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. Thomas[1,3], Zachary Thompson[1,3], Zachary Mount[1,3], James Pippin[1,2], Peter A. Morawski[4], 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.

51 **INTRODUCTION**

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

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

75 Chromatin conformation assays allow for identification of putative target genes of autoimmune
76 variant-containing cRE in close spatial proximity to gene promoters. Previous work using

77 promoter-capture HiC and HiC in combination with other epigenetic marks have implicated sets
78 of autoimmune variants and effector genes that may participate in T cell activation[17–19]. In addition
79 to chromatin conformation approaches, multiple complementary approaches have been
80 developed to link disease associated SNPs to their downstream effector genes. Expression
81 quantitative trait mapping (eQTL) and chromatin co-accessibility data can be used to implicate
82 effector genes through statistical association of genotypes with readouts of expression or other
83 molecular markers[20]. However, these approaches do not account for the relevant 3D structure of
84 the genome in the nucleus, and are highly susceptible to *trans*-effects and other confounding
85 factors.

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

98 **RESULTS**

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

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

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

124 **Supplementary File 1**). Genes upregulated early upon activation (cluster 1; n=1621 genes) are
125 enriched for pathways involved in the unfolded protein response, cytokine signaling, and
126 translation (*e.g.*, *IL2, IFNG, TNF, IL3, IL8, IL2RA, ICOS, CD40LG, FASLG, MYC, FOS, JUNB,*
127 *REL, NFKB1, STAT5A,* **Supplementary File 3**). Genes downregulated late (cluster 2; n=1600
128 genes) are moderately enriched for receptor tyrosine kinase signaling, cytokine signaling, and
129 extracellular matrix organization. Genes monotonically increasing (cluster 3; n=2676 genes) are
130 highly enriched for pathways involved in infectious disease and RNA stability, translation and
131 metabolism, and moderately enriched for pathways involved in the unfolded protein response,
132 cellular responses to stress, regulation of apoptosis, and DNA repair (*e.g.*, *TBX21, BHLHE40,*
133 *IL12RB2, STAT1, CCND2, CDK4, PRMT1, ICAM1, EZH2,* **Supplementary File 3**). Genes
134 downregulated early (cluster 4; n=1628 genes) are enriched for pathways involved in inositol
135 phosphate biosynthesis, neutrophil degranulation, and metabolism of nucleotides (*e.g.*, *KLF2,*
136 *IL7R, RORA, IL10RA*), and genes upregulated late (cluster 5; n=2154 genes) are highly enriched
137 for pathways involved in cell cycle, DNA unwinding, DNA repair, chromosomal maintenance, beta-
138 oxidation of octanoyl-CoA, and cellular response to stress (*e.g.*, *CDK2, E2F1, CDK1, CCNE1,*
139 *CCNA2, PCNA, WEE1, CDC6, ORC1, MCM2*). These patterns are consistent with known
140 changes in the cellular processes that operate during T cell activation, confirming that *in vitro*
141 stimulation of CD4+ T cells recapitulates gene expression programs known to be engaged during
142 a T cell immune response.
143

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

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

150 ~500 cubic micron nucleus of a cell in a hierarchically ordered manner. This degree of compaction
151 results in only ~1% of genomic DNA being accessible to the machinery that regulates gene
152 transcription[25], therefore a map of open chromatin regions (OCR) in a cell represents its potential
153 gene regulatory landscape. Open chromatin mapping of human CD4+ T cells at all states
154 identified a consensus *cis*-regulatory landscape of 181,093 reproducible OCR (**Supplementary**
155 **File 4**). Of these, 14% (25,291) exhibited differential accessibility following 8 hours of stimulation
156 (FDR<0.05). Most differentially accessible regions (DAR) became more open (18,887), but some
157 DAR (6,629) showed reduced accessibility (**Figure 2A**, **Supplementary File 5**). The change in
158 accessibility over the next 16 hours of stimulation showed the opposite dynamic, with 6,629
159 regions exhibiting reduced accessibility, and only 4,417 DAR becoming more open (total of 11,046
160 DAR, **Figure 2A**, **Supplementary File 5**). These OCR represent the set of putative cRE with
161 dynamic activity during T cell activation.

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

175 compartmentalized into A (42%) and B (52%), indicating that activation does not cause a major
176 shift in the large-scale organization of the genome within the nucleus (**Figure 2B**).

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

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

201 regulatory cRE-gene interactions, we identified high confidence loop contacts across all cell states
202 using Fit-HiC (merged 1kb, 2kb, and 4kb resolutions, **Figure 2D**). This approach detected
203 933,755 loop contacts in quiescent naïve CD4+ T cells, 900,267 loop contacts in 8-hour
204 stimulated cells, and 551,802 contacts in 24-hour stimulated cells (2,099,627 total unique loops).
205 Approximately 23% of these loops involved a gene promoter at one end and an OCR at the other,
206 and these promoter-interacting OCR were enriched for enhancer signatures based on flanking
207 histone marks from CD4+ T cells in the epigenome roadmap database[32] (**Figure 2 - Figure**
208 **Supplement 1E**). T cell activation resulted in significant reorganization of the open chromatin-
209 promoter interactome, as 907 promoter-OCR exhibited increased contact and 1333 showed
210 decreased contact following 8 hours of stimulation (**Figure 2E**). Continued stimulation over the
211 next 16 hours was associated with an increase in the contact frequency of 41 promoter-OCR
212 pairs, while only 4 pairs exhibited decreased contact (**Figure 2E**). Activation-induced changes in
213 chromatin architecture and gene expression were highly correlated, as genomic regions exhibiting
214 increased promoter connectivity became more accessible at early stages of stimulation, which
215 was associated with increased gene transcription from connected promoters (**Figure 2F,**
216 **Supplementary File 9**). The accessibility of promoter-connected OCR and the expression of their
217 associated genes decreased globally from 8 to 24 hours of stimulation (**Figure 2F**). We compared
218 these loop calls to a prior chromatin capture analysis in CD4+ T cells by Burren et al.[18] and found
219 that roughly 40% of stable loops in both stimulated and unstimulated cells were identical in both
220 studies, despite differing in approach (HiC vs. PCHiC), analysis (HiC vs. CHiCAGO), sample
221 (naïve CD4+ vs. total CD4+), timepoint (8 vs. 4 hour), and donor individuals (**Figure 2 - Figure**
222 **Supplement 1F**). As expected, unstimulated samples were more similar than activated samples.
223 We next focused on the 5 sets of genes with the dynamic expression patterns defined in
224 **Figure 1D**, and identified 57,609 OCR that contact dynamic gene promoters in at least one stage.
225 Most dynamic genes contacted between 1 and ~35 OCR, with a median of 10 OCR per gene, but
226 a handful of dynamic genes were observed in contact with over 100 distinct open chromatin

227 regions (**Figure 2 - Figure Supplement 1G**). Similarly, most OCR were connected to a single
228 dynamic gene, but many contacted more than one gene (median 2 genes per OCR), suggesting
229 that most dynamic genes have a complex regulatory architecture. Increased gene expression
230 upon activation correlated with an increase in the accessibility and promoter contact frequency of
231 distant cRE (**Figure 2 - Figure Supplement 1H**), as exemplified by *GEM* and *IRF4* (**Figure 2 -**
232 **Figure Supplement 3A**,**B**). Conversely, the 3D regulatory architecture of genes like KLF2 and
233 DPEP2, which were downregulated following stimulation, exhibited decreased contact and
234 accessibility (**Figure 2 - Figure Supplement 3C**,**D**).
235

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

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

253 also cooperates with CTCF to regulate chromatin looping[36], KLF2, which controls genes involved
254 in naïve CD4+ T cell quiescence and homing[2,37], NRF1, a factor implicated in age-associated T
255 cell hypofunction[38], and EGR2 and 3, which are known to oppose T cell activation and promote
256 tolerance[34,39,40] (**Figure 3B** and **C**).

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

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

279 in this network were enriched for pathways involved in immune signaling, DNA replication and
280 repair, protein secretion, and programmed cell death (**Figure 3 – Figure Supplement 1B**).
281 Costimulation through the TCR and CD28 induced a set of core network genes active at both time
282 points with known roles in T cell activation and differentiation (*NFKB1, JUNB, MYC, IRF4, STAT5,*
283 *STAT1, LEF1, ATF4*). Also part of this set is *PLAGL2*, an oncogene in the Wnt pathway that
284 regulates hypoxia-induced genes[48] with no prior defined role in T cell activation. Additional nodes
285 specifically implicated at 8 hours post-activation are HIF1A, the major sensor of cellular hypoxia[49],
286 and XBP1, a major transcriptional mediator of the unfolded ER protein response with defined
287 roles in T cell activation, differentiation, and exhaustion[50,51]. Factors specifically implicated at 24
288 hours post-activation include E2F1, a transcriptional regulator of both cell cycle and apoptosis in
289 T cells[52,53], BHLHE40, a factor known to control multiple aspects of T cell metabolism and
290 differentiation[54], and the Myc cofactor MAZ that has not been previously studied in the context of
291 T cell function. Genes connected to factors in the activated T cell networks were enriched for
292 pathways involved in cytokine signaling, the interferon response, transcription, cell cycle, DNA
293 replication and repair, and programmed cell death (**Figure 3 – Figure Supplement 1B**).
294 Together, these data indicate that concurrent but separable stimulation-induced gene programs
295 are the result of the activity of distinct sets of DNA binding factors mobilized by antigen and
296 costimulatory receptor signaling in naïve CD4+ T cells.
297

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

300 Following our established variant-to-gene (V2G) mapping approach to implicate functional
301 SNPs and their effector genes using the combination of GWAS and chromatin conformation
302 capture data[55–59], we intersected promoter-interacting OCR with autoimmune SNPs from the 95%
303 credible set derived from 15 autoimmune diseases (**Figure 4A**, **Supplementary Files 12** and **13**).
304 Constraining the GWAS SNPs in this way reduced the credible set size from an average of 14

305 variants per sentinel to 3 variants per sentinel. To determine whether open chromatin in physical
306 contact with dynamically regulated genes in CD4+ T cells is enriched for autoimmune disease
307 heritability, we performed a partitioned LD score regression analysis. This landscape was
308 enriched for variants associated with susceptibility to inflammatory bowel disease (IBD), ulcerative
309 colitis (UC), type I diabetes (T1D), lupus (SLE), celiac disease (CEL), allergy (ALG), eczema
310 (ECZ), and rheumatoid arthritis (RA), but not for variants associated with psoriasis (PSO) or
311 juvenile idiopathic arthritis (JIA) (**Figure 4B**, **Supplementary File 14**). The OCR connected to
312 genes upregulated early and/or progressively upon activation (clusters 1 and 3) were most
313 strongly enriched for ALG, CEL, IBD/UC, RA and T1D heritability, while SLE and ECZ heritability
314 was most enriched in OCR connected to genes upregulated later post-activation (clusters 3 and
315 5, **Figure 4B**). SLE was also the only disease (besides PSO and JIA) that was not enriched in
316 open chromatin connected to down-regulated genes (cluster 4, **Figure 4B**).

