–86 (2007).
43. Thomas, R. M. *et al.* Ikaros Enforces the Costimulatory Requirement for IL2 Gene Expression and Is Required for Anergy Induction in CD4+ T Lymphocytes. *J Immunol* **179**, 7305–7315 (2007).
44. Thomas, R. M. *et al.* Ikaros Silences T-bet Expression and Interferon-γ Production during T Helper 2 Differentiation. *J Biol Chem* **285**, 2545–2553 (2010).

45. Weber, B. N. *et al.* A critical role for TCF-1 in T-lineage specification and differentiation. *Nature* **476**, 63–68 (2011).
46. Johnson, J. L. *et al.* Lineage-Determining Transcription Factor TCF-1 Initiates the Epigenetic Identity of T Cells. *Immunity* **48**, 243-257.e10 (2018).
47. Wang, W. *et al.* TCF-1 promotes chromatin interactions across topologically associating domains in T cell progenitors. *Nat Immunol* 1–11 (2022) doi:10.1038/s41590-022-01232-z.
48. Furukawa, T. *et al.* Involvement of PLAGL2 in activation of iron deficient- and hypoxia-induced gene expression in mouse cell lines. *Oncogene* **20**, 4718–4727 (2001).
49. Wu, Q. *et al.* Hypoxia-inducible factors: master regulators of hypoxic tumor immune escape. *J Hematol Oncol* **15**, 77 (2022).
50. Fu, W. *et al.* A multiply redundant genetic switch “locks in” the transcriptional signature of regulatory T cells. *Nat Immunol* **13**, 972–980 (2012).
51. Pramanik, J. *et al.* Genome-wide analyses reveal the IRE1a-XBP1 pathway promotes T helper cell differentiation by resolving secretory stress and accelerating proliferation. *Genome Med* **10**, 76 (2018).
52. Gao, X., Tewari, K., Svaren, J. & Suresh, M. Role of cell cycle regulator E2F1 in regulating CD8 T cell responses during acute and chronic viral infection. *Virology* **324**, 567–576 (2004).
53. DeRyckere, D. & DeGregori, J. E2F1 and E2F2 Are Differentially Required for Homeostasis-Driven and Antigen-Induced T Cell Proliferation In Vivo. *J Immunol* **175**, 647–655 (2005).
54. Cook, M. E., Jarjour, N. N., Lin, C.-C. & Edelson, B. T. Transcription Factor Bhlhe40 in Immunity and Autoimmunity. *Trends Immunol* **41**, 1023–1036 (2020).
55. Chesi, A. *et al.* Genome-scale Capture C promoter interactions implicate effector genes at GWAS loci for bone mineral density. *Nat Commun* **10**, 1260 (2019).
56. Su, C. *et al.* Mapping effector genes at lupus GWAS loci using promoter Capture-C in follicular helper T cells. *Nat Commun* **11**, 3294 (2020).
57. Pahl, M. C. *et al.* Cis-regulatory architecture of human ESC-derived hypothalamic neuron differentiation aids in variant-to-gene mapping of relevant complex traits. *Nat Commun* **12**, 6749 (2021).
58. Pahl, M. C. *et al.* Implicating effector genes at COVID-19 GWAS loci using promoter-focused Capture-C in disease-relevant immune cell types. *Genome Biol* **23**, 125 (2022).
59. Su, C. *et al.* 3D chromatin maps of the human pancreas reveal lineage-specific regulatory architecture of T2D risk. *Cell Metab* **34**, 1394-1409.e4 (2022).
60. Feng, Y. *et al.* The Cdc42-interacting Protein-4 (CIP4) Gene Knock-out Mouse Reveals Delayed and Decreased Endocytosis*. *J Biol Chem* **285**, 4348–4354 (2010).
61. Darling, N. J. & Cohen, P. Nuts and bolts of the salt-inducible kinases (SIKs). *Biochem J* **478**, 1377–1397 (2021).
62. Gate, R. E. *et al.* Genetic determinants of co-accessible chromatin regions in activated T cells across humans. *Nat Genet* **9**, 1 (2018).
63. Schmiedel, B. J. *et al.* Single-cell eQTL analysis of activated T cell subsets reveals activation and cell type–dependent effects of disease-risk variants. *Sci Immunol* **7**, eabm2508 (2022).
64. Ye, C. J. *et al.* Intersection of population variation and autoimmunity genetics in human T cell activation. *Science* **345**, 1254665 1254665 (2014).
65. Tangye, S. G. *et al.* Human Inborn Errors of Immunity: 2022 Update on the Classification from the International Union of Immunological Societies Expert Committee. *J. Clin. Immunol.* **42**, 1473–1507 (2022).
66. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature* 1–6 (2021) doi:10.1038/s41586-021-03446-x.

67. Fulco, C. P. *et al.* Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
68. Cerosaletti, K. *et al.* Multiple autoimmune-associated variants confer decreased IL-2R signaling in CD4+ CD25(hi) T cells of type 1 diabetic and multiple sclerosis patients. *Plos One* **8**, e83811 (2013).
69. Spolski, R., Li, P. & Leonard, W. J. Biology and regulation of IL-2: from molecular mechanisms to human therapy. *Nat Rev Immunol* **18**, 648 659 (2018).
70. Abbas, A. K., Trotta, E., Simeonov, D. R., Marson, A. & Bluestone, J. A. Revisiting IL-2: Biology and therapeutic prospects. *Sci Immunol* **3**, eaat1482 (2018).
71. Joosse, M. E. *et al.* Duplication of the IL2RA locus causes excessive IL-2 signaling and may predispose to very early onset colitis. *Mucosal Immunol* 1–11 (2021) doi:10.1038/s41385-021-00423-5.
72. Fujita, T., Shibuya, H., Ohashi, T., Yamanishi, K. & Taniguchi, T. Regulation of human interleukin-2 gene: Functional DNA sequences in the 5′ flanking region for the gene expression in activated T lymphocytes. *Cell* **46**, 401–407 (1986).
73. Durand, D. B., Bush, M. R., Morgan, J. G., Weiss, A. & Crabtree, G. R. A 275 basepair fragment at the 5' end of the interleukin 2 gene enhances expression from a heterologous promoter in response to signals from the T cell antigen receptor. *J Exp Medicine* **165**, 395 407 (1987).
74. Novak, T. J., White, P. M. & Rothenberg, E. V. Regulatory anatomy of the murine interleukin-2 gene. *Nucleic Acids Res* **18**, 4523–4533 (1990).
75. Mehra, P. & Wells, A. D. Long-Range Transcriptional Control of the Il2 Gene by an Intergenic Enhancer. *Mol Cell Biol* **35**, 3880–3891 (2015).
76. Consortium, T. E. P. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–640 (2004).
77. Lin, X. *et al.* Nested epistasis enhancer networks for robust genome regulation. *Science* **377**, 1077–1085 (2022).
78. Hedlund, G., Dohlsten, M., Ericsson, P. O. & Sjögren, H. O. Rapid response to Con A by CD4+CD45R− rat memory lymphocytes as compared to CD4+CD45R+ lymphocytes. *Cell Immunol* **119**, 317–326 (1989).
79. Picker, L. *et al.* Direct demonstration of cytokine synthesis heterogeneity among human memory/effector T cells by flow cytometry. *Blood* **86**, 1408–1419 (1995).
80. Sojka, D. K., Bruniquel, D., Schwartz, R. H. & Singh, N. J. IL-2 Secretion by CD4+ T Cells In Vivo Is Rapid, Transient, and Influenced by TCR-Specific Competition. *J Immunol* **172**, 6136–6143 (2004).
81. Powrie, F., Correa-Oliveira, R., Mauze, S. & Coffman, R. L. Regulatory interactions between CD45RBhigh and CD45RBlow CD4+ T cells are important for the balance between protective and pathogenic cell-mediated immunity. *J Exp Medicine* **179**, 589–600 (1994).
82. Morawski, P. A., Mehra, P., Chen, C., Bhatti, T. & Wells, A. D. Foxp3 Protein Stability Is Regulated by Cyclin-dependent Kinase 2. *J Biol Chem* **288**, 24494–24502 (2013).
83. Smigiel, K. S. *et al.* CCR7 provides localized access to IL-2 and defines homeostatically distinct regulatory T cell subsets. *J Exp Medicine* **211**, 121–136 (2014).
84. Mostafavi, H., Spence, J. P., Naqvi, S. & Pritchard, J. K. Systematic differences in discovery of genetic effects on gene expression and complex traits. *Nat. Genet.* 1–10 (2023) doi:10.1038/s41588-023-01529-1.
85. Kovanen, P. E. *et al.* T-cell development and function are modulated by dual specificity phosphatase DUSP5. *J Biological Chem* **283**, 17362–9 (2008).
86. Chuang, H.-C. & Tan, T.-H. MAP4K Family Kinases and DUSP Family Phosphatases in T-Cell Signaling and Systemic Lupus Erythematosus. *Cells* **8**, 1433 (2019).
87. Danileviciute, E. *et al.* PARK7/DJ-1 promotes pyruvate dehydrogenase activity and maintains Treg homeostasis during ageing. *Nat Metabolism* **4**, 589–607 (2022).

88. McHugh, D. *et al.* N-arachidonoyl glycine, an abundant endogenous lipid, potently drives directed cellular migration through GPR18, the putative abnormal cannabidiol receptor. *Bmc Neurosci* **11**, 44 (2010).
89. Kan, J. L., Jannatipour, M., Taylor, S. M. & Moran, R. G. Mouse cDNAs encoding a trifunctional protein of de novo purine synthesis and a related single-domain glycinamide ribonucleotide synthetase. *Gene* **137**, 195–202 (1993).
90. Varney, M. J. *et al.* G protein–coupled receptor kinase 6 (GRK6) regulates insulin processing and secretion via effects on proinsulin conversion to insulin. *J Biol Chem* **298**, 102421 (2022).
91. Gilic, M. B. & Tyner, A. L. Targeting protein tyrosine kinase 6 in cancer. *Biochimica Et Biophysica Acta Bba - Rev Cancer* **1874**, 188432 (2020).
92. Slattery, M. L., Lundgreen, A. & Wolff, R. K. MAP kinase genes and colon and rectal cancer. *Carcinogenesis* **33**, 2398–2408 (2012).
93. Knackmuss, U. *et al.* MAP3K11 is a tumor suppressor targeted by the oncomiR miR-125b in early B cells. *Cell Death Differ* **23**, 242–252 (2016).
94. Hammond, R. K. *et al.* Biological constraints on GWAS SNPs at suggestive significance thresholds reveal additional BMI loci. *Elife* **10**, e62206 (2021).
95. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).
96. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
97. Lun, A. T. L. & Smyth, G. K. csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res* **44**, e45–e45 (2016).
98. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinform Oxf Engl* **26**, 139–40 (2009).
99. Wingett, S. *et al.* HiCUP: pipeline for mapping and processing Hi-C data. *F1000research* **4**, 1310 (2015).
100. Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinform Oxf Engl* **36**, 311–316 (2019).
101. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods* **9**, 999 1003 (2012).
102. Nora, E. P. *et al.* Molecular basis of CTCF binding polarity in genome folding. *Nat Commun* **11**, 5612 (2020).
103. Cresswell, K. G. & Dozmorov, M. G. TADCompare: An R Package for Differential and Temporal Analysis of Topologically Associated Domains. *Frontiers Genetics* **11**, 158 (2020).
104. Kaul, A., Bhattacharyya, S. & Ay, F. Identifying statistically significant chromatin contacts from Hi-C data with FitHiC2. *Nat Protoc* **15**, 991–1012 (2020).
105. Stansfield, J. C., Cresswell, K. G., Vladimirov, V. I. & Dozmorov, M. G. HiCcompare: an R-package for joint normalization and comparison of HI-C datasets. *Bmc Bioinformatics* **19**, 279 (2018).
106. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinform Oxf Engl* **31**, 166–9 (2014).
107. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *Bmc Bioinformatics* **14**, 7–7 (2013).
108. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst* **1**, 417–425 (2015).
109. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47–e47 (2015).
110. Li, Z. *et al.* Identification of transcription factor binding sites using ATAC-seq. *Genome Biol* **20**, 45 (2019).

111. Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **48**, D87–D92 (2020).
112. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228–35 (2015).
113. Su, C. *et al.* 3D promoter architecture re-organization during iPSC-derived neuronal cell differentiation implicates target genes for neurodevelopmental disorders. *Prog Neurobiol* 102000 (2021) doi:10.1016/j.pneurobio.2021.102000.
114. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *Plos Comput Biol* **9**, e1003118 (2013).
115. Lopez-Delisle, L. *et al.* pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinform Oxf Engl* **37**, 422–423 (2020).
116. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187–W191 (2014).
117. Muñoz-Fuentes, V. *et al.* The International Mouse Phenotyping Consortium (IMPC): a functional catalogue of the mammalian genome that informs conservation. *Conservation Genetics Print* **19**, 995–1005 (2017).
118. Wagner, A. H. *et al.* DGIdb 2.0: mining clinically relevant drug–gene interactions. *Nucleic Acids Res* **44**, D1036–D1044 (2016).
119. Consortium, T. Gte. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
120. Wells, A. D., Gudmundsdottir, H. & Turka, L. A. Following the fate of individual T cells throughout activation and clonal expansion. Signals from T cell receptor and CD28 differentially regulate the induction and duration of a proliferative response. *J Clin Invest* **100**, 3173 3183 (1997).

Figure 1

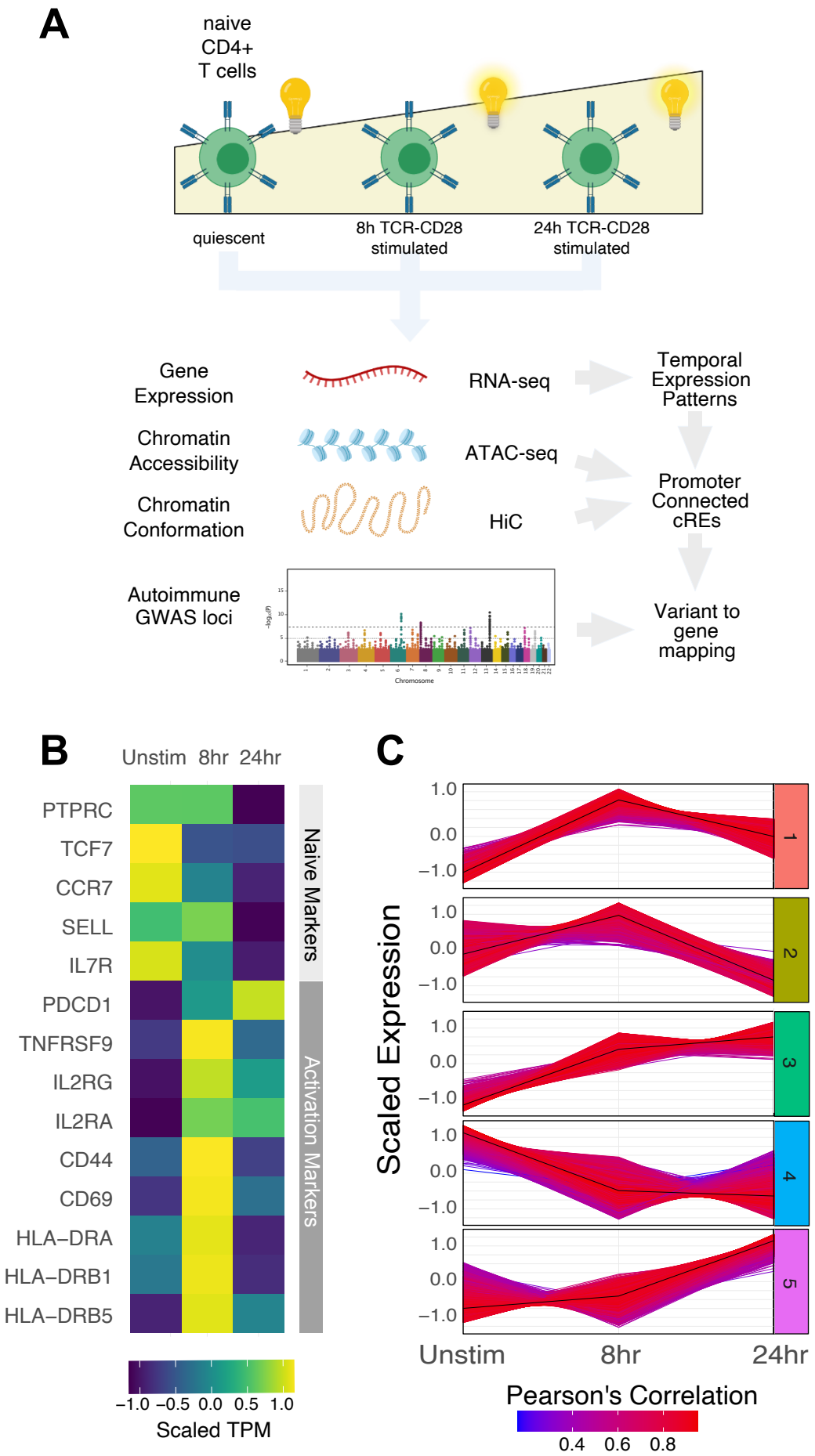

A
naive
CD4+
T cells
quiescent
8h TCR-CD28
stimulated
24h TCR-CD28
stimulated
Gene
Expression
RNA-seq
Temporal
Expression
Patterns
Chromatin
Accessibility
ATAC-seq
Chromatin
Conformation
HiC
Promoter
Connected
cREs
Autoimmune
GWAS loci
Chromosome
Variant to
gene
mapping
B
Unstim 8hr 24hr
PTPRC
TCF7
CCR7
SELL
IL7R
PDCD1
TNFRSF9
IL2RG
IL2RA
CD44
CD69
HLA–DRA
HLA–DRB1
HLA–DRB5
Naive Markers
Activation Markers
–1.0 –0.5 0.0 0.5 1.0
Scaled TPM
C
Scaled Expression
1.0
0.0
–1.0
1
2
3
4
5
Unstim
8hr
24hr
Pearson's Correlation
0.4 0.6 0.8

Figure 2

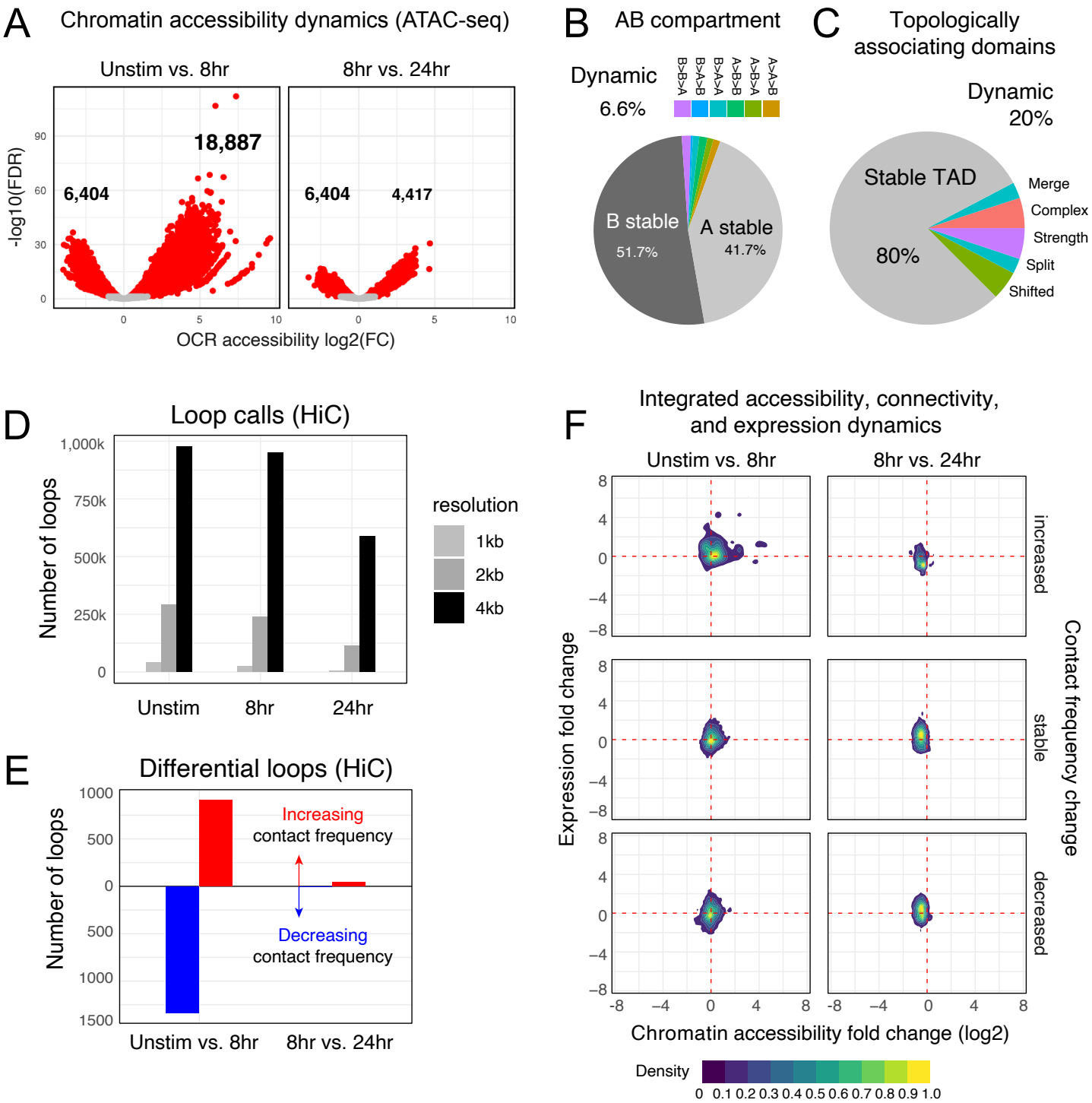

A
Chromatin accessibility dynamics (ATAC-seq)
Unstim vs. 8hr
8hr vs. 24hr
-log10(FDR)
18,887
6,404
6,404
4,417
OCR accessibility log2(FC)
B
AB compartment
Dynamic
6.6%
B>B>A
B<A>B
B>A>A
A>B>B
A<B>A
A>A>B
B stable
51.7%
A stable
41.7%
C
Topologically associating domains
Dynamic
20%
Stable TAD
80%
Merge
Complex
Strength
Split
Shifted
D
Loop calls (HiC)
Number of loops
resolution
1kb
2kb
4kb
Unstim
8hr
24hr
E
Differential loops (HiC)
Number of loops
Increasing contact frequency
Decreasing contact frequency
Unstim vs. 8hr
8hr vs. 24hr
F
Integrated accessibility, connectivity, and expression dynamics
Unstim vs. 8hr
8hr vs. 24hr
Expression fold change
Contact frequency change
increased
stable
decreased
Chromatin accessibility fold change (log2)
Density

