Chromatin conformation dynamics during CD4+ T cell activation implicates autoimmune disease-associated genes and regulatory elements

Matthew C. Pahl*1,2, Prabhat Sharma*1,3, Rajan M. Thomas1,3, Zachary Thompson1,3, Zachary Mount1,3, James Pippin1,2, Peter A. Morawski4, Chun Su1,2, Daniel J. Campbell4,5, Struan F.A. Grant1,2,6,7,8, Andrew D. Wells1,3,9,10

1Center for Spatial and Functional Genomics, The Children’s Hospital of Philadelphia
2Division of Human Genetics, The Children’s Hospital of Philadelphia
3Department of Pathology, The Children’s Hospital of Philadelphia
4Benaroya Research Institute at Virginia Mason, Seattle, WA
5Department of Immunology, University of Washington School of Medicine, Seattle, WA
6Department of Genetics, Perelman School of Medicine, University of Pennsylvania
7Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania
8Division of Endocrinology and Diabetes, The Children’s Hospital of Philadelphia
9Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania
10Institute 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: wells@a.chop.edu
Phone: 215-590-8710
ABSTRACT

Genome-wide association studies (GWAS) have identified hundreds of genetic loci associated with autoimmune disease, but the causal variants and their contribution to immune dysregulation remain largely unknown. We measured the dynamics of chromosome conformation, chromatin accessibility, and gene transcription across three phases of naive human CD4+ T cell activation and co-localized active cis-regulatory elements with the 95% credible set of variants from 15 autoimmune GWAS. Reorganization of chromosome structure placed these elements in direct contact with ~1,200 protein-coding genes, at least one-third of which were dynamically regulated by stimulation. The set of implicated genes is enriched for high-throughput CRISPR screen targets that control multiple aspects of CD4+ T cell activation, and we pharmacologically validated eight novel gene products as potent regulators of T cell proliferation. These maps also allowed the identification and functional validation of a novel stretch of intergenic enhancers for the autoimmune gene IL2 whose activity is influenced by autoimmune-associated genetic variation. This study represents a powerful strategy and resource for assigning physiologic relevance to autoimmune-risk variants and identify novel genes that control T cell activation and function.
INTRODUCTION

GWAS has revealed hundreds of regions of the human genome robustly associated with 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 their corresponding effector genes molecularly responsible for increased disease risk is critical for identifying targets for downstream molecular study and therapeutic intervention. Identifying and prioritizing disease-causal variants is limited by a lack of genetic fine-mapping and functional studies that define the mechanisms of control over cRE activity and gene expression.

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 NFκB, 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 therefore 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 physical 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\(^{17,18}\). In addition to chromatin conformation approaches, multiple complementary approaches have been developed to link disease-associated variants 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\(^{19}\). 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 define the impact of TCR-CD28 activation on the autoimmune-associated cis-regulatory architecture of CD4+ helper T cells, leading to the discovery of a stretch of distal enhancers required for normal expression of the canonical T cell activation gene \(IL2\) 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 current study and by CRISPR-based screens reported in orthogonal studies\(^{20–22}\).

**RESULTS**

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

To identify the set 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 CD3+CD4+CD45RA+CD45RO- 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, ATAC-seq, and HiC (Fig. 1A). ATAC-seq and RNA-seq libraries were prepared using CD4+ T cells derived from N=3 donors, while HiC libraries were prepared from N=2 donors. 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 (Fig. S1A-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 (Fig 1B). 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 (Fig. 1C, Table S1).

To define global gene expression dynamics during the course of CD4+ T cell activation, we performed pairwise comparisons and k-means clustering across the three cell states assayed. This approach identified 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, Fig. S1D) that could be further separated into five clusters based on their distinct trajectories: upregulated early (cluster 1; n=1621 genes), downregulated late (cluster 2; n=1600 genes), monotonically increasing (cluster 3; n=2676 genes), downregulated early (cluster 4; n=1628 genes) and upregulated late (cluster 5; n=2154 genes, Fig. 1D, Fig. S1E-G, Table S2).