317 The promoter-connected open chromatin landscape for all CD4+ T cell states in this study
318 contains 2606 putatively causal variants linked to ~half of the sentinel signals (423) for the 15
319 diseases analyzed, and are in contact with a total of 1836 genes (**Supplementary File 13**). A
320 total of 1151 autoimmune variants localized to the promoters of 400 genes (-1500/+500bp from
321 the TSS, **Supplementary File 15**). These variants were on average 103 kb from the TSS of their
322 connected gene (**Figure 4C**), and each variant contacted an average of 5 genes (**Figure 4D**).
323 The majority of linked SNPs interact with genes in addition to the nearest gene, and ~half of linked
324 SNPs 'skip' the nearest gene to target only distant genes (**Figure 4E**). Approximately 60% of
325 connected genes were implicated across all timepoints (**Figure 4F**, **Supplementary File 13**),
326 while ~40% (753) are dynamically regulated (clusters 1-5) in response to TCR/CD28 co-
327 stimulation. Examples of SNP-genes pairs that exhibit dynamic accessibility, chromosome
328 contact, and expression in response to T cell activation are the *SIK1*, *PARK7*, *DUSP5*, *CLEC2D*,
329 *TRIP10*, *GPR108* and *IL2* loci (**Figure 4G**, **Supplementary File 13**). The *TRIP10* and *GPR108*
330 promoters were each captured in contact with a high confidence variant rs1077667 (PP>0.99),

331 which is located in an intron of *TNFSF14* and is associated with multiple sclerosis (**Figure 4H**).
332 The accessibility of this SNP and it's contact with *TRIP10* and *GPR108* increased following
333 activation (**Figure 4H**). Conversely, the allergy-associated SNP rs7380290 is accessible and
334 contacts the *SIK1* promoter in resting cells, but shows reduced accessibility and promoter
335 connectivity upon activation (**Figure 4I**). *TRIP10* encodes a cytoskeletal binding protein involved
336 in endocytosis that suppresses glucose uptake in response to insulin signaling[60], *GPR108*
337 encodes an orphan G-protein coupled receptor, and *SIK1* encodes a salt-inducible kinase with
338 roles in cancer, epilepsy, and myeloid signaling[61]. None of these genes have been previously
339 studied in the context of T cell activation.
340
341 **Comparative predictive power against orthogonal variant-to-gene mapping approaches**
342 In an attempt to gauge the predictive power of our approach relative to other V2G approaches,
343 we compared our chromatin capture-based autoimmune GWAS effector gene predictions to the
344 predictions of four other chromosome capture-based studies in human CD4+ T cells[17–19,62], four
345 single-cell eQTL studies in human CD4+ T cells[13,62–64], and a set of 449 genes that when mutated
346 in humans cause inborn errors of immunity[65]. As part of the comparison, we harmonized the
347 chromatin-based datasets using our chromatin loop calling approach and integrated with the
348 same 95% credible set of autoimmune variants to create comparable physical contact maps. A
349 total of 6936 unique genes were implicated by all studies. Our HiC-ATAC-seq approach in naïve
350 and activated human CD4+ implicated 1947 effector genes, 400 of which through proxies in open
351 promoters and 1836 through contacts with distal variants, of which 752 were dynamically
352 regulated throughout activation. The HiC dataset from Gate et al. implicated 110 genes, and the
353 capture-HiC datasets from Burren et al., Yang et al., and Javierre et al. implicated 1408, 2572,
354 and 2368 genes, respectively (**Supplementary File 16**). A total of 369 autoimmune GWAS
355 effector genes were predicted in common by our HiC approach and the pcHiC-based approaches.
356 The eQTL studies by Soskic et al., Gate et al., Ye et al., and Schmiedel et al. (DICE) implicated

357 171, 15, 15, and 221 genes, and the co-accessibility data from Gate et al. implicated 45 genes.
358 Concordance between our gene predictions and the pcHiC predictions ranged from 15-24%
359 (**Figure 5A**), while concordance between our predictions and the eQTL predictions was low
360 (0.6%~5%, **Figure 5A**), and concordance among eQTL studies was also low (1.6%-13%, **Figure**
361 **5A**). However, when we co-localized the Soskic variant-gene pairs with autoimmune GWAS
362 SNPs, 41 of these eQTL were also predicted by our study, representing a concordance 10-fold
363 higher than obtained by a random sampling of genes within 500 kb of autoimmune variants
364 (**Figure 5B**). A total of 1519 of the genes implicated in our study were not predicted by eQTL
365 approaches, while 644 genes from our CD4 T cell V2G data were not implicated by the other
366 chromatin-based datasets, and altogether 562 putative autoimmune effector genes were uniquely
367 predicted by our study. Potential sources of variation between the results of the studies are
368 outlined in the Discussion. A total of 17 genes were implicated by all CD4+ T cell-based
369 approaches that nominated at least 150 genes (APIP, BCL2L11, ERAP1, ERAP2, CD5, FIGNL1,
370 IL18R1, METTL21B, MRPL51, PPIF, PTGER4, PXK, RMI2, RNASET2, SERPINB1, TAPBPL,
371 VAMP1).

372 To measure the predictive power of the sets of genes implicated by each approach above, a
373 precision-recall analysis was performed against the 'gold standard' human inborn errors in
374 immunity gene set, based on the hypothesis that genes under the control of autoimmune-
375 associated variants would cause monogenic immune disease phenotypes if subjected to
376 deleterious coding mutations. The set of genes from our study with disease-associated SNPs
377 located in their open promoters overlapped with only 5% of the genes from the gold standard set
378 (**Supplementary File 16**, **Figure 5C**). However, leveraging our HiC data to include distal
379 autoimmune variants captured interacting with open gene promoters increased the sensitivity of
380 the approach 3-fold to 16% (71 overlapping genes, **Figure 5C**), confirming the relevance of long-
381 range promoter-cRE/SNP contacts. This level of sensitivity is comparable to that of other
382 chromosome capture-based datasets for predicting HIEI genes in CD4+ T cells (**Figure 5C**).

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

393 We also analyzed the relative predictive power of a recently proposed 'activity-by-contact'
394 (ABC) model[66] that uses a multi-tissue averaged HiC dataset instead of cell type-matched 3D
395 chromatin-based data to link variants to genes. This model constrains the input data by removing
396 bidirectional, antisense, and small RNAs, and potentially underweights distal elements connected
397 to ubiquitously expressed genes due to stronger average activity scores near their promoters.
398 Also, while most human genes have multiple alternative transcriptional start sites (median = 3),
399 ABC only annotates only one promoter per gene. To apply the activity-by-contact gene nomination
400 model to our CD4+ T cell chromatin-based V2G data, we used our ATAC-seq data and publicly
401 available CD4+ T cell H3K27ac ChIP-seq data as input, and integrated this with GWAS and the
402 averaged HiC dataset from the original ABC study[66,67]. The ABC model nominated 650 genes
403 compared to 1836 genes using our cell type-matched HiC data and analysis pipeline. Only 357
404 genes were nominated by both approaches, while 1479 genes nominated by our approach were
405 not implicated by the ABC model, and 293 genes not implicated by our approach were newly
406 implicated by ABC. To measure the predictive power of the ABC approach, we re-ran the
407 precision-recall re-analysis with all datasets subjected to the ABC gene-promoter filter (**Fig. 5D**).
408 We found that applying the restricted ABC model promoter annotation to all datasets did not have

409 a large effect on recall, however, the precision of several of the datasets were affected. For
410 example, using the restricted promoter/gene set reduced the precision of our V2G approach and
411 artificially inflated the precision of the 'nearest gene to SNP' metric. The precision-recall analysis
412 also shows that the ABC score-based approach is only half as powerful at predicting HIEI genes
413 as the chromatin-based V2G approaches (**Fig. 5D**). This indicates that informing GWAS data with
414 cell type- and state-specific 3D chromatin-based data brings more target gene predictive power
415 than application of the multi-tissue-averaged HiC used by the ABC model. Together, these
416 analyses indicate that chromosome capture-based V2G is a more sensitive approach for
417 identifying 'true' effector genes than eQTL or ABC approaches, but comes with additional
418 predicted genes that represent either false positives or true effector genes for which monogenic
419 LOF/GOF mutations have not yet been characterized in humans.
420