Figure 3

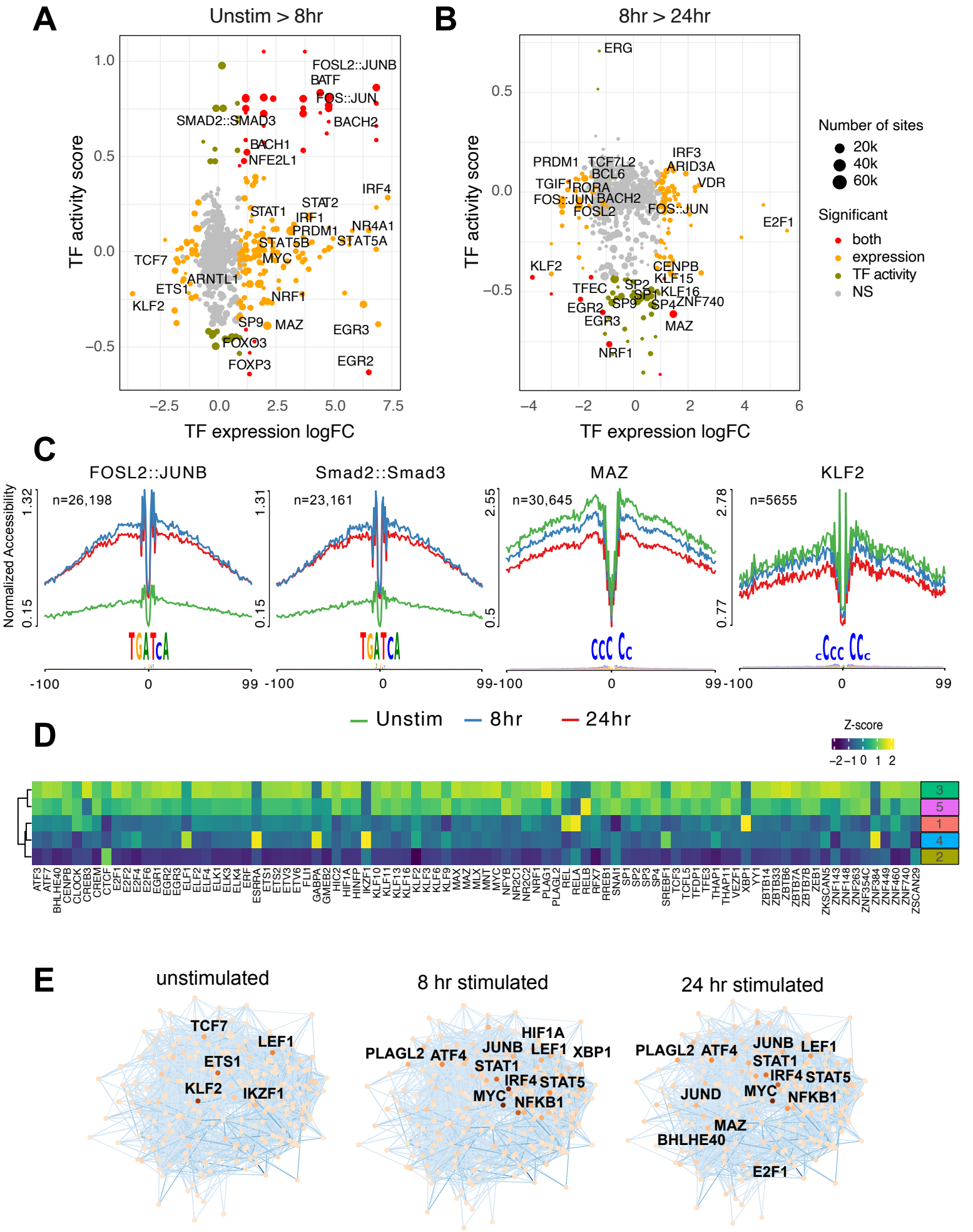

A
Unstim > 8hr
TF activity score
TF expression logFC
FOSL2::JUNB
BATF
FOS::JUN
BACH2
SMAD2::SMAD3
BACH1
NFE2L1
IRF4
STAT2
STAT1
IRF1
PRDM1
NR4A1
STAT5B
STAT5A
TCF7
MYC
ARNTL1
ETS1
NRF1
KLF2
SP9
MAZ
EGR3
FOXO3
FOXP3
EGR2
B
8hr > 24hr
ERG
PRDM1
TCF7L2
BCL6
IRF3
ARID3A
VDR
TGIF1
RORA
FOS::JUN
BACH2
FOSL2
FOS::JUN
E2F1
KLF2
CENPB
KLF15
TFEC
SP2
SP1
KLF16
ZNF740
EGR2
SP9
SP4
EGR3
MAZ
NRF1
Number of sites
20k
40k
60k
Significant
both
expression
TF activity
NS
C
FOSL2::JUNB
Smad2::Smad3
MAZ
KLF2
Normalized Accessibility
n=26,198
n=23,161
n=30,645
n=5655
Unstim
8hr
24hr
D
Z-score
E
unstimulated
8 hr stimulated
24 hr stimulated
TCF7
LEF1
ETS1
KLF2
IKZF1
HIF1A
PLAGL2
ATF4
JUNB
LEF1
XBP1
STAT1
IRF4
STAT5
MYC
NFKB1
JUND
MAZ
BHLHE40
E2F1

Figure 4

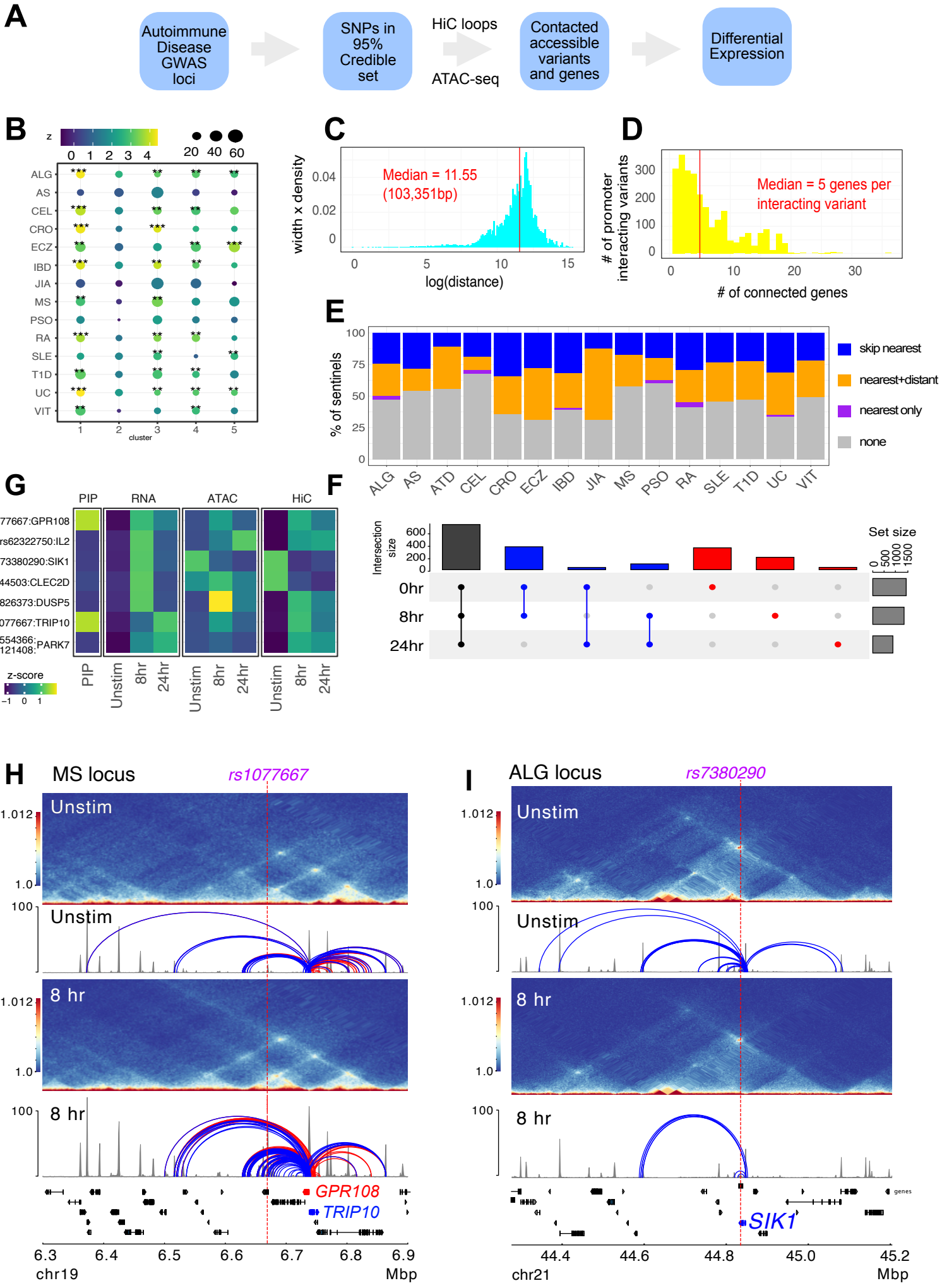

A
Autoimmune Disease GWAS loci
SNPs in 95% Credible set
HiC loops
ATAC-seq
Contacted accessible variants and genes
Differential Expression
B
z
0 1 2 3 4
20 40 60
ALG
AS
CEL
CRO
ECZ
IBD
JIA
MS
PSO
RA
SLE
T1D
UC
VIT
1 2 3 4 5
cluster
C
width x density
Median = 11.55 (103,351bp)
log(distance)
D
# of promoter interacting variants
Median = 5 genes per interacting variant
# of connected genes
E
% of sentinels
skip nearest
nearest+distant
nearest only
none
ALG AS ATD CEL CRO ECZ IBD JIA MS PSO RA SLE T1D UC VIT
F
Intersection size
Set size
0hr
8hr
24hr
G
PIP
RNA
ATAC
HiC
rs1077667:GPR108
rs62322750:IL2
rs73380290:SIK1
rs10844503:CLEC2D
rs180826373:DUSP5
rs1077667:TRIP10
rs35554366: rs12121408:PARK7
z-score
PIP
Unstim
8hr
24hr
H
MS locus
rs1077667
Unstim
8 hr
GPR108
TRIP10
chr19
Mbp
I
ALG locus
rs7380290
Unstim
8 hr
SIK1
chr21
Mbp