Genes upregulated early (cluster 1; e.g., IL2, IFNG, TNF, IL3, IL8, IL2RA, IL15RA, IL21R, ICOS, CD40LG, FASLG, MYC, FOS, FOSB, FOSL1, FOSL2, JUNB, REL, NFKBID, NFkB1, RELB, NFkbia, STAT5A, BATF, BTLA) are enriched for pathways involved in the unfolded protein response, cytokine signaling, and translation (Fig 1E; Table S3), and genes downregulated late (cluster 2) are moderately enriched for receptor tyrosine kinase signaling, cytokine signaling, and extracellular matrix organization. Genes monotonically increasing upon activation (cluster 3; e.g., TBX21, BHLHE40, LTA, IL12RB2, STAT1, SLAMF7, CCND2, CDK4, PRMT1, ICAM1, NFkB2, EZH2) 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. Genes downregulated early upon stimulation (cluster 4; e.g., KLF2, IL7R, NOG, PIK3IP1, TSC22D3, RORA, IL10RA) are enriched for pathways involved in inositol phosphate biosynthesis, neutrophil degranulation, and metabolism of nucleotides, and genes upregulated late (cluster 5; e.g., CDK2, E2F1, CDK1, CCNE1, CCNA2, PCNA, WEE1, CDC6, ORC1, MCM2) 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 (Fig 1E; Table S3).

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\(^2\), 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 (Table S4). 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 (Fig. 2A, Table S5). 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, Fig. 2A, Table S5). 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. 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 (Fig S2A; Table S6), a trend observed across all cell states. Likewise, genes located in A compartments show higher average expression than those located in B compartments (Fig S2B). 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. However, ~6% of the CD4+ T cell genome did experience a transient (A>B>A or B>A>B) or stable (A>B>B, A>A>B, B>B>A or B>A>A) shift in A/B compartmentalization in response to activation (Fig. 2B). Genes that switched from A to B compartments are enriched for processes involved in fatty acid receptors and metabolism, eicosanoid and leukotriene biosynthesis, Runx and Yap activity, and IL-20 signaling, while genes that switched from B to A compartments are enriched for pathways involved in cell-ECM and cell-cell interactions, vitamin metabolism, and MECP2 activity (Fig. 2B), and include activation-induced genes known to regulate T cell function like EGR2 and IL12RB2.
Within each A or B compartment, the genome is further organized into topologically associating domains (TADs). 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 (Fig. 2C, Fig. S2C and D, Table S7). 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 (Fig. 2C).

We 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. 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 (Fig S3A). 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 (Fig. S3B-D).