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

422 Our 3D chromatin-based analysis specifically predicts dynamic, disease-associated
423 regulatory elements in intergenic space at the *IL2* locus. The *IL2* gene encodes a cytokine with
424 crucial, pleotropic roles in the immune system, and dysregulation of IL-2 and IL-2 receptor
425 signaling leads to immunodeficiency and autoimmune disorders in mice and humans[68–71].
426 Activation-induced transcription of *IL2* involves an upstream regulatory region (URR) ~375 bp
427 from the TSS that has served as a paradigm of tissue-specific, inducible gene transcription for
428 nearly four decades[72–74]. However, the presence of evolutionarily conserved non-coding
429 sequences (CNS) in the ~150 kb of intergenic space 46, 51, 80, 83, 85, 96, 122, and 128 kb
430 upstream of the TSS suggest that additional regulatory elements may have evolved to control
431 *IL2*[75] (**Figure 6A**). The -51 kb CNS contains a SNP linked to T1D, IBD, PSO, CEL and allergy
432 (rs72669153), while the -85 kb CNS contains a SNP linked to RA (rs6232750) and the -128 kb
433 CNS contains two SNPs linked to T1D, JIA, and SLE (rs1512973 and rs12504008, **Figure 6A**).
434 In TCR/CD28-stimulated naïve CD4+ T cells, these CNS are remodeled to become highly

435 accessible (**Figure 6A**), and they loop to interact physically with the *IL2* URR (**Figure 6B**) at both
436 time points. ChIP-seq analyses in human T cells[76] show that the URR and all distal CNS except
437 -85 are occupied by TF such as Jun/Fos (AP-1), NFAT, and NFkB that are known regulators of
438 *IL2* (**Figure 6 – Figure Supplement 1**), and the -85, -122, and -128 CNS are occupied by
439 additional TF not previously thought to be direct regulators of *IL2*, such as MYC, BCL6, and
440 STAT5 (**Figure 6 – Figure Supplement 1**). Recombinant reporter assays in primary activated
441 human CD4+ T cells showed that the -46, -51, -83, and -128 CNS/OCR can enhance transcription
442 from the URR (**Figure 6C**). To determine whether the native elements contribute to the expression
443 of *IL2*, we targeted each CNS/OCR individually in primary human CD4+ T cells or Jurkat T cells
444 using CRISPR/CAS9 (**Figure 6D**) and measured IL-2 secretion following TCR/CD28 stimulation
445 (**Figure 6E**). Deletion of the -46, -51, -83, -85, -122, and -128 kb elements in primary human
446 CD4+ T cells each resulted in a ~50% reduction in IL-2 production, while deletion of the -80 kb
447 element had little effect. A very similar pattern of impact was observed when these elements were
448 deleted individually in Jurkat T cells (**Figure 6E**). The URR has a stronger contribution to IL-2
449 production than any individual intergenic element, as deletion of the URR almost completely
450 abrogated activation-induced IL-2 production by both primary CD4+ or Jurkat T cells (**Figure 6E**).
451 To determine whether these intergenic enhancers exist in synergistic epistasis[77] necessary for
452 *IL2* transactivation, we generated Jurkat T cell clones in which the stretch of all 7 elements located
453 46 kb to 128 kb upstream of *IL2* was deleted using CRISPR/CAS9 (**Figure 6F**). Despite the URR
454 and 46 kb of upstream sequence being intact in these clones, loss of the 81.3 kb stretch of
455 intergenic enhancers renders these cells incapable of expressing *IL2* at both the mRNA and
456 protein level in response to stimulation (**Figure 6G**). These results show that the URR is not
457 sufficient for activation-induced expression of *IL2*, and that *IL2* has a previously unappreciated,
458 complex, and autoimmune disease-associated regulatory architecture that was accurately
459 predicted by our 3D epigenomic V2G approach. Importantly, we find that the distal *IL2* cRE are
460 highly accessible in quiescent memory T cell subsets (Th1, Th2, Th17, Th1-17, Th22) isolated

461 directly *ex vivo* from human blood, whereas naïve CD4+ T cells and non-IL-2-producing Treg
462 showed little accessibility at these elements (**Figure 6 – Figure Supplement 1D**). This suggests
463 that stable remodeling of distal *IL2* cRE can persist *in vivo* after TCR signals cease, and that this
464 epigenetic imprinting contributes to the immediate, activation-induced production of IL-2 exhibited
465 by memory, but not naïve or regulatory, CD4+ T cells[78–80].
466

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

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

482 To test whether the distal -83 kb *Il2* enhancer contributes to physiologic immune responses
483 *in vivo*, we immunized wild-type and *Il2*-83-cRE-ko mice with the model antigen chicken
484 ovalbumin. *Il2*-83-cRE-ko animals showed a nominal increase in the differentiation of CD4+ T
485 cells into CXCR5+PD-1^{hi} follicular helper T cells (Tfh) in the spleen compared to wild-type animals
486 (**Figure 7D**), a process known to be antagonized by IL-2, and generated significantly elevated

487 levels of ovalbumin-specific IgG antibody following immunization (**Figure 7E**). To determine
488 whether the 83 kb *Il2* enhancer contributes to auto-inflammatory disease susceptibility *in vivo*, we
489 used an adoptive transfer model of IBD in which the development of colitis is determined by the
490 balance between conventional and regulatory CD4+ T cell activities[81,82]. Upon transfer into
491 lymphocyte-deficient Rag1-ko recipients, wild-type CD4+ T cells respond to microbiota in the gut
492 by inducing inflammatory colitis that leads to weight loss (**Figure 7F**, WT Tconv), while co-transfer
493 of wild-type Treg effectively controls inflammation (**Figure 7F**, WT Tconv + WT Treg). Transfer of
494 *Il2*-83-cRE-ko Tconv led to significantly less colitis (**Figure 7F**, *Il2*-83-cRE-ko Tconv), and was
495 associated with a nominal decrease in *Il2*-83-cRE-ko Tconv in the spleen, and a significant
496 decrease in the accumulation of *Il2*-83-cRE-ko Tconv in the mesenteric lymph nodes (MLN) that
497 drain the intestines (**Figure 7F** inset). Treg require IL-2 from Tconv for their function and
498 homeostasis[83], and despite their reduced numbers, *Il2*-83-cRE-ko Tconv were able to support the
499 homeostasis and function of co-transferred wild-type Treg (**Figure 7F**, *Il2*-83-cRE-ko Tconv + WT
500 Treg). That the -83 kb ortholog of the human -128 kb *IL2* enhancer contributes to physiologic
501 immune responses and inflammatory disease susceptibility *in vivo.*
502
503 **Impact of autoimmune risk-associated genetic variation on *cis*-regulatory element activity**
504 The chromatin conformation approach used here employs GWAS variants as 'signposts' to
505 identify disease-relevant regulatory elements and connect them to their target genes, but does
506 not *per se* determine the effect of disease-associated genetic variation on enhancer activity or
507 gene expression. We experimentally measured variant effects at the *IL2* locus, where the -128
508 enhancer defined above contains two SNPs linked to T1D, JIA, and SLE (rs1512973 and
509 rs12504008). Using a recombinant reporter assay in primary activated CD4+ T cells, we
510 confirmed that disease-associated genetic variation influences intergenic *IL2* enhancer activity at
511 the -128 kb element, in that the risk allele contributes significantly less transcriptional activity than
512 the reference allele (**Figure 8A**).

513 Autoimmune variants are likely to influence disease risk by altering the activity of *cis*-
514 regulatory elements in T cells. At genome scale, we identified over 1000 cRE likely impacted by
515 autoimmune disease-associated genetic variation, in that they contain autoimmune risk SNPs
516 that are predicted to decrease or increase transcription factor binding affinity. Overall, 1370
517 autoimmune risk variants in open chromatin were predicted to influence the activity of 495 DNA
518 binding factors (**Supplementary File 17**), including PLAG1, PRDM1, BACH2, MYC, TBX21,
519 BHLHE40, LEF1, TCF7, BCL6, IRF, p53, STAT (**Figure 8 – Figure Supplement 1A**), and
520 hundreds of NFkB, EGR, KLF, FKH/FOX, and FOS-JUN sites (**Figure 8B**). For example, the T1D
521 SNP rs3024505 (PP = 0.200) connected to the promoters of *IL9* and *FAIM3* (**Figure 8 – Figure**
522 **Supplement 1B**) and the celiac SNP rs13010713 (PP = 0.154) contacting the *ITGA4* promoter
523 (**Figure 8 – Figure Supplement 1C**) are predicted to disrupt binding sites for *MZF1* (**Figure 8 –**
524 **Figure Supplement 1D**) and SOX4 (**Figure 8 – Figure Supplement 1E**). Similarly, the MS SNP
525 rs1077667 contacting the promoters of *GPR108* and *TRIP10* (**Figure 4H**) is predicted to reduce
526 affinity for TP53, TP63, and OCT2/POU2F2 (**Figure 8 – Figure Supplement 1F**).
527

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

529 To determine whether genes identified *via* their physical interaction with autoimmune variants
530 in CD4+ T cell chromatin contact maps tend to be directly involved in T cell activation and function,
531 we compared the set of autoimmune genes implicated by chromatin contacts in this study to sets
532 of genes identified in CRISPR-based screens that control aspects of CD4+ T cell activation like
533 proliferation and expression of the inflammatory genes *IFNG, CTLA4, IL2, IL2RA,* and *TNF*[21–23].
534 The set of all V2G-implicated genes was highly enriched for genes shown to regulate IL-2, IL-2
535 receptor, CTLA-4, and proliferation (**Figure 8C, Figure 8 – Figure Supplement 2A,**
536 **Supplementary File 18**). For example, 202 genes shown to regulate IL-2 production and 166
537 genes shown to regulate proliferation were also implicated in our autoimmune V2G set (**Figure 8**
538 **– Figure Supplement 2A, Supplementary File 16**). Genes implicated by V2G in activated CD4+

539 T cells were moderately enriched for genes known to control the production of IFN-G, but at the
540 individual disease level, only genes connected to CRO- and ATD-associated variants were
541 enriched for IFNG regulatory genes (**Figure 8C, Figure 8 – Figure Supplement 2A**). Genes
542 contacting CRO, PSO, RA, SLE, T1D and VIT variants were moderately enriched for TNF
543 regulatory genes, but the set of all V2G genes was not enriched for TNF genes. We also queried
544 the orthologs of our V2G-implicated genes from the international mouse phenotype consortium
545 (IMPC) database and identified 97 genes that when knocked out give an immune phenotype and
546 126 V2G genes that result in a hematopoietic phenotype (**Figure 8D**, **Figure 8 – Figure**
547 **Supplement 2B**, **Supplementary File 18**). This frequency of observed immune/hematopoietic
548 phenotypes represents a significant (adjP<0.05) ~30% enrichment over expected. This gene set
549 was also enriched for mortality, homeostasis/metabolism, growth/body size, skeleton, and
550 embryonic phenotypes (**Figure 8D**, **Supplementary File 18**).