Figure 5

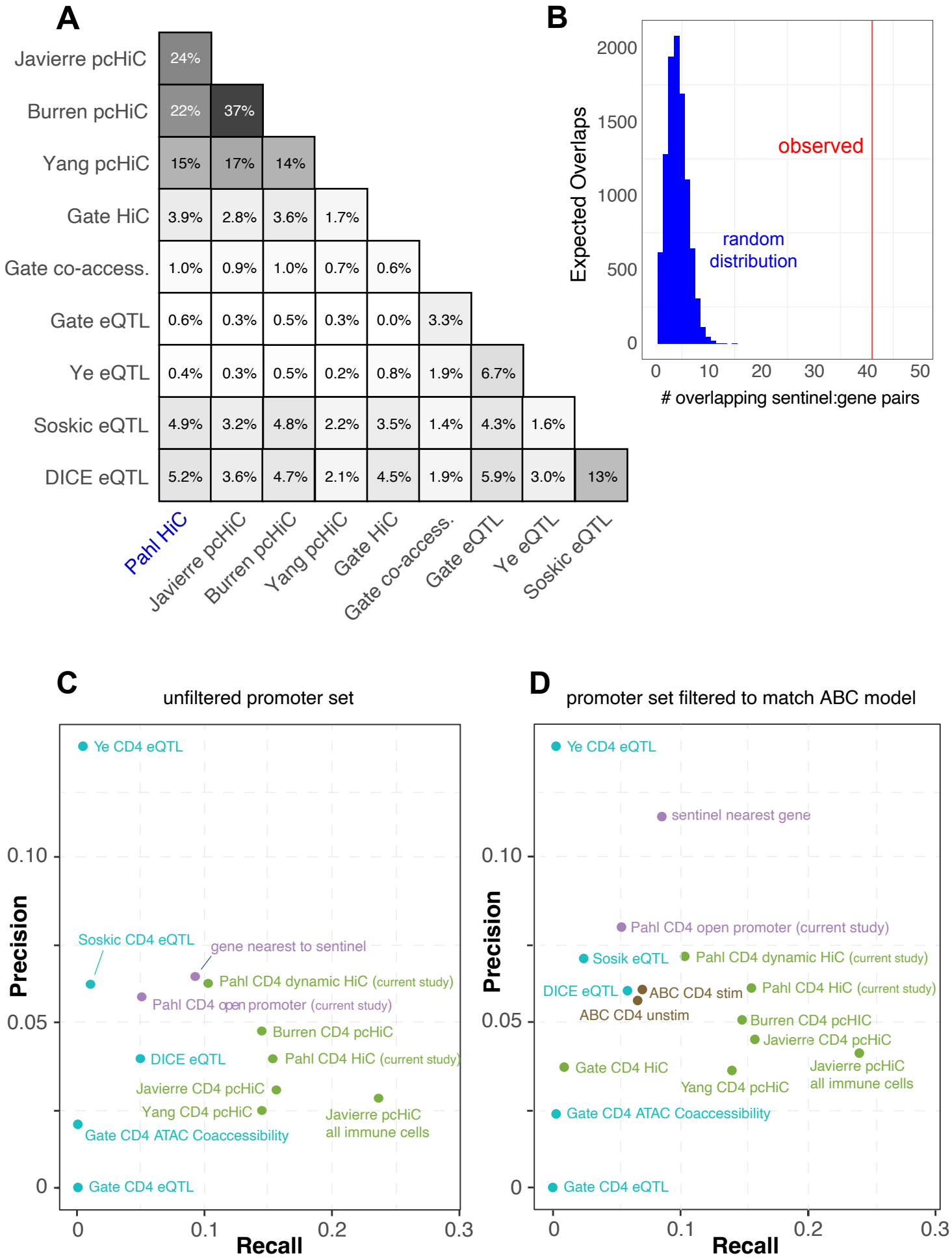

A
Javierre pcHiC
Burren pcHiC
Yang pcHiC
Gate HiC
Gate co-access.
Gate eQTL
Ye eQTL
Soskic eQTL
DICE eQTL
Pahl HiC
24%
22% 37%
15% 17% 14%
3.9% 2.8% 3.6% 1.7%
1.0% 0.9% 1.0% 0.7% 0.6%
0.6% 0.3% 0.5% 0.3% 0.0% 3.3%
0.4% 0.3% 0.5% 0.2% 0.8% 1.9% 6.7%
4.9% 3.2% 4.8% 2.2% 3.5% 1.4% 4.3% 1.6%
5.2% 3.6% 4.7% 2.1% 4.5% 1.9% 5.9% 3.0% 13%
B
Expected Overlaps
observed
random distribution
overlapping sentinel:gene pairs
C
unfiltered promoter set
Precision
Recall
Ye CD4 eQTL
Soskic CD4 eQTL
gene nearest to sentinel
Pahl CD4 dynamic HiC (current study)
Pahl CD4 open promoter (current study)
Burren CD4 pcHiC
DICE eQTL
Pahl CD4 HiC (current study)
Javierre CD4 pcHiC
Yang CD4 pcHiC
Javierre pcHiC all immune cells
Gate CD4 ATAC Coaccessibility
Gate CD4 eQTL
D
promoter set filtered to match ABC model
Precision
Recall
Ye CD4 eQTL
sentinel nearest gene
Pahl CD4 open promoter (current study)
Sosik eQTL
Pahl CD4 dynamic HiC (current study)
DICE eQTL
ABC CD4 stim
Pahl CD4 HiC (current study)
ABC CD4 unstim
Burren CD4 pcHIC
Javierre CD4 pcHiC
Gate CD4 HiC
Yang CD4 pcHiC
Javierre pcHiC all immune cells
Gate CD4 ATAC Coaccessibility
Gate CD4 eQTL

Figure 6

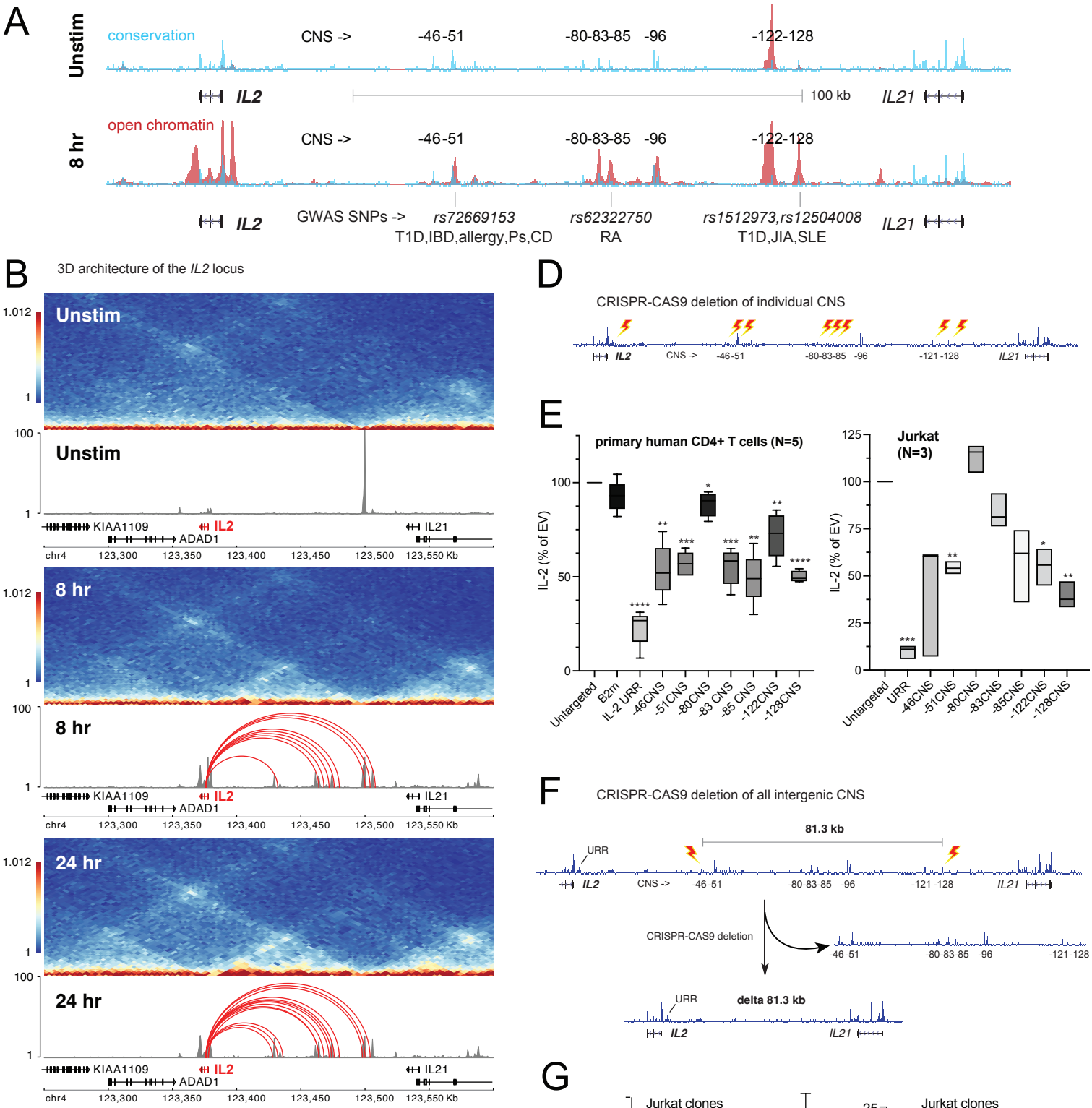

A
Unstim
conservation
CNS ->
-46-51
-80-83-85
-96
-122-128
IL2
100 kb
IL21
8 hr
open chromatin
GWAS SNPs ->
rs72669153
T1D,IBD,allergy,Ps,CD
rs62322750
RA
rs1512973,rs12504008
T1D,JIA,SLE
B
3D architecture of the IL2 locus
Unstim
8 hr
24 hr
KIAA1109
ADAD1
IL21
chr4
123,300
123,350
123,400
123,450
123,500
123,550 Kb
C
reporter assay for IL2 CNS enhancer activity
Luciferase activity (fold over pGL4)
Luciferase activity (fold over URR)
PR
URR
-46
-51
-80
-83
-128
D
CRISPR-CAS9 deletion of individual CNS
E
primary human CD4+ T cells (N=5)
Jurkat (N=3)
IL-2 (% of EV)
Untargeted
B2m
IL-2 URR
-46CNS
-51CNS
-80CNS
-83 CNS
-85 CNS
-122CNS
-128CNS
F
CRISPR-CAS9 deletion of all intergenic CNS
81.3 kb
URR
CRISPR-CAS9 deletion
delta 81.3 kb
G
Jurkat clones (N=3)
Empty vector
IL2-21 del81.3 kb
IL-2 (pg/mL)
PMA+I+a-CD28 stim (hrs)
IL2 transcript (normalized to 18s)
EV
del81.3kb
PMA+I+a-CD28 6h