Within active topological structures, transcriptional enhancers can regulate the expression of distant genes by looping to physically interact with gene promoters. 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, Fig. 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\textsuperscript{30} (Fig. S2E). 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 (Fig. 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 (Fig. 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 (Fig. 2F, Table S8). The accessibility of promoter-connected OCR and the expression of their associated genes decreased globally from 8 to 24 hours of stimulation (Fig. 2F). We compared these loop calls to a prior chromatin capture analysis in CD4+ T cells by Burren et al.\textsuperscript{17} 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 (Fig. S2F). 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 Fig. 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 (Fig. S2G). 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 (Fig. S2H), as exemplified by GEM and IRF4 (Fig. S4A,B). Conversely, the 3D regulatory architecture of genes like KLF2 and DPEP2, which were downregulated following stimulation, exhibited decreased contact and accessibility (Fig. S4C,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 (Table S9). 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 (Fig. 3A and C, Table S9), 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 (Fig. 3A). The activity of NFE2 and SMAD was increased without increased expression (Fig. 3A and C), likely due to post-translational regulation of these factors by phosphorylation. 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 (Fig. 3A). By 24 hours post-activation, bZIP activity remained largely unchanged compared to 8 hours (Fig. 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, KLF2, which controls genes involved in naïve CD4+ T cell quiescence and homing, NRF1, a factor implicated in age-associated T cell hypofunction, and EGR2 and 3, which are known to oppose T cell activation and promote tolerance (Fig. 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 (Fig. S5A). 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 (Fig. 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, Fig. 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\textsuperscript{39–42}. 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, Fig. 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 (Fig. 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 (Fig. 3E, Table S10). 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\textsuperscript{43–45} 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 (Fig S5B). 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\textsuperscript{46} 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\textsuperscript{47}, and XBP1, a major transcriptional mediator of the unfolded ER protein response with defined roles in T cell activation, differentiation, and exhaustion\textsuperscript{48,49}. Factors specifically implicated at 24 hours post-activation include E2F1, a transcriptional regulator of both cell cycle and apoptosis in T cells\textsuperscript{50,51}, BHLHE40, a factor known to control multiple aspects of T cell metabolism and differentiation\textsuperscript{52}, 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 (Fig S5B). 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 cREs and putative 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\textsuperscript{53–57}, we intersected promoter-interacting OCR with autoimmune SNPs from the 95% credible set derived from 15 autoimmune diseases (Fig. 4A, Tables S11 and S12). 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) (Fig. 4B, Table S13). 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, Fig. 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, Fig. 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 (Table S12). A total of 1151 autoimmune variants localized to the promoters of 400 genes (~1500/+500bp from the TSS, Table S14). These variants were on average 103 kb from the TSS of their connected gene (Fig. 4C), and each variant contacted an average of 5 genes (Fig. 4D). The majority of contacted SNPs interact with genes in addition to the nearest gene, and ~half of linked SNPs ‘skip’ the nearest gene to target only distant genes (Fig. 4E). Approximately 60% of connected genes were implicated across all timepoints (Fig. 4F, Table S12), 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 (Fig. 4G, Table S12). 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 (Fig. 4H). The accessibility of this SNP and it’s contact with TRIP10 and GPR108 increased following activation (Fig. 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 (Fig. 4I). TRIP10 encodes a cytoskeletal binding protein involved in endocytosis that suppresses glucose uptake in response to insulin signaling.\textsuperscript{58}
GPR108 encodes an orphan G-protein coupled receptor, and SIK1 encodes a salt-inducible kinase with roles in cancer, epilepsy, and myeloid signaling\textsuperscript{59}. None of these genes have been previously implicated in T cell activation.

**Functional validation of autoimmune V2G-implicated cis-regulatory elements**

This analysis specifically implicated potential 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\textsuperscript{60–63}. Activation-induced transcription of *IL2* involves an upstream regulatory region (URR) \textasciitilde 375 bp from the TSS that has served as a paradigm of tissue-specific, inducible gene transcription for nearly four decades\textsuperscript{64–66}. However, the presence of evolutionarily conserved non-coding sequences (CNS) in the \textasciitilde 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*\textsuperscript{67} (Fig. 5A). 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, Fig. 5A). In TCR/CD28-stimulated naïve CD4+ T cells, these CNS are remodeled to become highly accessible (Fig. 5A), and they loop to interact physically with the *IL2* URR (Fig. 5B) at both time points. ChIP-seq analyses in human T cells (ENCODE\textsuperscript{68}) 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* (Fig. S6), 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 (Fig. S6). 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 (Fig. 5C). 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 (Fig. 5D) and measured IL-2 secretion following TCR/CD28 stimulation (Fig. 5E). 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 (Fig. 5E). 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 (Fig. 5E). To determine whether these intergenic enhancers exist in synergistic epistasis 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 (Fig. 5F). 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 (Fig. 5G). 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 (Fig. S6). 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.

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. Expression quantitative trait loci (eQTL) studies identify variant-gene pairs through statistical associations between gene expression and genotype, offering an orthogonal line of evidence to compare with chromatin conformation capture approaches that identify variant-gene pairs by biophysical association. We compared our variant-gene pairs to those from a recent scRNA-seq-based eQTL study in activated CD4+ T cells and identified 41 (out of 127) pairs implicated by both methods, an overlap 10-fold higher than obtained by random sampling of genes within 500 kb of autoimmune variants (Fig. 6A, Fig S7A). We experimentally validated 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 (Fig. 5H).