551 An important application of this V2G approach is the identification of novel regulators of T cell
552 activation and their potential as drug targets, as nearly 20% of implicated genes have at least one
553 chemical modulator currently available (**Figure 8E**, **Supplementary File 19**). As shown above,
554 this approach validates genes that are well-studied regulators of T cell function, however, a
555 significant portion of implicated genes are not well-studied and are not currently known to regulate
556 T cell activation (**Figure 8F**). We observed a trend that genes expressed more highly in immune
557 tissues (GTEX) have in general been better investigated, and identified several less-studied
558 genes that could be novel targets, including the kinases GRK6, PTK6, SIK1, and MAP3K11, the
559 G protein-coupled receptors OXER1, GPR183, GPR18, and KISS1R, the acetylcholine receptor
560 CHRNB1, and the *de novo* purine pathway enzyme GART. To determine whether these V2G-
561 implicated genes are novel regulators of T cell activation, we used commercially available
562 pharmacologic modulators in dose-response assays of activation-induced T cell proliferation.
563 Stimulation of T cells in the presence of ligands for CHRNB1, KISS1R and OXER1 did not
564 significantly affect T cell proliferation, however, small molecules targeting GRK6, PTK6,

565 MAP3K11, GPR183, GART, and SIK1 inhibited T cell activation in the nanomolar to micromolar
566 range (**Figure 8G**). Together, these data show that maps of dynamic, 3D variant-gene chromatin
567 contacts in stimulated CD4+ T cells are able to identify genes with *bona fide* roles in T cell
568 activation.
569

570 **DISCUSSION**

571 By measuring dynamic changes in chromosome folding, chromatin accessibility, and gene
572 expression in naïve CD4+ T cells as a function of TCR-CD28 costimulation, we identified the
573 putative *cis*-regulatory landscape of autoimmune disease-associated genetic variation and
574 physically connected these elements to their putative effector genes. This and prior chromosome
575 capture-based studies show that most cRE and their variants interact with, and therefore have
576 the potential to regulate, more than one gene (median 5 in this study), supporting a scenario in
577 which multiple effector genes are operative at a GWAS locus. We validated a stretch of novel
578 distal elements predicted by our V2G approach as *bona fide* enhancers for the canonical immune
579 gene *IL2*, and showed that autoimmune-associated genetic variation at one of these elements
580 influences it's activity. We also conducted pharmacologic targeting experiments and compared
581 our results with CRISPR-based studies to validate sets of *bona fide* effector genes with a
582 confluence of multiple orthogonal lines of evidence supporting their role in CD4+ T cell activation.
583 We also observe that the vast majority (87-95%) of GWAS effector genes predicted by
584 chromosome capture-based approaches are not the genes nearest to the sentinel SNPs queried
585 in this study, while roughly one-third of eGenes predicted by eQTL approaches are the nearest to
586 a sentinel. This and the low concordance between 3D chromatin vs. eQTL eGene predictions is
587 consistent with the view that eQTL-based and GWAS-based approaches inherently implicate
588 different types of genes. While eQTLs cluster strongly near transcription start sites of genes with
589 simple regulatory structures and low enrichment for functional annotations, GWAS variants are
590 generally far from the TSS of genes with complex regulatory landscapes[84]. Moreover, eGene

591 discovery by eQTL studies using scRNA-seq approaches are significantly limited by the low gene
592 detection power inherent to these methods, particularly in rare cell types.

593 Our systematic comparison of 3D chromatin-based target gene nomination studies in human
594 CD4+ T cells revealed significant variability between datasets, with the highest concordance
595 exhibited by the Javierre and Burren datasets (37%) and the next highest exhibited by our V2G
596 and the Javierre dataset (24%). Although we harmonized the comparisons as much as possible
597 by subjecting each dataset to the same HiC loop calling and GWAS integration steps, there are
598 several potential sources for the observed discrepancy between the studies. The modes of
599 stimulation are largely comparable, but timepoints and donors varied, and ours was the only study
600 that sorted naïve CD4+ T cells prior to stimulation. The higher concordance among promoter-
601 capture datasets compared to our HiC dataset is likely due to their lower resolution compared to
602 our HiC and their greater sequencing depth focused at promoters compared to HiC. The lower
603 resolution of HindIII-based capture-HiC results in loops called to the wrong promoters[24], and will
604 miss distal SNP interactions at any promoters excluded from the capture set. While HiC is
605 unbiased in this regard, high resolution HiC will fail to call some SNP-promoter loops because the
606 sequencing depth is spread across the whole genome instead of focused at promoters. However,
607 despite variation between studies, the results show the clear value of tissue-matched chromatin
608 conformation maps compared to tissue-averaged HiC for understanding the complex genetic and
609 epigenetic mechanisms that regulate gene expression and for predicting autoimmune effector
610 genes.

611 Why do the V2G approaches not capture a larger proportion of genes from the human inborn
612 errors in immunity 'truth set'? Low recall/sensitivity could result from the fact that a substantial
613 portion of the mutations in the full HIEI gene set result in immunodeficiency but not autoimmune
614 phenotypes. However, restricting the HIEI truth set to only those disease gene mutations that
615 result in an autoimmune phenotype (143 genes) only increased recall from 16% to 17.5%,
616 suggesting that this is not a major factor reducing sensitivity. Also, our method uses GWAS

617 signals as an input, and unlike GWAS for height, BMI, blood traits, etc., most autoimmune GWAS
618 signals are likely not saturated, which limits discoverability. Another reason for reduced sensitivity
619 is the likelihood that many of the genes in the HIEI truth set operate in cells other than CD4+ T
620 cells, and consistent with this, when the Javierre pcHiC V2G is extended to all immune cell types
621 tested, recall increases from 17% to 27%. This observation emphasizes how limited inclusion cell
622 types and states in a study can significantly limit the power to detect autoimmune effector genes.
623 Why are so few of our 3D chromatin-based V2G genes present in the HIEI truth set? Low precision
624 could be due to false positives in the V2G gene set; *i.e.*, contacts between disease-associated
625 cRE and genes that are not in fact involved in disease susceptibility. This does not argue *per se*
626 against the biological relevance of the cRE or the gene, only that the linkage to disease
627 susceptibility is not relevant. Alternatively, the absence of chromatin-GWAS-implicated genes in
628 the truth set could be a false negative from the point of view of the truth set; *i.e.*, monogenic,
629 disease-causing mutations in these genes may exist in the human population, but have not yet
630 been discovered. For example, in the year 2000 approximately 100 monogenic mutations causing
631 human inborn errors of immunity were known, but this increased by ~11 disease mutations per
632 year until 300 human inborn errors of immunity had been identified by 2017. The advent of next-
633 generation exome/genome sequencing and the COVID-19 pandemic resulted in an increase in
634 the rate of HIEI discovery between 2018-2021 to ~40 per year to the current recognized HIEI of
635 ~450. If we assume a conservative continued rate of HIEI discovery of 25 disease genes per year,
636 the next 10 years will show us an additional ~250 disease genes that are not currently contained
637 in the truth set.

638 Many of the genes identified in this 3D epigenomic V2G screen have known roles in T cell
639 activation and function. An example is *IL2*, and we used the resulting maps to identify and validate
640 a stretch of previously unknown distal enhancers whose activity is required for *IL2* expression and
641 is influenced by autoimmune genetic variation. Another example is the phosphatase DUSP5 that
642 regulates MAPK signaling during T cell activation[85,86]. However, roles for many of the genes

643 implicated here in T cell activation are not known. For example, one of the top implicated genes,
644 PARK7, is a deglycase studied in the context of Parkinson's disease, but has a recently been
645 shown to modulate regulatory T cell function[87]. The orphan G protein-coupled receptor GPR108
646 is another top gene uniquely implicated by our chromatin-based V2G that has not been studied
647 in T cells, but was identified in a recent CRISPR screen for genes affecting IL-2 levels[23]. Also co-
648 implicated by our study and recent CRISPR screens are the cannabidiol receptor GPR18[88] and
649 the purine biosynthetic enzyme GART[89]. The GPR18 agonist arachidonyl glycine inhibited CD4+
650 T cell activation above 10 uM, while the GPR18 antagonist O-1918 slightly enhanced T cell
651 activation. This GPR was implicated by both chromatin- and eQTL-based approaches.
652 Antagonism of GART, an enzyme we previous identified as a V2G effector gene in COVID19
653 severity[58], with the FDA-approved drug lometrexol inhibited T cell activation in the 10 nM range.
654 Antagonism of GRK6, a member of the G-coupled receptor kinase family associated with insulin
655 secretion and T2D susceptibility[90], and PTK6, an oncogenic kinase studied in the context of
656 cancer[91], led to inhibition of T cell activation in the nM to uM range. These targets were implicated
657 by the other chromatin-based approaches but not by eQTL. Inhibition of MAP3K11, a kinase that
658 facilitates signaling through JNK1 and IkappaB kinase in cancer[92,93] inhibited stimulation-induced
659 CD4+ T cell proliferation in the 100 nM range, as did 25-hydroxycholesterol, a ligand of the G
660 protein-coupled receptor GPR183 that was implicated by both chromatin- and eQTL-based
661 approaches. SIK1 is a member of the salt-inducible kinase family uniquely implicated in our study
662 that negatively regulates TORC activity[61], and a small molecule SIK1 inhibitor potently
663 antagonized stimulation-induced CD4+ T cell activation in the pM range.