Figure 7

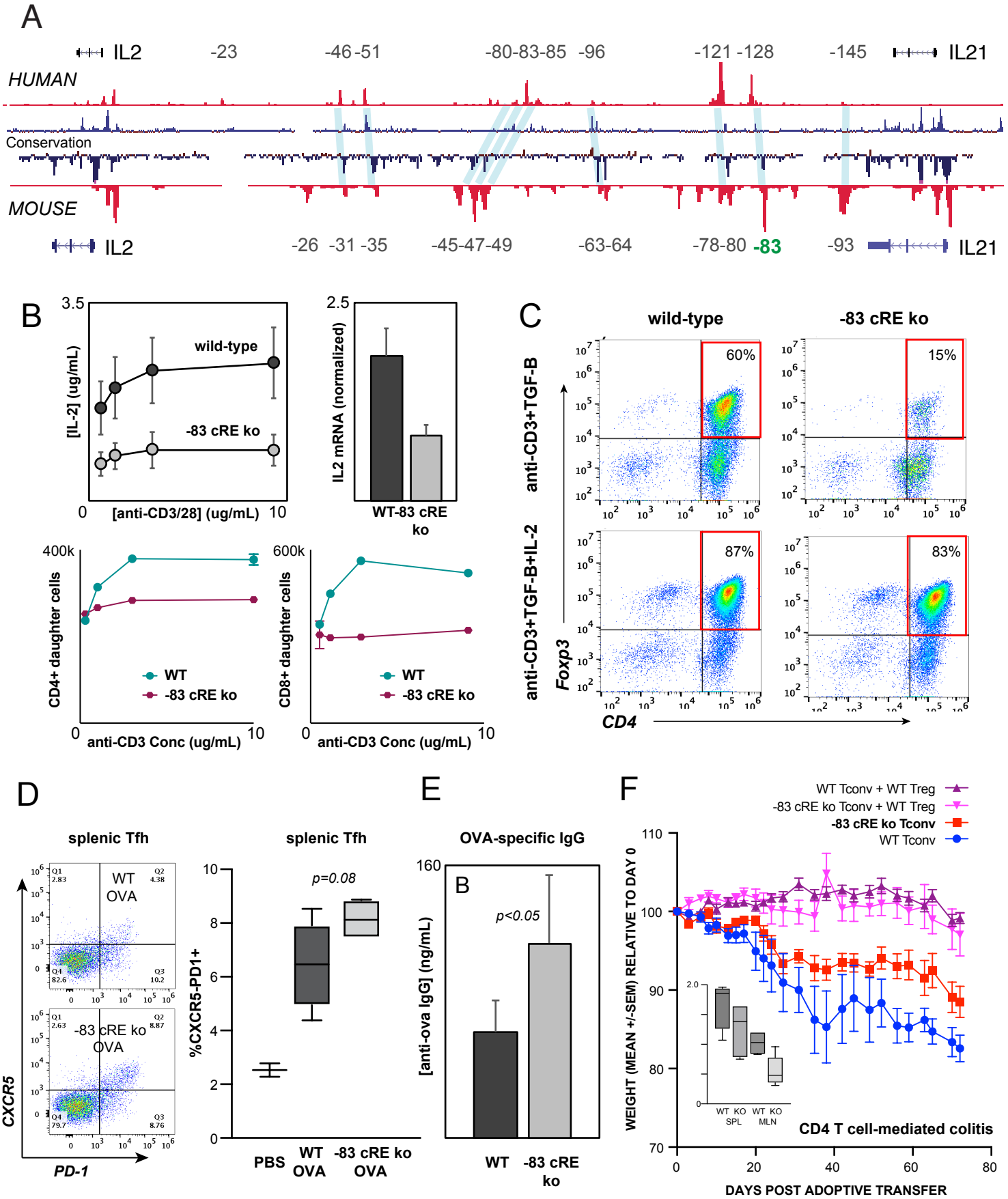

A
IL2
-23
-46 -51
-80-83-85 -96
-121 -128
-145
IL21
HUMAN
Conservation
MOUSE
IL2
-26 -31 -35
-45-47-49
-63-64
-78-80 -83
-93
IL21
B
[IL-2] (ug/mL)
wild-type
-83 cRE ko
[anti-CD3/28] (ug/mL)
IL2 mRNA (normalized)
WT-83 cRE ko
CD4+ daughter cells
CD8+ daughter cells
WT
-83 cRE ko
anti-CD3 Conc (ug/mL)
C
wild-type
-83 cRE ko
anti-CD3+TGF-B
anti-CD3+TGF-B+IL-2
Foxp3
CD4
60%
15%
87%
83%
D
splenic Tfh
WT OVA
-83 cRE ko OVA
CXCR5
PD-1
%CXCR5-PD1+
p=0.08
PBS
WT OVA
-83 cRE ko OVA
E
OVA-specific IgG
[anti-ova IgG] (ng/mL)
p<0.05
WT
-83 cRE ko
F
WT Tconv + WT Treg
-83 cRE ko Tconv + WT Treg
-83 cRE ko Tconv
WT Tconv
WEIGHT (MEAN +/-SEM) RELATIVE TO DAY 0
WT KO SPL
WT KO MLN
CD4 T cell-mediated colitis
DAYS POST ADOPTIVE TRANSFER

Figure 8

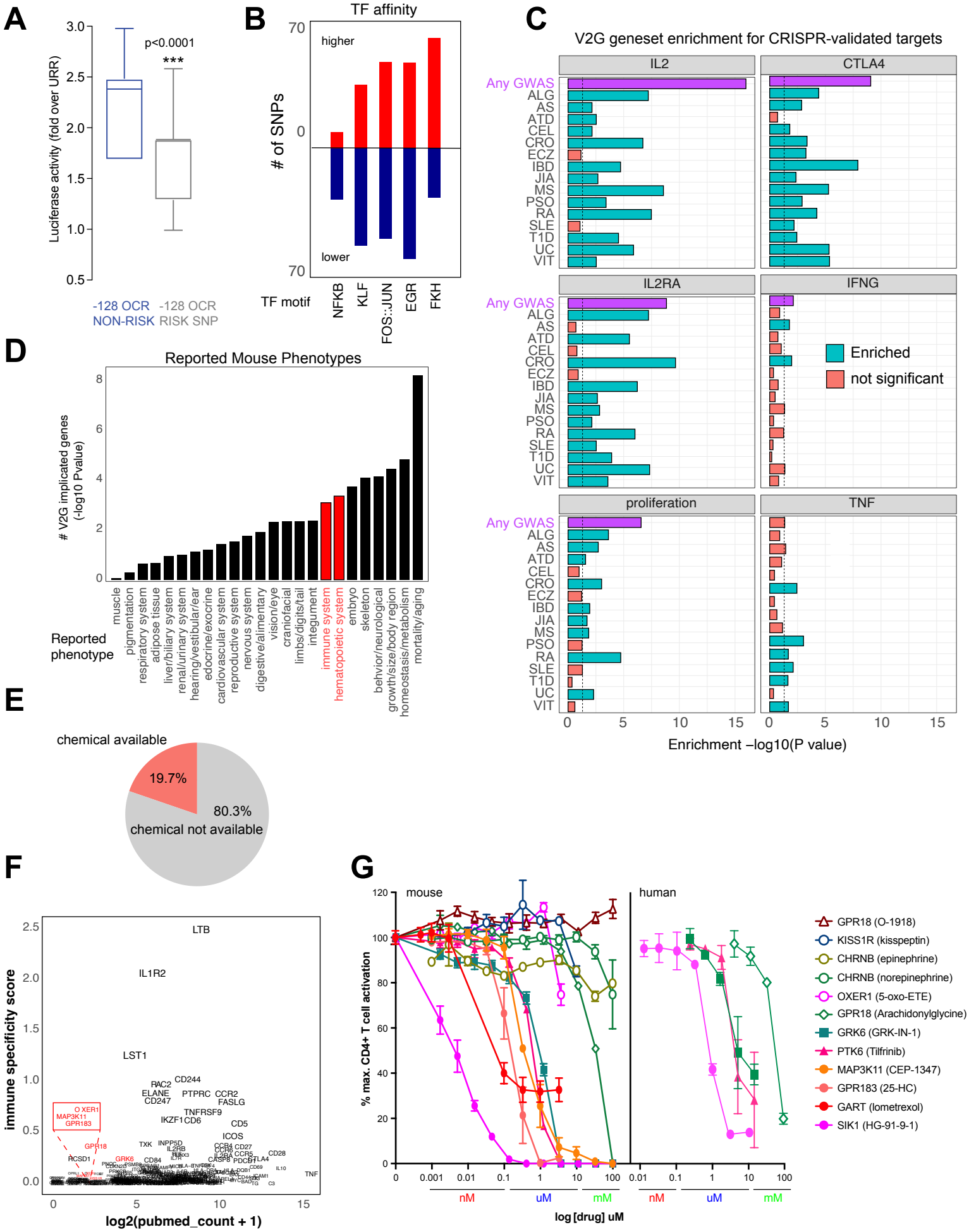

A
Luciferase activity (fold over URR)
p<0.0001
***
-128 OCR NON-RISK
-128 OCR RISK SNP
B
TF affinity
higher
lower
# of SNPs
TF motif
NFKB
KLF
FOS::JUN
EGR
FKH
C
V2G geneset enrichment for CRISPR-validated targets
IL2
CTLA4
IL2RA
IFNG
proliferation
TNF
Any GWAS
ALG
AS
ATD
CEL
CRO
ECZ
IBD
JIA
MS
PSO
RA
SLE
T1D
UC
VIT
Enriched
not significant
Enrichment –log10(P value)
D
Reported Mouse Phenotypes
V2G implicated genes (-log10 Pvalue)
Reported phenotype
muscle
pigmentation
respiratory system
adipose tissue
liver/biliary system
renal/urinary system
hearing/vestibular/ear
edocrine/exocrine
cardiovascular system
reproductive system
nervous system
digestive/alimentary
vision/eye
craniofacial
limbs/digits/tail
integument
immune system
hematopoietic system
embryo
skeleton
behvior/neurological
growth/size/body region
homeostasis/metabolism
mortality/aging
E
chemical available
19.7%
80.3%
chemical not available
F
immune specificity score
log2(pubmed_count + 1)
LTB
IL1R2
LST1
CD244
RAC2
ELANE
CD247
PTPRC
CCR2
FASLG
TNFRSF9
IKZF1
CD6
CD5
ICOS
O XER1
MAP3K11
GPR183
GPR18
GRK6
CD84
CD28
TNF
G
mouse
human
% max. CD4+ T cell activation
log [drug] uM
nM
uM
mM
GPR18 (O-1918)
KISS1R (kisspeptin)
CHRNB (epinephrine)
CHRNB (norepinephrine)
OXER1 (5-oxo-ETE)
GPR18 (Arachidonylglycine)
GRK6 (GRK-IN-1)
PTK6 (Tilfrinib)
MAP3K11 (CEP-1347)
GPR183 (25-HC)
GART (lometrexol)
SIK1 (HG-91-9-1)