Autoimmune variants are likely to influence disease risk by altering the activity of cis-regulatory elements in T cells. Indeed, 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 (Table S15), including PLAG1, PRDM1, BACH2, MYC, TBX21, BHLHE40, LEF1, TCF7, BCL6, IRF, p53, STAT (Fig. S8A), and hundreds of NFkB, EGR, KLF, FKH/FOX, and FOS-JUN sites (Fig. 6B). For example, the MS-associated SNP rs1077667 is located in an element that exhibits increased accessibility and contact frequency with the promoters of GPR108 and TRIP10 upon activation. This variant was predicted to reduce affinity for TP53, TP63, and OCT2/POU2F2 (Fig. S8B). Similarly, the T1D SNP rs3024505 (PP = 0.200) connected to the
promoters of *IL9* and *FAIM3* (Fig. S8C) was predicted to disrupt binding of *MZF1*, while the celiac SNP rs13010713 (PP = 0.154) contacting the *ITGA4* promoter (Fig. S8D) likely disrupts a SOX4 binding site.

**Functional validation of V2G-implicated effector gene sets**

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*.[20–22]. The set of all V2G-implicated genes was highly enriched for genes shown to regulate IL-2, IL-2 receptor, CTLA-4, and proliferation (Fig. 6C, Fig. S7B, Table S16). 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 (Table S16). 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 (Fig. 6C, Fig. S7B). 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 (Fig. 6D, Fig. S7C, Table S16). 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 (Fig. 6D, Table S16).
Identification of novel regulators of T cell activation

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 gene products have at least one chemical modulator currently available (Fig 6E, Table S17). 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 (Fig. 6F). 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 encode novel targets, including the kinases GRK6, PTK6, SIK1, and MAP3K11, the G protein-coupled receptors OXER1, GPR183, GPR18, and KISS1R, the acetylcholine receptor CHRN1, 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 CHRN1, KISS1R and OXER1 did not significantly affect T cell proliferation, however, small molecules targeting GRK6, PTK6, MAP3K11, GPR183, GART, and SIK1 inhibited activation of both human and mouse T cells in the nanomolar to micromolar range (Fig. 6G). 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, evolutionarily conserved 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. We conducted pharmacologic targeting experiments and compared our results with orthogonal eQTL and
CRISPR-based studies to validate sets of effector genes with a confluence of evidence supporting their role in CD4+ T cell activation. This study and prior 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), and the majority of cRE/variants ‘skip’ the nearest gene to interact with distant genes. For these reasons, chromatin conformation mapping in disease-relevant cell types brings a crucial added dimension to genome-scale 1D data for understanding the complex genetic and epigenetic mechanisms that regulate gene expression, and for predicting novel therapeutic targets.

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. 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. The orphan G protein-coupled receptor GPR108 is another top gene in our study that has not been studied in T cells, but was identified in a recent CRISPR screen for genes affecting IL-2 levels. Also co-implicated by our study and recent CRISPR screens are the cannabidiol receptor GPR18 and the purine biosynthetic enzyme GART. The GPR18 agonist arachidonyl glycine inhibited CD4+ T cell activation above 10 uM, while the GPR18 antagonist O-1918 slightly enhanced T cell activation. Antagonism of GART, an enzyme we previous identified as a V2G effector gene in COVID19 severity, 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, and PTK6, an oncogenic kinase studied in the context of cancer, led to inhibition of T cell activation in the nM to uM range. Inhibition of MAP3K11, a kinase that facilitates signaling...
through JNK1 and IkappaB kinase in cancer\textsuperscript{82,83}, and 25-hydroxycholesterol, a ligand of the G protein-coupled receptor GPR183, inhibited stimulation-induced CD4+ T cell proliferation in the 100 nM range. SIK1 is a member of the salt-inducible kinase family that negatively regulates TORC activity\textsuperscript{59}, 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 \textasciitilde 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\textsuperscript{84}, 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 \textit{cis}-regulatory elements, transcription factors, and genetic variants.

**MATERIALS AND METHODS**

**T cell isolation and \textit{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 EasySep\textsuperscript{TM} 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.
RNA-seq library generation and sequencing

RNA was isolated from ~ 1 million of each cell stage using Trizol Reagent (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 Bioanalyzer 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 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 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 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, and duplicate reads were removed from the alignment. Narrow peaks were called independently for each replicate using macs2 (-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 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\textsuperscript{88} 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\textsuperscript{89}, 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-dcc/pairix), and eventually converted to pre-binned Hi-C matrix in .cool format by cooler v0.8.10\textsuperscript{90} 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\textsuperscript{91}. Replicate similarity was determined by HiCRep v1.12.2\textsuperscript{90} at 10K resolution. For each sample, eigenvectors were determined from an ICE balanced Hi-C matrix with 40kb resolution using cooltools v0.3.2\textsuperscript{92} 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)\textsuperscript{93}. 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\textsuperscript{94} (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\textsuperscript{95} 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.