664 Our integration of high-resolution Hi-C, ATAC-seq, RNA-seq and GWAS data in a single
665 immune cell type across multiple activation states identified hundreds of autoimmune variant-
666 gene pairs at ~half of all GWAS loci studied, and application of this technique to additional immune
667 cell types will likely identify effector genes at many of the remaining loci. This study highlights the
668 value of chromosome conformation data as a powerful biological constraint for focusing variant-

669 to-gene mapping efforts[94], and shows that dynamic changes in the spatial conformation of the
670 genome that accompany cell state transitions alter gene expression by exposing promoters to a
671 varying array of *cis*-regulatory elements, transcription factors, and genetic variants.
672

673 **MATERIALS AND METHODS**

674 T cell isolation and *in vitro* stimulation

675 Human primary CD4+ T cells were purified from the apheresis products obtained from healthy,
676 screened human donors through University of Pennsylvania Human Immunology Core (HIC).
677 Naïve CD4+ T cells were purified using EasySepTM human naïve CD4+ T cell isolation kit II
678 (STEM cells Technologies, cat#17555) by immunomagnetic negative selection as per
679 manufacturer's protocol. Isolated untouched, highly purified (93-98%) naïve human CD4 T cells
680 were activated using anti-CD3+anti-CD28 Dynabeads (1:1, Thermofisher scientific, cat # 11161D)
681 for 8-24 hours. Cells were then used to prepare sequencing libraries. The human leukemic T cell
682 line Jurkat was obtained from ATCC, cloned by limiting dilution, and clones with high activation-
683 induced secretion of IL-2 were selected for further study. Single-cell mouse lymphocyte
684 suspensions were prepared from spleen and lymph nodes isolated from 6-week-old female wild-
685 type or *Il2*-83cRE-ko mice (C57BL6 background). Mouse CD4+ T cells were purified by negative
686 selection using a CD4+ T cell isolation kit (Miltenyi Biotec, cat. # 130-104-454). For CD8 depletion,
687 CD8+ T cells were positively stained using CD8a (Ly-2) microbeads (Miltenyi Biotec, cat. # 130-
688 117-044) and the negative flow-through fraction was collected. Tregs were purified from the
689 mouse lymphocytes using a CD4+CD25+ regulatory T cell isolation kit (Miltenyi Biotec, cat. # 130-
690 091-041). The purity of the isolated cells was checked by flow cytometry, showing approximately
691 95% cell purity. For iTreg induction, CD4+ CD25- T cells (1x 10^6) were activated in a 24-well plate
692 pre-coated with anti-CD3 (1 µg/mL) and the cells were cultured for 72 hours in RPMI culture
693 medium with soluble anti-CD3 (0.5 µg/mL), IL-2 (25 units/mL), TGF-beta (3 ng/mL), anti-IFN-
694 gamma (5 µg/mL), and anti-IL-4 (5 µg/mL). iTreg induction cultures were also set up without

695 adding IL-2. iTreg cultures were harvested after 72 hours, and cells were stained for Foxp3
696 expression.

697 *Il2*-83-cRE ko Mice

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

707 T7_guideRNA#1_scaffold#1:

708 TTAATACGACTCACTATAGGTTTTCCACGGATCTGCTCGGGTTTTAGAGCTAGAAATAGC

709 T7_guideRNA#1_scaffold#2:

710 TTAATACGACTCACTATAGGTGCTTTCTAGGTGAAGCCCCGTTTTAGAGCTAGAAATAGC

711 T7_guideRNA#1_scaffold#3:

712 TTAATACGACTCACTATAGGTCATTTGAGCCTAACTACTCGTTTTAGAGCTAGAAATAGC

713 Along with the above-cited sequences, a reverse primer sequence
714 (AGCACCGACTCGGTGCCACT) from PX335 was used to amplify a PCR-generated template
715 for sgRNA *in vitro* transcription. The PCR product (~ 117 bp) was verified on a 2% agarose gel
716 and then gel-purified using the QIAQuick gel extraction kit. The T7 sgRNA PCR product (500 ng)
717 was used as a template for *in vitro* transcription with the T7 High Yield RNA synthesis kit (NEB,
718 cat # E2040S). The reaction was incubated at 37°C for 4 hours, and then the sgRNA was purified
719 using the MEGAclear kit (Life Technologies, cat # AM1908) according to the kit protocol. The
720 sgRNA was eluted with elution buffer preheated to 95°C. The eluted sgRNA was centrifuged at

721 13,000 rpm for 20 minutes at 4°C, and the suspension was transferred into a new RNase-free
722 tube. The sgRNA quality was checked by a bioanalyzer using an RNA nano Chip. The sgRNA
723 was diluted to 500 ng/µl in injection buffer (sterile 10 mM Tris/0.1 mM EDTA, pH 7.5) and stored
724 in a -80°C freezer until use. An injection mixture (final volume 30 µl) was prepared in injection
725 buffer by mixing 500 ng/µl of each of the sgRNAs (left and right) with Cas9 mRNA (1 µg/µl, TriLink,
726 cat # L-6125). Fertilized eggs collected from B6/129 mice were microinjected at the CHOP
727 transgenic core and transferred into pseudo-pregnant B6 females. The pups were genotyped
728 using primers flanking the targeted sequence (Forward primer: TTAGGACCTCACCCATCACAA
729 and reverse primer: CATGCCCAGCTACTCTGACAT). The PCR product was cloned into the
730 Promega PGEM T Easy TA cloning vector, and plasmid DNA was Sanger sequenced to
731 determine the size of the deletion. The targeting resulted in mice with a 500 bp and 583 bp deletion
732 at the targeted -83 CNS site and these mutant mice showed same phenotype. The *Il2*-83-cRE
733 mutant heterozygous male mice were backcrossed with B6 females for 10 successive
734 generations.

735 ELISA

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

740 *In vivo* ovalbumin immunization

741 Eight week-old female WT and Il2-CNS-83 KO mice were injected intraperitoneally with 50 µg of
742 ovalbumin (Sigma cat # A5503-5G) mixed with IFA (Sigma cat # F5506). Blood was collected
743 from the immunized mice after 10 days, and serum was separated. The level of ovalbumin-
744 specific IgG in the serum was determined by ELISA using an ovalbumin-specific IgG ELISA kit
745 purchased from MyBiosource (cat # MBS763696). The immunized mice were sacrificed on day
746 10 of immunization, and spleen and lymph nodes were collected. Splenocytes and lymph node

747 cells were stained with CD4 BV785, CD25 BV650, CD44 Percp Cy5.5, CD62L APC eFL780,
748 CXCR5 BV421, PD-1 APC, Bcl-6 PE antibodies. TFH frequency was determined by flow
749 cytometry analysis.

750 *In vivo* inflammatory colitis model

751 Conventional CD4+CD25-ve T cells and CD4+CD25+ Tregs were purified from spleens and
752 lymph nodes of male wild-type or Il2-CNS-83 ko mice (6-8 weeks of age). T cells were transferred
753 into Rag1-ko male mice (5 per group) by retro-orbital injection of 1 million wild-type or Il2-CNS-83
754 ko Tconv cells alone or along with (0.25×10^6) wild-type Tregs. Experimental Rag1-ko recipients
755 were weighed 3 times per week and scored for IBD-induced clinical symptoms. Mice were
756 sacrificed 72 days post-transfer, and spleen, mesenteric lymph nodes, and colon were collected.
757 Single-cell suspensions were prepared, cells were stained for CD4, CD8, CD25, CD44, and
758 Foxp3, and analyzed by flow cytometry. The absolute T cell count was estimated using cell count
759 and cell frequency derived from the flow cytometry analysis.

760 RNA-seq library generation and sequencing

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

773 ATAC-seq library generation and sequencing

774 A total of 50,000 to 100,000 sorted cells were centrifuged at 550 g for 5 min at 4 °C. The cell pellet
775 was washed with cold PBS and resuspended in 50 µL cold lysis buffer (10 mM Tris-HCl, pH 7.4,
776 10 mM NaCl, 3 mM MgCl2, 0.1% NP-40/IGEPAL CA-630) and immediately centrifuged at 550 g
777 for 10 min at 4 °C. Nuclei were resuspended in the Nextera transposition reaction mix (25 µL 2x
778 TD Buffer, 2.5 µL Nextera Tn5 transposase (Illumina Cat #FC-121-1030), and 22.5 µL nuclease
779 free H2O) on ice, then incubated for 45 min at 37 °C. The tagmented DNA was then purified using
780 the Qiagen MinElute kit eluted with 10.5 µL Elution Buffer (EB). Ten microliters of purified
781 tagmented DNA was PCR amplified using Nextera primers for 12 cycles to generate each library.
782 PCR reaction was subsequently cleaned up using 1.5x AMPureXP beads (Agencourt), and
783 concentrations were measured by Qubit. Libraries were paired-end sequenced on the Illumina
784 HiSeq 4000 platform (100 bp read length).

785 Hi-C library preparation

786 Hi-C library preparation on FACS-sorted CD4+ T cells was performed using the Arima-HiC kit
787 (Arima Genomics Inc), according to the manufacturer's protocols. Briefly, cells were crosslinked
788 using formaldehyde. Crosslinked cells were then subject to the Arima-HiC protocol, which utilizes
789 multiple restriction enzymes to digest chromatin. Arima-HiC sequencing libraries were prepared
790 by first shearing purified proximally-ligated DNA and then size-selecting 200-600 bp DNA
791 fragments using AmpureXP beads (Beckman Coulter). The size-selected fragments were then
792 enriched using Enrichment Beads (provided in the Arima-HiC kit), and then converted into
793 Illumina-compatible sequencing libraries with the Swift Accel-NGS 2SPlus DNA Library Kit (Swift,
794 21024) and Swift 2S Indexing Kit (Swift, 26148). The purified, PCR-amplified DNA underwent
795 standard QC (qPCR, Bioanalyzer, and KAPA Library Quantification [Roche, KK4824]) and was
796 sequenced with unique single indexes on the Illumina NovaSeq 6000 Sequencing System using
797 200 bp reads.