Figure 1 - Figure Supplement 1

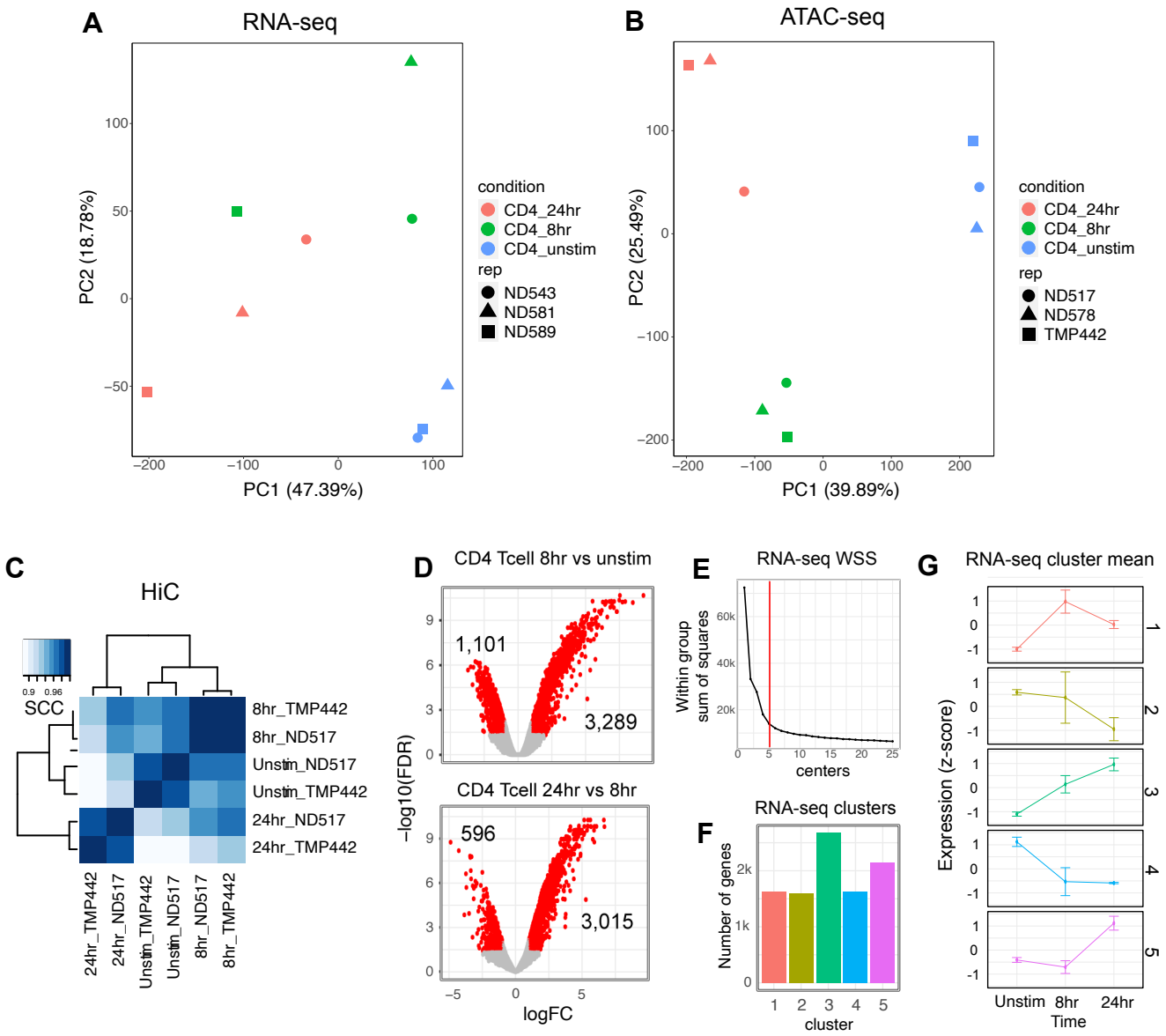


Figure 2 - Figure Supplement 1

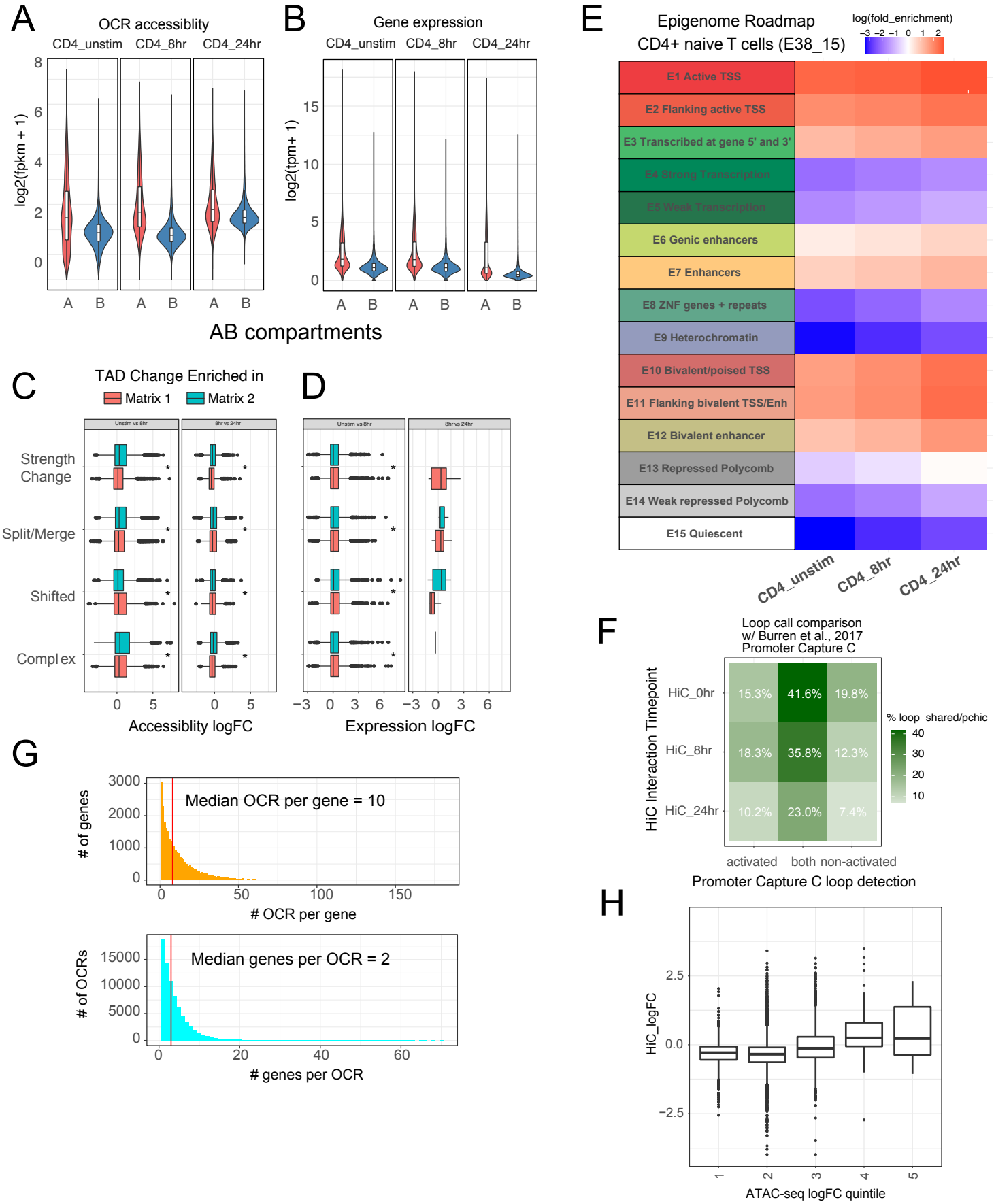


Figure 2 - Figure Supplement 2

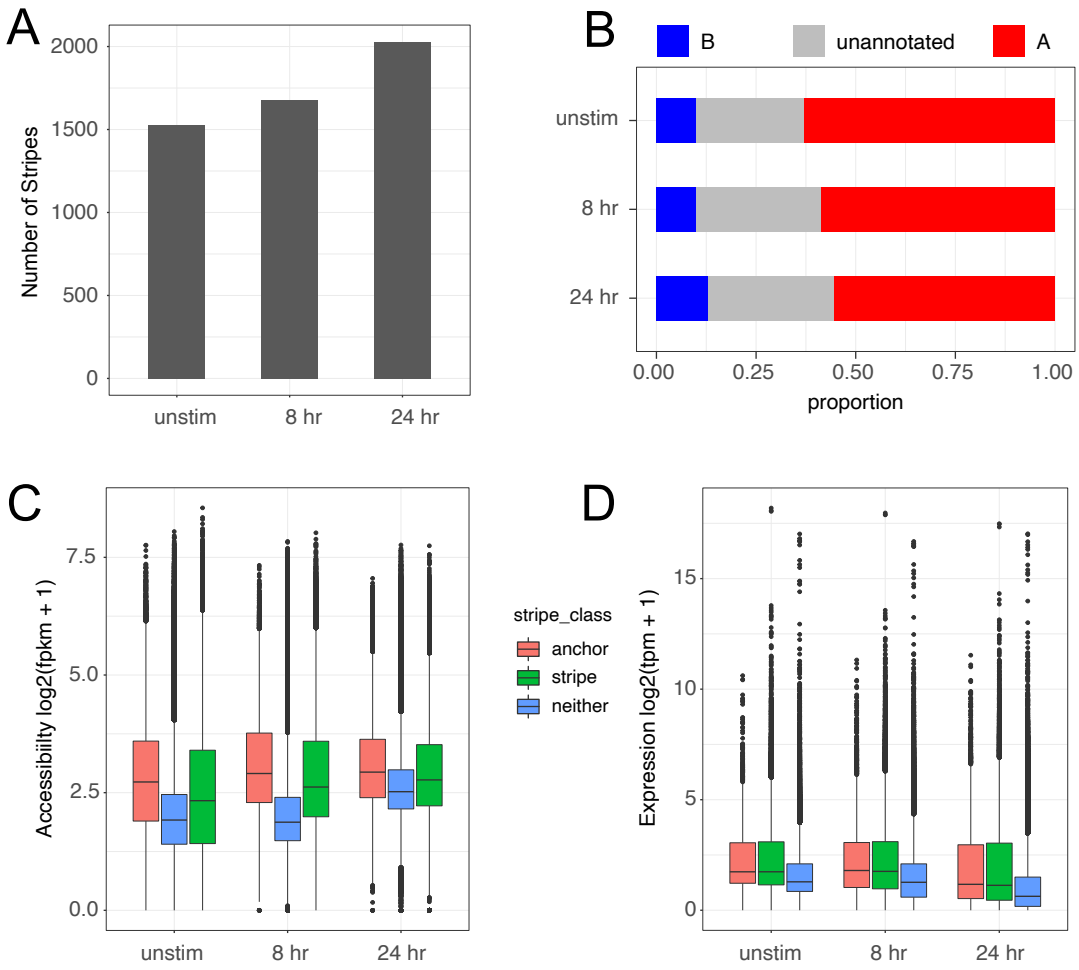
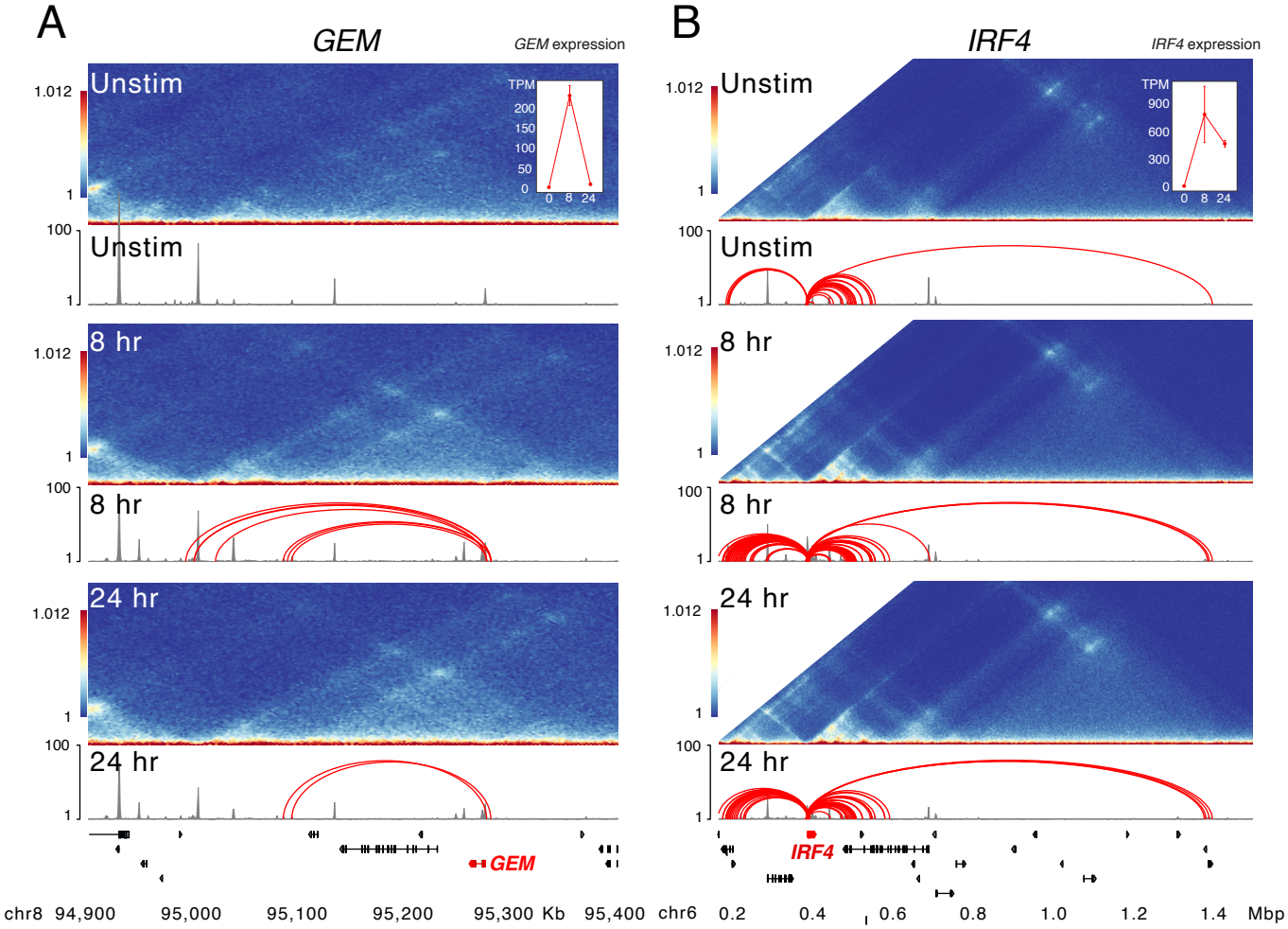