593

**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)\(^96\) 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\(^97\) with the MSigDB hallmark
geneset\(^98\), with resulting scores analyzed using limma (limma_3.50.3)\(^99\). Low abundance peaks
(CPM>1) were excluded from the analysis. Testing for differential expression was conducted with
the glmQLFit approach implemented in edgeR\(^88\). 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.

610

**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\(^100\). 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. 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 (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 the supplemental table 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. 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).
Genomic reference and visualizations

All analyses were performed using the hg19 reference genome using genencodeV19 as the gene reference. Genomic tracks were visualized with pyGenomeTracks v3.5. 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.

Colocalized eQTL comparisons

Enrichment of sentinel-gene assignments was conducted similarly as described previously. Briefly, a null distribution was constructed by randomly selecting genes within 1 Mb of the sentinel compared to the set of colocalized cis-eQTL 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. 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). 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 (Table S17). 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.109

**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:

<table>
<thead>
<tr>
<th>CNS region</th>
<th>CRISPR gRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>-46 CNS</td>
<td>gRNA #1 AGGATGCCTACCTCCAAATG</td>
</tr>
<tr>
<td></td>
<td>gRNA #2 AGGTGACAACATTTAGTCAG</td>
</tr>
<tr>
<td>-51 CNS</td>
<td>gRNA #1 GGCAACGAAATTCACTGTGA</td>
</tr>
<tr>
<td></td>
<td>gRNA #2 ATTCTAACAGGAATCATTGC</td>
</tr>
<tr>
<td>-80 CNS</td>
<td>gRNA #1 GTTCTACCTATGCGCATTG</td>
</tr>
<tr>
<td></td>
<td>gRNA #2 GAGATTTACTCAGTCCAAATG</td>
</tr>
<tr>
<td>-83 CNS</td>
<td>gRNA #1 GTGACAAGCATGACTCTACA</td>
</tr>
<tr>
<td></td>
<td>gRNA #2 GTGATGGTGAATTAAGCTGA</td>
</tr>
<tr>
<td>-85 CNS</td>
<td>gRNA #1 AGGGTTTTCTAGTTACGAGA</td>
</tr>
<tr>
<td></td>
<td>gRNA #2 ATGGTTAGTTAGCTCCCAAG</td>
</tr>
<tr>
<td>-96 CNS</td>
<td>gRNA #1 TGGGAAAAACATCTTACCTG</td>
</tr>
<tr>
<td></td>
<td>gRNA #2 TGGCCCATGAACCATCAAAG</td>
</tr>
<tr>
<td>-122 CNS</td>
<td>gRNA #1 GTTATTAATCTAAGCGGAGA</td>
</tr>
<tr>
<td></td>
<td>gRNA #2 GGAAGTTAGGCAATGATCG</td>
</tr>
<tr>
<td>-128 CNS</td>
<td>gRNA #1 CTTCAATCATTCGATTCAC</td>
</tr>
<tr>
<td></td>
<td>gRNA #2 TGACACCACCCCCTGCTTGA</td>
</tr>
</tbody>
</table>
HEK 293T cells were grown in RPMI1640 complete medium (RPMI1640 + 1X P/S, 1x L-Glu, 10% FBS), 37°C, 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 37°C 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 x10^6/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.