798 ATAC-seq data analysis

799 ATAC-seq peaks from libraries unstimulated and stimulated Naïve CD4+ T cells were called using
800 the ENCODE ATAC-seq pipeline (https://www.encodeproject.org/atac-seq/). Briefly, pair-end
801 reads from three biological replicates for each cell type were aligned to hg19 genome using
802 bowtie2[95], and duplicate reads were removed from the alignment. Narrow peaks were called
803 independently for each replicate using macs2[96] (-p 0.01 --nomodel --shift -75 --extsize 150 -B --
804 SPMR --keep-dup all --call-summits) and ENCODE blacklist regions (ENCSR636HFF) were
805 removed from peaks in individual replicates. Reproducible peaks, peaks called in at least two
806 replicates, were used to generate a consensus peakset. Signal peaks were normalized using
807 csaw[97] in 10kb bins background regions and low abundance peaks (CPM>1) were excluded from
808 the analysis. Tests for differential accessibility was conducted with the glmQLFit approach
809 implemented in edgeR[98] using the normalization factors calculated by csaw. OCRs with FDR <
810 0.05 and abs($\log_2$FC > 1) between stages were considered differentially accessible.

811 Hi-C data analysis

812 Paired-end reads from two replicates were pre-processed using the HICUP pipeline v0.7.4[99], with
813 bowtie as aligner and hg19 as the reference genome. The alignment .bam file were parsed to
814 .pairs format using pairtools v0.3.0 (https://github.com/open2c/pairtools) and pairix v0.3.7
815 (https://github.com/4dn-dcic/pairix), and eventually converted to pre-binned Hi-C matrix in .cool
816 format by cooler v0.8.10[100] with multiple resolutions (500bp, 1kbp, 2kbp, 2.5kbp, 4kbp, 5kbp,
817 10kbp, 25kbp, 40kbp, 50kbp, 100kbp, 250kbp, 500kbp, 1Mbp and 2.5Mbp) and normalized with
818 ICE method[101]. Replicate similarity was determined by HiCRep v1.12.2[100] at 10K resolution. For
819 each sample, eigenvectors were determined from an ICE balanced Hi-C matrix with 40kb
820 resolution using cooltools v0.3.2[102] and first principal components were used to determine A/B
821 compartments with GC% of genome region as reference track to determine the sign. Differential
822 TAD comparison was performed using TADcompare with the default settings for each
823 chromosome (v1.4.0)[103]. Finally, for each cell type, significant intra-chromosomal interaction loops
824 were determined under multiple resolutions (1kb, 2kb and 4kb) using the Hi-C loop caller Fit-Hi-

825 C2 v2.0.7[104] (FDR<1e-6) on merged replicates matrix. The consensus chromatin loops within
826 resolution were identified by combining all three stages. These sets of loops were used as
827 consensus for quantitative differential analysis explained below. The final consensus interaction
828 loops for visualization were collected by merging loops from all the resolutions with preference to
829 keep the highest resolution. Quantitative loop differential analysis across cell types was performed
830 on fast lasso normalized interaction frequency (IF) implemented in multiCompareHiC v1.8.0[105] for
831 each chromosome at resolution 1kb, 2kb and 4kb independently. The contacts with zero
832 interaction frequency (IF) among more than 80% of the samples and average IF less than 5 were
833 excluded from differential analysis. The QLF test based on a generalized linear model was
834 performed in cell type-pairwise comparisons, and p-values were corrected with FDR. The final
835 differential loops were identified by overlapping differential IF contacts with consensus interaction
836 loops.

837 Bulk RNA-seq data analysis

838 Bulk RNA-seq libraries were sequenced on an Illumina Novaseq 6000 instrument. The pair-end
839 fastq files were mapped to the genome assembly hg19 by STAR (v2.6.0c) independently for each
840 replicate. The GencodeV19 annotation was used for gene feature annotation and the raw read
841 count for gene feature was calculated by htseq-count (v0.6.1)[106] with parameter settings -f bam -
842 r pos -s yes -t exon -m union. The gene features localized on chrM or annotated as rRNAs, small
843 coding RNA, or pseudo genes were removed from the final sample-by-gene read count matrix.
844 Gene set variation analysis was performed using GSVA 1.42.0[107] with the MSigDB hallmark
845 geneset[108], with resulting scores analyzed using limma (limma_3.50.3)[109]. Low abundance peaks
846 (CPM>1) were excluded from the analysis. Testing for differential expression was conducted with
847 the glmQLFit approach implemented in edgeR[98]. Genes with FDR<0.05 and abs(log_2FC>1)
848 between stages were considered differentially expressed. Differential genes were then clustered
849 using k-means clustering. The number of clusters was determined using the elbow method on
850 the weighted sum of squares, where was set to k=5. Score for how similar each gene followed

851 the clusters expression pattern was determined by calculating pearson correlation coefficients
852 between each gene in the cluster and the cluster centroid.

853 Transcription factor footprinting and motif analysis

854 Transcription factor footprints were called using Regulatory Analysis Toolbox HINT-ATAC
855 (v0.13.0) with pooled ATAC-seq data for each stage and consensus peak calls[110]. The rgt-hint
856 footprinting was run with parameters –atac-seq, --paired-end, and organism=hg19 set. The output
857 footprint coordinates were subsequently matched using rgt-motifanalysis matching with
858 parameters --organism hg19 and –pseudocount 0.8 set. The JASPAR2020 position weight matrix
859 database was used to match footprints[111]. Differential analysis of TF binding across conditions
860 was performed using rgt-hint differential with parameters –organism hg19, --bc, --nc 24 using the
861 motif matched transcription factor footprints. An activity score is then calculated based on the
862 accessibility surrounding the footprint.

863 Partitioned heritability LD score regression enrichment analysis
864
865 Partitioned heritability LD score regression[112] (v1.0.0) was used to identify heritability enrichment
866 with GWAS summary statistics and open chromatin regions annotated to genes. The baseline
867 analysis was performed using LDSCORE data
868 (https://data.broadinstitute.org/alkesgroup/LDSCORE) with LD scores, regression weights, and
869 allele frequencies from the 1000G phase 1 data. The summary statistics were obtained from
870 studies as described in **Supplementary File 14** and harmonized with the munge_sumstats.py
871 script. Annotations for Partitioned LD score regression were generated using the coordinates of
872 open chromatin regions that contact gene promoters through Hi-C loops for each cell type. Finally,
873 partitioned LD scores were compared to baseline LD scores to measure enrichment fold change
874 and enrichment p-values, which were adjusted with FDR across all comparisons.

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

876 95% credible sets were determined as previously described[113]. Briefly, P values from GWAS
877 summary statistics were converted to Bayes factors. Posterior probabilities were then calculated

878 for each variant. The variants were ordered from highest to lowest posterior probabilities added
879 to the credible set until the cumulative sum of posterior probabilities reached 95%. Variants in the
880 95% credible set were then intersected with the CD4+ T cell promoter interacting region OCRs
881 from the three timepoints using the R package GenomicRanges (v1.46.1)[114].

882 Genomic reference and visualizations

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

887 ABC model predictions

888 We used the ABC model (PMID: 31784727) using the our Tcell ATAC-seq data (unstimulated or
889 24 hours stimulated) generated from Naive CD4+ T cells merged acrossed replicates. For
890 H3K27ac we retrieved paired fastq files from H3K27ac MINT-ChIP data from ENCODE for resting
891 (unstimated) and activated (36hr stimulated) CD4 T cells derived from thymus. The deduplicated
892 bam files were used as signal files for the ABC model, the consensus peak set were used as the
893 input set of enhancers. The accession numbers for unstimulated cells are: ENCFF732NOS,
894 ENCFF658TLX, ENCFF605OGN, ENCFF012XYW, ENCFF407QZP, ENCFF556QWM,
895 ENCFF888IHV, ENCFF260FUF, ENCFF550EVW, ENCFF747HGZ, ENCFF067XWQ,
896 ENCFF939YBH, ENCFF749LGW, ENCFF788LWV, ENCFF610FRB, ENCFF538GQX,
897 ENCFF355PCA, ENCFF040OMT, ENCFF878SHZ, ENCFF210WNQ, ENCFF849ZJH,
898 ENCFF925ARM, ENCFF112YNU, ENCFF808JFV, ENCFF448KHE, ENCFF591BFV,
899 ENCFF366UTF, ENCFF172EWR. The accession numbers for stimulated cells are:
900 ENCFF442DFF, ENCFF870ZYC, ENCFF189UPS, ENCFF725KSA, ENCFF618YAV,
901 ENCFF765AJK, ENCFF854QUS, ENCFF180TKV, ENCFF761RCK, ENCFF449EAE,
902 ENCFF915VPK, ENCFF575XFH, ENCFF227AGX, ENCFF810XFH, ENCFF367OMH,
903 ENCFF114BBZ, ENCFF855XPP, ENCFF132AKN, ENCFF663MHJ, ENCFF244TPB,