Figure 2 - Figure Supplement 3



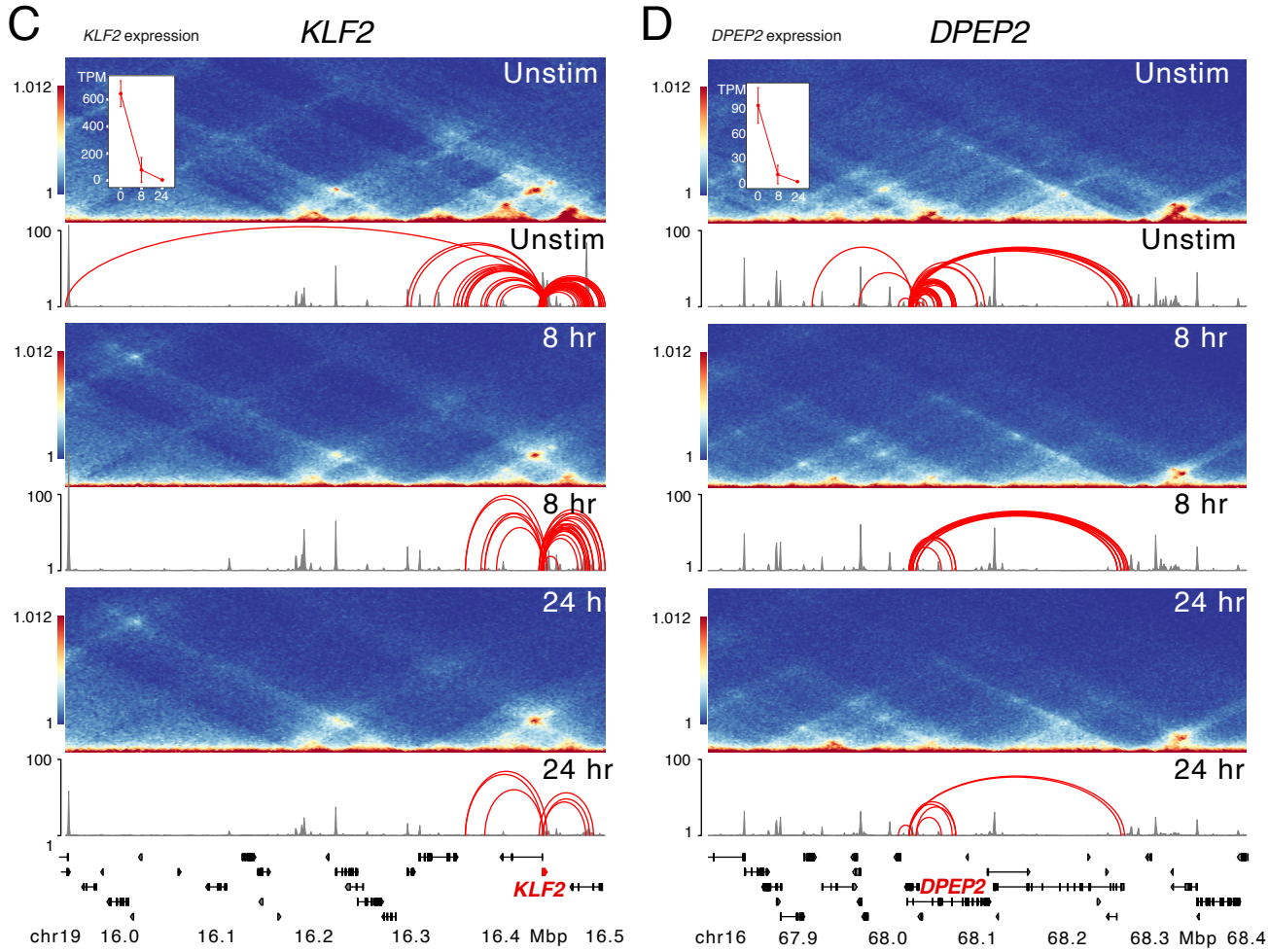


Figure 3 - Figure Supplement 1

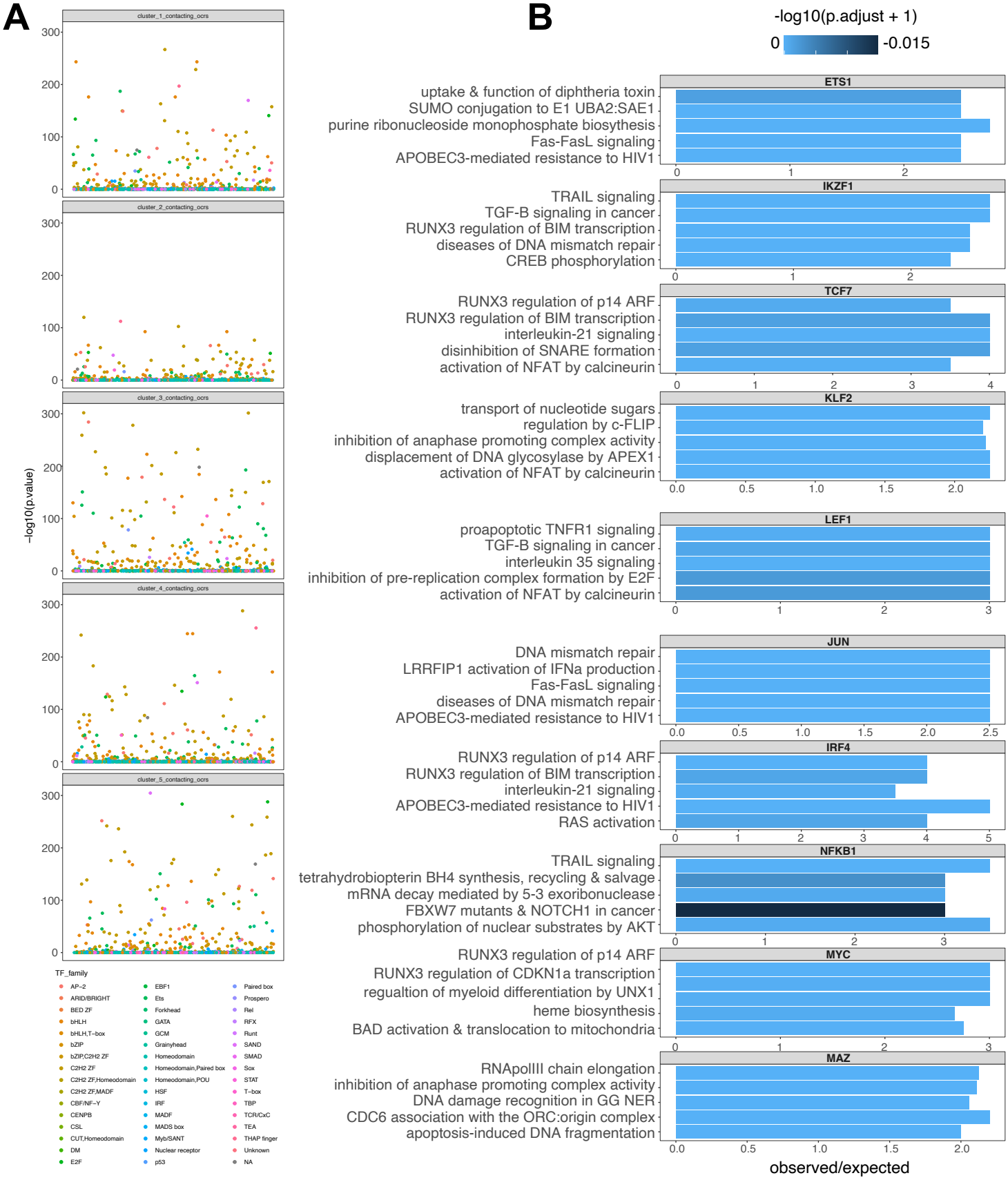


Figure 6 - Figure Supplement 1