<table>
<thead>
<tr>
<th>Name</th>
<th>20nt gRNA Seq</th>
</tr>
</thead>
<tbody>
<tr>
<td>46CNS-1</td>
<td>GACGTATATAGTCATCTGAT</td>
</tr>
<tr>
<td>46CNS-2</td>
<td>TCTGTGGAGCTGCTGCGTTA</td>
</tr>
<tr>
<td>46CNS-3</td>
<td>AGGATGCCTACCTCCAAATG</td>
</tr>
<tr>
<td>51CNS-1</td>
<td>ATTCTAACAGGAATCATTTCG</td>
</tr>
<tr>
<td>51CNS-2</td>
<td>GAGTTAAAAAGACGTGTTACC</td>
</tr>
<tr>
<td>51CNS-3</td>
<td>GGCAACGAAATTCACTGTGTA</td>
</tr>
<tr>
<td>80CNS-1</td>
<td>AATGCGGCATAGGTAGAACT</td>
</tr>
<tr>
<td>80CNS-2</td>
<td>GAGATTTACTCAGTCCAATG</td>
</tr>
<tr>
<td>80CNS-3</td>
<td>TGGTGTCACAGTAACTAGG</td>
</tr>
<tr>
<td>83CNS-1</td>
<td>GTGACAAGCAGTACTCTACA</td>
</tr>
<tr>
<td>83CNS-2</td>
<td>GTGATGGTGAATTAAGCTGA</td>
</tr>
<tr>
<td>83CNS-3</td>
<td>TGGGCTCTGACTCATTAGA</td>
</tr>
<tr>
<td>85CNS-1</td>
<td>GCATAACATTTGACTTCTCTAC</td>
</tr>
<tr>
<td>85CNS-2</td>
<td>AATCTATGCAAGGGGTGAAT</td>
</tr>
<tr>
<td>85CNS-3</td>
<td>GCATCATGATGAAGCTTATC</td>
</tr>
<tr>
<td>122CNS-1</td>
<td>GTTATTAATCTAAGCGGAGA</td>
</tr>
<tr>
<td>122CNS-2</td>
<td>GGAAGTTCAGGCAGTCAATCG</td>
</tr>
<tr>
<td>122CNS-3</td>
<td>CACTTTGTGCTCGGATGCTC</td>
</tr>
<tr>
<td>128CNS-1</td>
<td>CTTCAATCATTCATCCAC</td>
</tr>
<tr>
<td>128CNS-2</td>
<td>TCAAGGCAGGGGTGTTGCTAA</td>
</tr>
<tr>
<td>128CNS-3</td>
<td>TGGTGATTCATCTTTAGCAT</td>
</tr>
<tr>
<td>IL-2URR-1</td>
<td>TCCATTcacgtcagtctttgg</td>
</tr>
</tbody>
</table>
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 Fl-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\textsuperscript{110}.

**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\textsuperscript{hi}CD44\textsuperscript{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) Gene set variance analysis (GSVA, N=3) using log-transformed expression values to MSigDB Hallmark gene set. Red indicates higher activity and blue indicates lower activity. (D) 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. (E) Gene set enrichment for each cluster of differential genes using the REACTOME pathway database. Enriched pathways associated with differentially expressed genes in those regions. X-axis indicates ratio of observed pathway members compared to those expected by chance, the Y-axis indicates the -log10(p-value) of enrichment.
Figure 2: Chromatin accessibility and architecture changes across three stages of T cell activation. (A) Volcano plots depicting the log₂ change in accessibility in all reproducible OCRs (present in 2 of the 3 replicates) compared to the -log₁₀(FDR), significant points are indicated in red (FDR < 0.05; abs(log₂FC) > 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. Significantly enriched pathways of genes located in variable A/B compartments are shown as bubble plots. Color indicates the pattern of A/B compartment assignment across the three timepoints, and node size indicates the number of genes in that pathway. (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. Enriched pathways associated with differentially expressed genes in those regions are shown. (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 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 log₂FC 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 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: 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 naive 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=5 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=5, 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). (H) Recombinant reporter assay in activated primary CD4+ T cells
(N=5 donors) showing transcriptional activity of the reference vs. risk alleles of the IL2 -128 cRE relative to the URR alone (P<0.0001).