904 ENCFF230RIJ, ENCFF648FPC, ENCFF797CIK, ENCFF096JPW, ENCFF704KMT,
905 ENCFF385POD, ENCFF395RIG, ENCFF043RRJ, ENCFF027RNR, ENCFF333RGD,
906 ENCFF048PMV, ENCFF887XSA, ENCFF188CNA, ENCFF614IZY, ENCFF984TTZ,
907 ENCFF977RDC, ENCFF393EWO, ENCFF781VQT, ENCFF638UCZ, ENCFF873QZH,
908 ENCFF553JPM, ENCFF589HRL, ENCFF721ZCN, ENCFF496HYV, ENCFF455WBD,
909 ENCFF679UTJ, ENCFF200MYN, ENCFF459ARH.
910 The retrieved files were trimmed/preprocessed using fastp with the default settings. Following
911 this, matched R1 and R2 files were concatenated to a single R1/R2 file pair. The reads were
912 aligned to hg19 and duplicates were removed as described for the ATAC-seq data analysis. The
913 deduplicated bam files were used as input to the ABC model. We ran the ABC model with the
914 default settings for hg19 and provided reference files for the blacklist, promoter annotation,
915 chromosome sizes reference, and set of ubiquitiously expressed genes. The HiC data used was
916 the KR normalized averaged HiC dataset that was derived from a set of cell lines (GM12878,
917 NHEK, HMEC, RPE1, THP1, IMR90, HU- VEC, HCT116, K562, KBM7;
918 average_hic.v2.191020.tar.gz) at 5kb resolution then scaled by powerlaw distribution as
919 described in the original manuscript. The default threshold of 0.2 was used to link enhancers to
920 genes. The resulting elements were intersected with the credible set list used in this study for
921 precision-recall analyses.
922 Precision-recall analyses against HIEI genes
923 To benchmark our autoimmune effector gene predictions against prior eQTL and chromatin
924 capture studies, we curated a list of 449 expert curated genes from Human Errors in Inborn
925 Immunity 2022, where mutations have been reported to cause immune phenotypes in humans[65].
926 Most of these genes were identified through rare loss of function mutations in the coding region.
927 In addition to this we also took a subset of 145 genes reported to specifically result in autoimmune
928 phenotypes. These two sets of genes were treated as “gold standard” sets of genes to compare
929 different approaches to predict GWAS effector genes. We curated the results of several other

930 chromatin conformation or eQTL based methods of assigning variants to genes to compare with
931 our study. For the chromatin confirmation studies, we obtained loop calls from the associated
932 publications and intersected with the list variants in the 95% credible set from 15 GWAS used to
933 predict effector genes using GenomicRanges (v1.46.1)[114]. For eQTL studies, we obtained the
934 summary stats files and identified eGenes linked to any member of the credible set reported as
935 an eQTL in the transcriptome wide association analysis (TWAS; p-value < 1e-4). Precision was
936 calculated as the ratio of the number predicted genes in the truth set to the total number of
937 predicted genes, while recall was calculated as the ratio of the number predicted genes in the
938 truth set to the total number of genes in truth set.

939 Colocalized eQTL comparisons

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

948 International mouse phenotyping consortium comparisons

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

953 Identification of pharmacological agents

954 We queried the Drug-Gene Interaction Database with the set of V2G implicated genes for
955 chemicals using rDGIdb (v1.20.0)[118]. To identify the number of papers for each gene with at least

956 one drug annotated to target it, we queried pubmed titles and abstracts using the R package
957 RISmed (v2.3.0) with each gene's name and either "autoimmune" and the list of autoimmune
958 diseases (**Supplementary File 17**). A score to approximate expression specificity was computed
959 using the sum GTEX median expression values (v8) for whole blood or spleen divided by other
960 tissues[119].

961 Lentiviral-based CRISPR/CAS9 targeting in Jurkat cells

962 LentiCRISPRv2-mCherry vectors encoding gRNA-CAS9 and the fluorescent reporter mCherry
963 were used for Jurkat targeting. CRISPR guide RNAs (sgRNA) targeting human IL-2-21 intergenic
964 -46, -51, -80, -83, -85, -122, -128 CNS regions were designed using http://crispr.tefor.net and
965 cloned into lentiCRISPRv2-mCherry. Empty vector without gRNA insert was used as a control.
966 Below is the list of CRISPR gRNA for causing deletion of human IL2-21 intergenic regions:

967 CNS region	CRISPR gRNA	
968 -46 CNS	gRNA #1	AGGATGCCTACCTCCAAATG
969	gRNA # 2	AGGTGACAACATTTAGTCAG
970 -51 CNS	gRNA #1	GGCAACGAAATTCACTGTGA
971	gRNA # 2	ATTCTAACAGGAATCATTCG
972 -80 CNS	gRNA # 1	GTTCTACCTATGCCGCATTG
973	gRNA # 2	GAGATTTACTCAGTCCAATG
974 -83 CNS	gRNA # 1	GTGACAAGCATGACTCTACA
975	gRNA # 2	GTGATGGTGAATTAAGCTGA
976 -85 CNS	gRNA #1	AGGGTTTTCTAGTTACGAGA
977	gRNA # 2	ATGGTTAGTTAGCTCCCAAG
978 -96 CNS	gRNA#1	TGGGAAAAACATCTTACCTG
979	gRNA # 2	TGGCCCATGAACCATCAAAG
980 -122 CNS	gRNA# 1	GTTATTAATCTAAGCGGAGA
981	gRNA# 2	GGAAGTTAGGCAGTCAATCG

982 -128 CNS gRNA#1 CTTCAATCATTGCATTCCAC

983 gRNA# 2 TGACACCACCCCTGCTTGAG

984 HEK 293T cells were grown in RPMI1640 complete medium (RPMI1640 + 1X P/S, 1x L-Glu, 10%
985 FBS), 37C, 7%CO2. 293T cells were transfected with 10 ug of lenti-CRISPR-V2-CRE construct
986 along with packaging plasmid 6 ug of PsPAX2 (Addgene, Cat #12260) and 3.5 ug of PmD2.G
987 (Addgene, Cat #12259) using Lipofectamine 2000 transfection reagent (Invitrogen cat #
988 11668019). After 6 hours, the transfection medium was replaced with complete culture medium.
989 Transfected cells were incubated at 37C for 48-72 hours in a cell culture incubator. and then the
990 Lentiviral supernatants were harvested and spun at 300g for 5 minutes to remove cellular debris.
991 Lentiviral supernatants were concentrated using Lenti-X^{TM} Concentrator (Takara Bio, Cat #
992 631232) and then centrifuged at 1500g for 30 minutes at 4°C and supernatant was discarded.
993 The lentiviral pellet was resuspended at a ratio of 1:20 of the original volume using RPMI media
994 and concentrated virus supernatant aliquots were prepared and stored until use at -80°C. To
995 achieve high transduction efficiency, the viral supernatant was titrated in Jurkat cells through
996 transduction using various dilutions of the viral supernatants and transduction efficiency was
997 determined by mcherry expression analyzed through flow cytometry. Jurkat cells were seeded in
998 a 24 well plate at 0.5 x106/well in culture media, viral supernatant with 8 ug/mL of polybrene was
999 added to each well. Spinfection was performed for 90 min. at 2500 rpm, and transduced cells
1000 were equilibrated at 37C for ~6 hrs, followed by incubation at 37C 5%CO2 for ~72 hours culture.
1001 Transduced Jurkat Cells were then harvested and stimulated using PMA (15 ng/mL) + Ionomycin
1002 (1 uM) + human anti-CD28 (2 ug/ml), BioXcell cat # BE0248 in 96 well culture plates (TPP, cat #
1003 92097) in triplicates. Cell culture supernatants were collected at the end of culture and analyzed
1004 for IL-2 by ELISA using a kit (cat # 88-7025-76) purchased from Thermofisher Scientific.

1005 <u>gRNA-CAS9 RNP-based targeting in primary human CD4+ T cells</u>

1006 Primary Human CD4 T cells derived from 5 normal healthy donors were obtained from Human
1007 Immunology core (University of Pennsylvania). Alt-R S.p. HiFi Cas9 Nuclease V3 cat # 1081061

1008 CAS9 and following list of Alt-R® CRISPR-Cas9 sgRNA targeting IL-2-21 CRE were purchased

1009 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

1010

1011 Primary Human CD4 T (5-10e6) were incubated with sgRNA and CAS9 protein complex and
1012 electroporation was done using P3 Primary Cell 4D-Nucleofector™ X Kit L (Lonza, cat # V4XP-
1013 3024) and Lonza 4D nucleofector system. B2M gene was CRISPR targeted as a positive control.
1014 As per manufacturer's protocol cells were electroporated pulse code Fl-115 in 100ul cuvette
1015 format. After nucleofection, cells were allowed to rest in the complete media for ~2 days. Cells
1016 were then harvested, washed with PBS and aliquots of cells were used for further experimentation
1017 such as flow staining and cell activation. Primary human CD4 T Cells (0.1e6/well) were seeded
1018 in triplicates (for each experimental condition) in 96 well plate format in RPMI complete medium
1019 and stimulated using ImmunoCult™ Human CD3/CD28 T Cell Activator (STEM cells
1020 Technologies, cat # 10971). Cell culture supernatants were collected at 4h of stimulation and
1021 stored at -80C until assayed. IL-2 ELISA was performed using Thermofisher Scientific IL-2 Human
1022 Uncoated ELISA Kit with Plates, cat # 88-7025-76.

1023 Recombinant reporter assays

1024 The *IL2* URR was cloned into Xho I and Hind III sites of PGL4 Luc vector (Promega) and then
1025 ~500 bp individual sequence representing *IL2-IL21* intergenic CNS at -46, -51, -80, -83, -85, -128
1026 were cloned upstream to IL-2 URR at the Xho I site of the pGL4 vector. Primary human CD4 T
1027 cells obtained from 5 normal healthy donors were transfected with PGL4-cRE-URR constructs.
1028 Briefly, primary human CD4 T cells were activated with anti-CD3+ anti-CD28 Dynabeads (1:1)
1029 (Thermofisher scientific, cat # 11161D) + IL-2 overnight and then 1 million aliquots of cells
1030 (triplicates) were electroporated using Nucleofector 2b, human T cell Nuclefactor kit (Lonza VPA
1031 # 1002, program # T-020) with 2ug of PGL4 firefly vector constructs along with 0.2 ug of PGL4

1032 Renilla vector; cells were allowed to rest ON in RPMI + IL-2 and then re-stimulated with plate-
1033 bound anti-CD3+ anti-CD28 (2 ug/ml each) for 5 hrs. Dual luciferase assay was performed with
1034 cell lysate prepared using Promega dual luciferase assay kit. Cell lysate was prepared in PLB as
1035 per manufacturer's protocols and then firefly and renilla luciferase activities were analyzed by
1036 spectramax ID5 (Molecular Devices). Firefly luciferase activity was normalized against the internal
1037 control renilla luciferase activity.