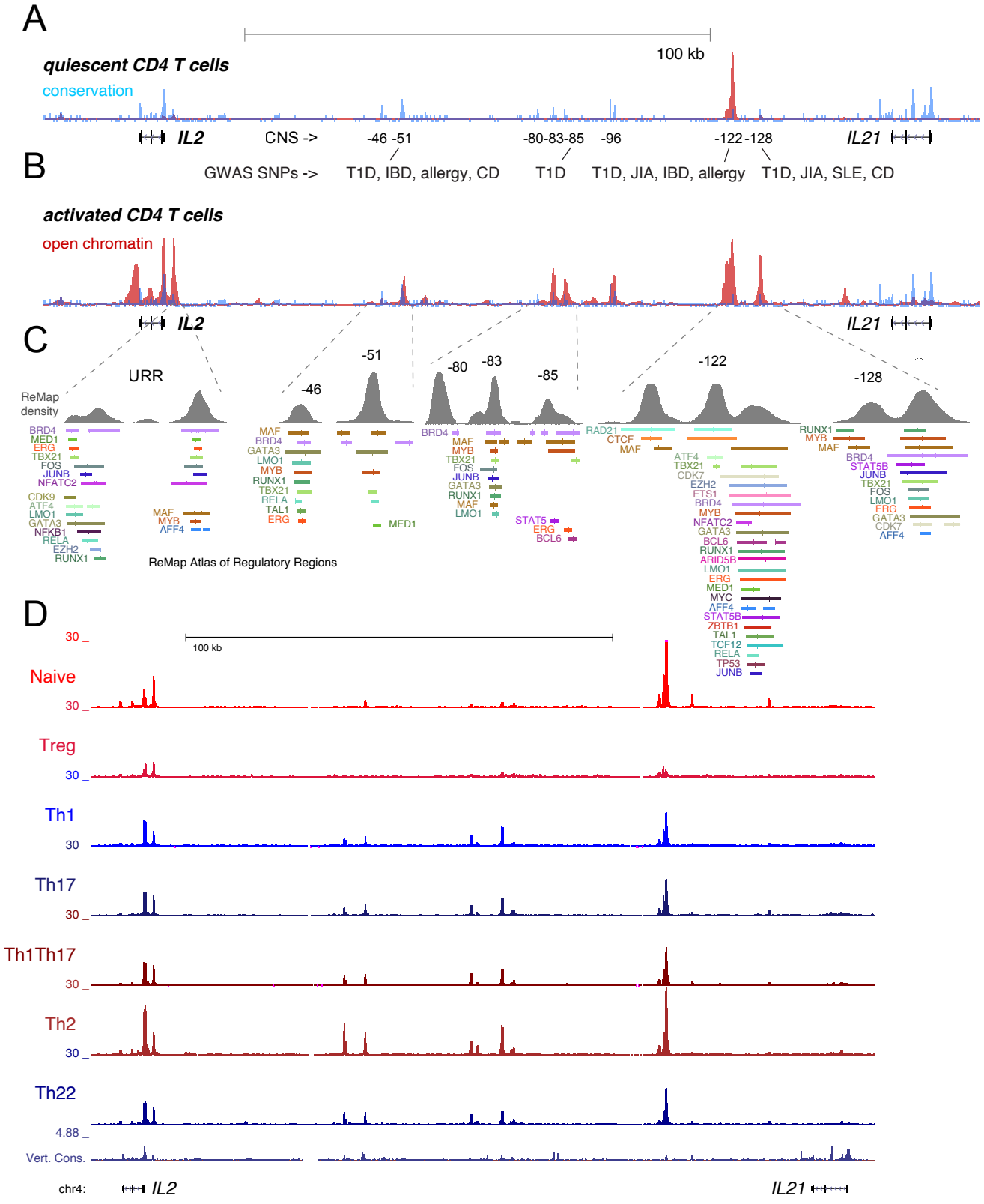


Figure 8 - Figure Supplement 1

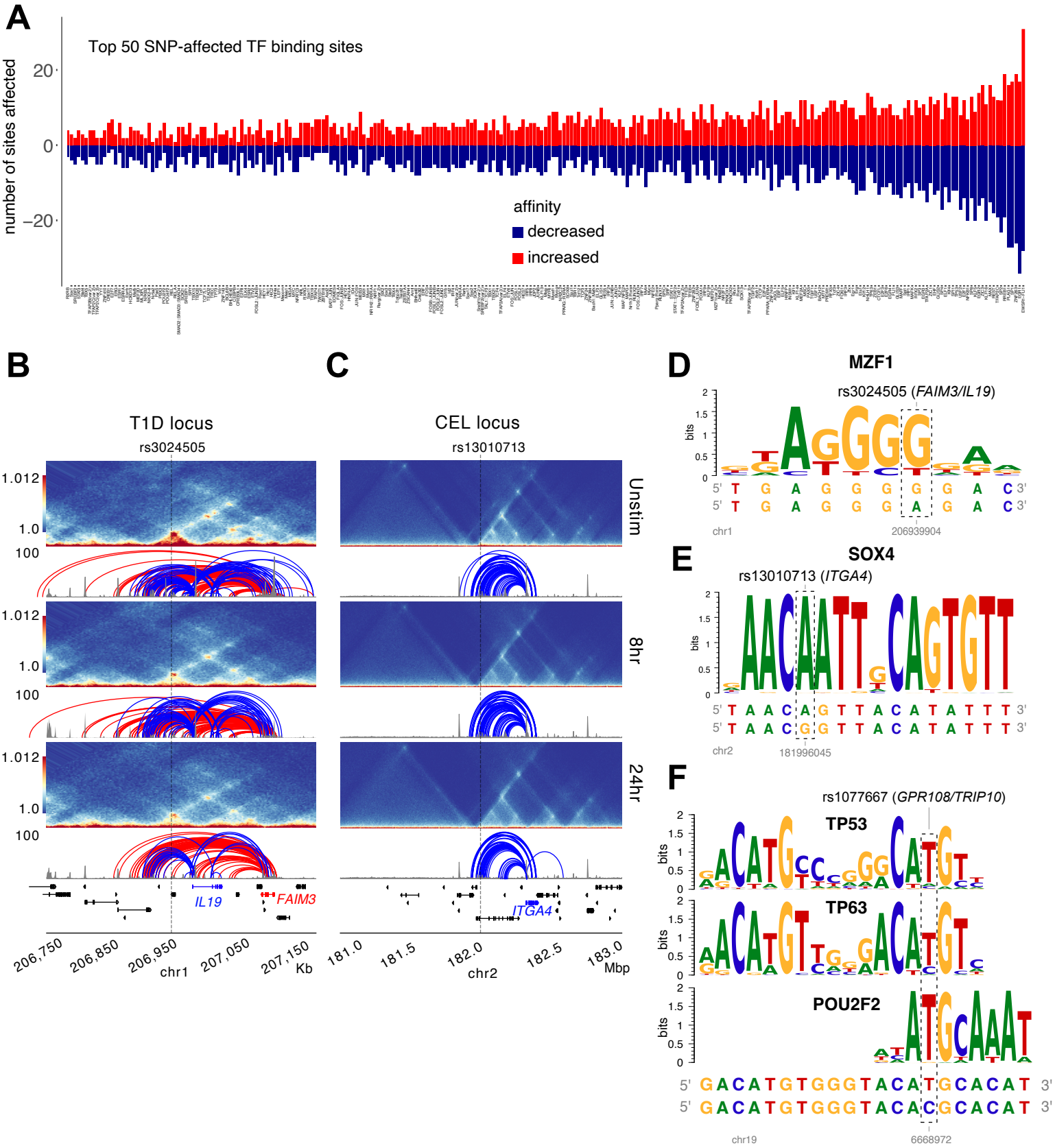
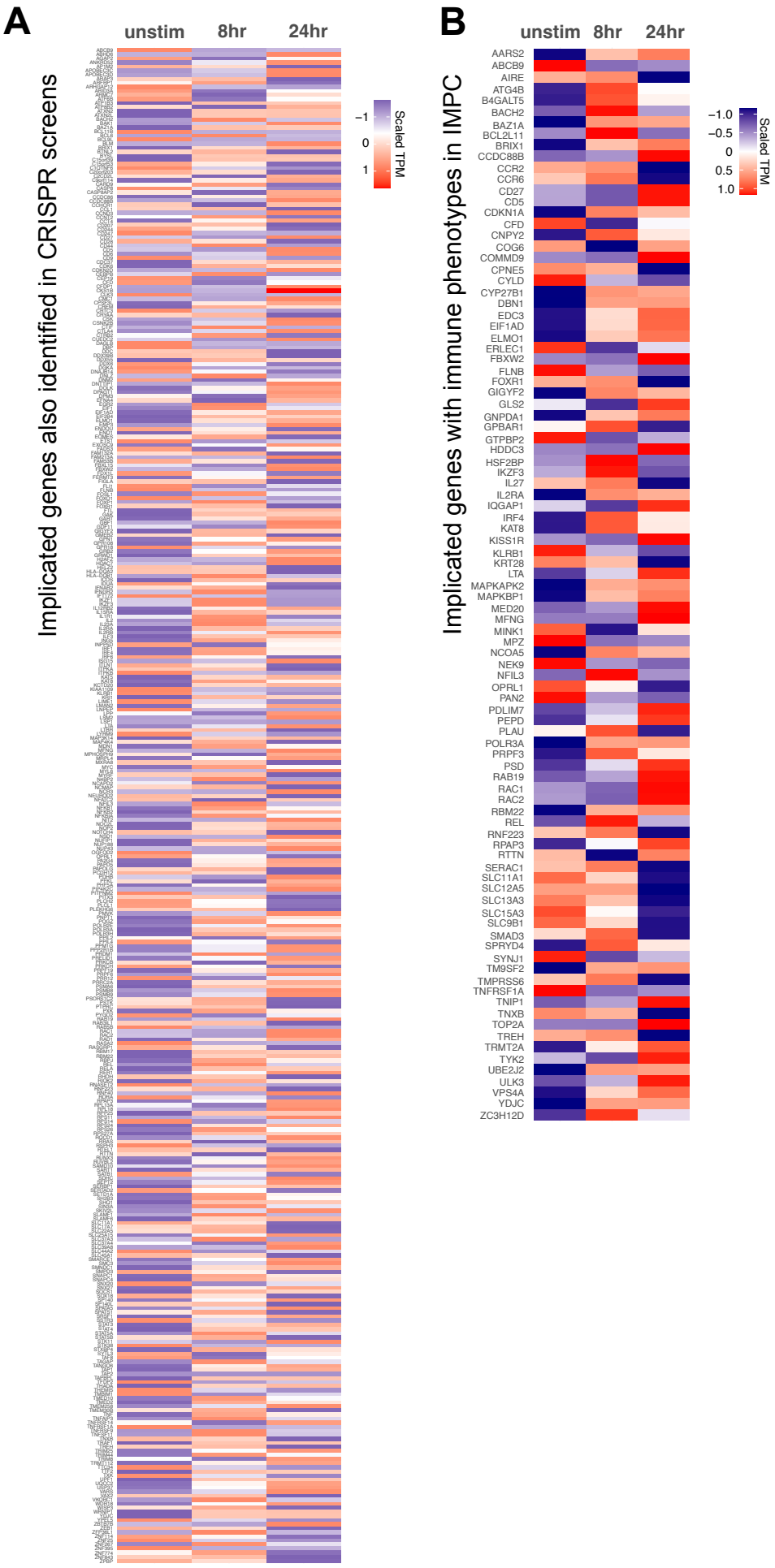



Figure 8 - Figure Supplement 2