**Figure 6: Functional validation of autoimmune V2G-implicated genes.** (A) Physical variant-gene associations (HiC) are highly 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. (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, CHRN, OXER1, GPR18, GRK6, PTK6, MAP3K11, GPR183, GART, and SIK1 on proliferation of anti-CD3/28 activated murine (left) and human (right) CD4+ T cells *in vitro* (N=4).
Supplemental Tables

Table S1: Expression and clustering of differentially expressed genes.
Table S2: Pathway enrichment of different RNA-seq genes
Table S3: Accessible chromatin regions
Table S4: Differential accessible open chromatin regions
Table S5: Differential contact frequency
Table S6: Differential TAD boundaries
Table S7: A/B compartment calls
Table S8: Stripe calls
Table S9: Summary of TF footprinting
Table S10: TF target gene pathway enrichment
Table S11: List of all GWAS studies included
Table S12: Variant to gene mapping across all timepoints
Table S13: Variant to gene mapping promoter OCRs
Table S14: Partitioned LD score regression output
Table S15: V2G genes implicated by orthogonal data
Table S16: Motifs predicted to disrupt transcription factor binding sites.
Table S17: V2G target gene drug repurposing results

REFERENCES

1028 52. Cook, M. E., Jarjour, N. N., Lin, C.-C. & Edelson, B. T. Transcription Factor Bhlhe40 in
1030 53. Chesi, A. et al. Genome-scale Capture C promoter interactions implicate effector genes at
1032 54. Su, C. et al. Mapping effector genes at lupus GWAS loci using promoter Capture-C in follicular
1034 55. Pahl, M. C. et al. Cis-regulatory architecture of human ESC-derived hypothalamic neuron
1035 differentiation aids in variant-to-gene mapping of relevant complex traits. Nat Commun 12,
1036 6749 (2021).
1037 56. Pahl, M. C. et al. Implicating effector genes at COVID-19 GWAS loci using promoter-focused
1039 57. Su, C. et al. 3D chromatin maps of the human pancreas reveal lineage-specific regulatory
1041 58. Feng, Y. et al. The Cdc42-interacting Protein-4 (CIP4) Gene Knock-out Mouse Reveals
1043 59. Darling, N. J. & Cohen, P. Nuts and bolts of the salt-inducible kinases (SIKs). Biochem J 478,
1045 60. Ceresaletti, K. et al. Multiple autoimmune-associated variants confer decreased IL-2R
1046 signaling in CD4+ CD25(hi) T cells of type 1 diabetic and multiple sclerosis patients. Plos
1047 One 8, e83811 (2013).
1048 61. Spolski, R., Li, P. & Leonard, W. J. Biology and regulation of IL-2: from molecular mechanisms
1050 62. Abbas, A. K., Trotta, E., Simeonov, D. R., Marson, A. & Bluestone, J. A. Revisiting IL-2:
1052 63. Joosse, M. E. et al. Duplication of the IL2RA locus causes excessive IL-2 signaling and may
1053 predispose to very early onset colitis. Mucosal Immunol 1–11 (2021) doi:10.1038/s41385-
1054 021-00423-5.
1055 64. Fujita, T., Shibuya, H., Ohashi, T., Yamanishi, K. & Taniguchi, T. Regulation of human
1056 interleukin-2 gene: Functional DNA sequences in the 5’ flanking region for the gene
1059 fragment at the 5’ end of the interleukin 2 gene enhances expression from a heterologous
1060 promoter in response to signals from the T cell antigen receptor. J Exp Medicine 165, 395
1062 66. Novak, T. J., White, P. M. & Rothenberg, E. V. Regulatory anatomy of the murine interleukin-
1064 67. Mehra, P. & Wells, A. D. Long-Range Transcriptional Control of the Il2 Gene by an Intergenic
1070 70. Hedlund, G., Dohlsten, M., Ericsson, P. O. & Sjögren, H. O. Rapid response to Con A by
1071 CD4+CD45R− rat memory lymphocytes as compared to CD4+CD45R+ lymphocytes. Cell
1073 71. Picker, L. et al. Direct demonstration of cytokine synthesis heterogeneity among human
1075 72. Sojka, D. K., Bruniquel, D., Schwartz, R. H. & Singh, N. J. IL-2 Secretion by CD4+ T Cells In
1076 Vivo Is Rapid, Transient, and Influenced by TCR-Specific Competition. J Immunol 172,


A Chromatin accessibility dynamics (ATAC-seq)

Unstim vs. 8hr

8hr vs. 24hr

Unstim
8hr
24hr

B AB compartment switching

Dynamic

6.6%

A stable

51.7%

B stable

41.7%

C

Stable TAD

80%

Merge

Complex

Shifted

Strength

Split

D Loop calls (HiC)

Number of loops

- 1000
- 500
- 0

Unstim
8hr
24hr

E Differential loops (HiC)

Number of loops

- 1500
- 750
- 250

Unstim vs. 8hr
8hr vs. 24hr

F Integrated accessibility, connectivity, and expression dynamics

Unstim vs. 8hr
8hr vs. 24hr

Chromatin accessibility fold change (log2)

Density

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

Expression fold change

Contact frequency change

increased

stable

decreased

unchanged
A

Autoimmune Disease GWAS loci

SNPs in 95% Credible set HiC loops Contacted accessible variants and genes Differential Expression

B

C

Median = 11.55 (103,351bp)

D

Median = 5 genes per interacting variant

E

% of interacting variants

F

Interaction size

G

H

MS locus

rs1077667

Unstim

8 hr

I

ALG locus

rs7380290

Unstim

8 hr

Figure 4

Figure 4

Figure 4

Figure 4
**Figure 5.**

**A.** Unstimulated and stimulated conservation of the IL2 locus with open chromatin and GWAS SNPs. GS7269153, rs6222750, rs1512973, rs12504008.

**B.** 3D architecture of the IL2 locus showing Unstimulated and stimulated conditions.

**C.** Reporter assay for IL2 CNS enhancer activity showing luciferase activity (fold over URR).

**D.** CRISPR-CAS9 deletion of individual CNS regions showing IL2 and IL21 expression.

**E.** Primary human CD4+ T cells (N=5) showing IL2 expression.

**F.** CRISPR-CAS9 deletion of all intergenic CNS regions showing a 81.3 kb deletion.

**G.** Jurkat clones (N=3) showing IL2 and IL21 transcript levels.

**H.** Jurkat clones (N=3) showing IL2 and IL21 transcript levels with PMA+I+a-CD28 stim (hrs).

The images and data support the findings of the study, demonstrating the impact of CRISPR-CAS9 deletions on IL2 and IL21 expression in both cellular and reporter assay contexts.
Figure 6

A. eQTL implicated genes

B. TF affinity

C. V2G geneset enrichment for CRISPR-validated targets

D. Reported Mouse Phenotypes

E. chemical available

F. T cell specificity score

G. mouse vs human % max. CD8+T cell activation

The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under CC-BY-NC-ND 4.0 International license.
Figure S1: 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 S2: Characterization of CD4+ T cell epigenomic data. (A) Accessibility of OCR and (B) expression of genes located in A vs. B compartments (log2 fpkm). 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-quartile range).
Figure S3: 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 S4: Early genome-wide epigenetic changes at dynamically expressed genes during CD4+ T cell activation. Heatmaps display 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 end of the chromosome annotation, and inset shows expression (TPM) at each time point for each gene.
Figure S5: 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 S6: 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 S7: Orthogonal validation of 3D chromatin-based V2G genes with expression changes in single-cell eQTLs (A), prior CRISPRi/a screens (B), and genes with known immune phenotypes in the international mouse phenotyping consortium. Red = increased, blue = decreased.

Implicated genes with immune phenotypes in IMPC

Implicated genes with immune phenotypes in MPC

Implicated genes with immune phenotypes in IMPC
Figure S8: Predicted impact of autoimmune disease-associated genetic variation at V2G-implicated loci. (A) Top 50 SNP-affected TF binding sites. Dashed box indicates the SNP affected by allelic variation. HiC contact frequency change (fold change) for all affected sites is shown. (B) Genetic variation at the MS rs1077667 SNP is predicted to disrupt TP53, TP63, and POU2F2 (OCT2) binding sites. Dashed box indicates the SNP affected by allelic variation. HiC contact frequency change (fold change) for all affected sites is shown. (C) and (D) SNP rs3024505 is predicted to disrupt a MZF1 motif and rs13010713 is predicted to disrupt a SOX4 binding motif.