1038 Pharmacologic validation of novel T cell activation regulatory genes

1039 The drugs lometrexol (GART antagonist), 25-hydroxycholesterol (GPR183 ligand), epinephrine &
1040 norepinephrine (CHRNB agonists), and arachidonoyl glycine (GPR18 agonist) were purchased
1041 from Sigma. The GPR18 antagonist O-1918 was purchased from ChemCruz. CEP1347
1042 (MAP3K11 antagonist), kisspeptin 234 (KISSIR ligand) and the PTK6 antagonist tilfrinib were
1043 purchased from Tocris. The GRK6 antagonist GRK-IN-1 was purchased from DC Chemicals. The
1044 SIK1 antagonist HG-91-9-1 was purchased from Selleckchem. The OXER1 agonist 5-Oxo-ete
1045 was purchased from Cayman chemicals. The drugs were dissolved in DMSO or ethanol as
1046 suggested by the vendor and working stocking concentrations were prepared in RPMI1640
1047 medium or PBS. The drug effect on T cell proliferation was assayed using murine lymphocytes
1048 cultured under TCR and CD28 activation conditions. Spleen and lymph nodes were collected from
1049 6-8 weeks old female C57BL/7 mice and single cell lymphocyte cell suspensions were prepared
1050 in RPMI 1640 complete medium. 20 million lymphocytes were labeled with CFSE and re-
1051 suspended in RPMI 1640 medium. Cells (0.5 million labeled cells/well) were loaded into 48 well
1052 culture plate and activated with mouse anti-CD3 and anti-CD28 agonistic antibodies (1μg mL
1053 each). Drugs at the indicated concentrations were added in the culture medium and both
1054 untreated and drug treated cell cultures were incubated at 37°C for 72 hours in a cell culture
1055 incubator. Cells were harvested after 3 days of culture, washed with PBS and then stained with
1056 live-dead aqua dye. After washing with FACS buffer (PBS containing 2% FBS), cells were stained

1057 with fluorochrome conjugated antibodies CD4-APC, CD8-PB, CD44-Percp Cy5.5 and CD25-
1058 BV650. Stained cells were analyzed on a Beckman Coulter Cytoflex S flow cytometer. The
1059 division profile of CD4+ CFSE+ T cells were gated on live populations. The flow data was
1060 analyzed using Flowjo10 software and the number of divided CD4+ T cells were determined as
1061 described previously[120].
1062
1063 **FIGURE LEGENDS**
1064 **Figure 1: Defined gene expression dynamics throughout activation of naïve CD4+ T cells.**
1065 (**A**) Diagram of study design: RNA-seq, ATAC-seq, and HiC libraries were prepared from sorted
1066 CD4+CD62L^{hi}CD44^{lo} naïve T cells (donor N=3) prior to or following 8hr or 24hr stimulation with
1067 anti-CD3/28 beads. cREs are classified to genes allowing for interrogation of the pattern of
1068 expression, accessibility, and chromatin structure changes of autoimmune-associated variants.
1069 (**B**) Heatmap showing normalized expression of known markers of T cell activation. (**C**) K-means
1070 clustering of differentially expressed genes into five groups using the elbow and within-cluster-
1071 sum of squares methods to select the number of clusters. The centroid of the cluster is depicted
1072 as a black line, and members of the cluster are depicted as colored lines. Color indicates the
1073 Pearson's correlation coefficient between the gene and the cluster centroid, with red indicating
1074 higher correlation, and blue lower correlation.
1075 **Figure 1 – Figure Supplement 1: Library sequencing reproducibility and expression**
1076 **clustering.** (A) PCA of RNA-seq libraries. (B) PCA of ATAC-seq libraries. (C) Stratified correlation
1077 (SCC) of HiC libraries from two donors by stage of activation. (D) Volcano plot of RNA-seq data
1078 for CD4+ T cells unstimulated vs. 8 hr stimulation (top) and 8 hr vs. 24 hr stimulation (number of
1079 DEG indicated). (E) Elbow plot of the within group sum of squares used to determine optimal
1080 cluster number. (F) Genes per expression cluster. (G) Centroids +/- standard deviation for each
1081 cluster.

1082 **Figure 2: Chromatin accessibility and architecture changes across three stages of T cell**
1083 **activation.** (**A**) Volcano plots depicting the $\log_2$ change in accessibility in all reproducible OCRs
1084 (present in 2 of the 3 replicates) compared to the $-\log_{10}$(FDR), significant points are indicated in
1085 red (FDR < 0.05; abs($\log_2$FC) > 1). (**B**) Genomic regions were classified by membership in A vs.
1086 B compartments at each time point. The pie chart depicts regions with differential A/B
1087 compartment assignment in response to activation. (**C**) TAD structure was determined at each
1088 time point, and regions exhibiting a shift, split/merge, change in strength, or more complex
1089 rearrangements are depicted in the pie chart. (**D**) Loop calls identified from HiC data for each
1090 timepoint called at three resolutions (1kb, 2kb, 4kb bins). (**E**) Differential loop-calls called across
1091 all resolutions. (**F**) Density plots of cRE accessibility and gene expression separated by whether
1092 contact frequency increased, decreased, or remained stable from the transition from unstimulated
1093 to 8 hour activation and 8 hour to 24 hour activation.
1094 **Figure 2 – Figure Supplement 1: Characterization of CD4+ T cell epigenomic data**. (A)
1095 Accessibility of OCR and (B) expression of genes located in
1096 A vs. B compartments (log2 fkpm). Differential accessibility (C) and expression (D) of cRE-gene
1097 pairs in regions exhibiting differential TAD structure. (E) Enrichment of promoter interacting region
1098 OCR for annotated regulatory elements in CD4+ T cells from the epigenome roadmap project
1099 (E038_15). (F) Overlap of cRE-promoter loops defined in this study to loops defined by Burren et
1100 al. in a prior promoter capture Hi-C study. (G) Distribution of OCR contacted per gene (top, median
1101 = 8 OCRs per gene). Distribution of genes contacted per OCR (bottom, median = 3). (H)
1102 Differential promoter-OCR connectivity by HiC as a function of differential OCR accessibility
1103 (central bar = median, boxplot edges = 25-75 quantiles, whiskers = 1.5x inter-quantile range).
1104 **Figure 2 – Figure Supplement 2: Characterization of chromatin stripes during CD4+ T cell**
1105 **activation**. (A) Stripes called per timepoint. (B) Proportion of SNPs annotated in A vs B
1106 compartments. Accessibility of OCR (C) and expression of genes (D) located in anchor (small
1107 more interactive region of stripe), stripe excluding the anchor (stripe), or outside stripes.

1108 **Figure 2 – Figure Supplement 3: Large-scale epigenetic changes at dynamically expressed**
1109 **genes during CD4+ T cell activation.** HiC contact frequency matrix (heatmap), chromatin
1110 accessibility (grey), and loop calls (red) for (A) GEM, (B) IRF4, (C) KLF2, and (D) DPEP2. The
1111 IRF4 matrix is truncated at by end of the chromosome annotation, and inset shows the expression
1112 (TPM) at each time point for each gene.

1113 **Figure 3: Transcription factors prediction potential regulators of chromatin status and**
1114 **expression changes**. Footprints were annotated to each TF motif by sequence matching to the
1115 JASPAR motif database, and the average accessibility of the region surrounding each motif was
1116 used as a measure of activity at each time point. Scatterplots depict the change in accessibility
1117 for each TF motif (activity score) and $\log_2$FC of TF gene expression between unstimulated and 8
1118 hour activation (**A**) or 8 hour and 24 hour activation (**B**). TF were classified as differentially
1119 expressed (orange), differentially active (brown), both (red), or neither (grey). Dot size indicates
1120 the number of predicted footprints occupied by each motif. (**C**) Average accessibility (normalized
1121 by depth and motif number) in a 200bp window surrounding motif footprints for FOSL2::JUNB,
1122 Smad2::Smad3, MAZ, and KLF2 from three timepoints (unstim: green; 8hr: blue; 24hr: red). (**D**)
1123 Z-score of TF motif enrichment for cRE connected to the five clusters compared to all OCR.
1124 Lighter color indicates higher specificity of enrichment for that TF cluster. (**E**) Connections
1125 between different TFs based on physical interactions and predicted regulation-based co-
1126 expression determined by GENIE3 for the three timepoints. Node color indicates expression
1127 (TPM; darker = higher expression, lighter = lower expression), edge color reflects confidence in
1128 the interaction called by GENIE3 (darker higher confidence).

1129 **Figure 3 – Figure Supplement 1: Enrichment of transcription family members and gene**
1130 **regulatory network construction**. (A) -log10(P) enrichment of transcription factors annotated by
1131 family for each expression cluster. (B) Top five pathways enriched for a subset of TF-regulated
1132 gene expression.

1133 **Figure 4: Variant-to-gene mapping of autoimmune-associated loci implicates genetic**
1134 **variants in the control of T cell activation**. (**A**) Identification of genes in contact with open
1135 chromatin regions harboring autoimmune-associated SNPs. Sentinel SNPs from 15 autoimmune
1136 GWAS (**Table S11**) were used to identify the 95% credible set of proxy SNPs for each lead SNP.
1137 SNP locations were integrated with ATAC-seq and HiC data to identify the 95% credible set of
1138 accessible SNPs in physical contact with a gene promoter. Genes are further refined based on
1139 expression dynamics over the time course of T cell activation. (**B**) Partitioned LD score regression
1140 for autoimmune GWAS studies using the OCR in contact with the genes in the five clusters
1141 defined by RNA-seq. Circle size = enrichment, color = significance, **FDR<0.01, ***FDR<0.001.
1142 (**C**) Log distribution of the 1D distance between each proxy SNP and its interacting gene based
1143 on 3D chromatin conformation (median=103 kbp). (**D**) Distribution of the number of genes
1144 contacted by each accessible variant (median=5). (**E**) Fraction of open disease-associated
1145 variants that interact with no gene promoters (grey), only with the nearest gene promoter (purple),
1146 with the nearest gene and a distant gene(s) (orange), and with only a distant gene(s) (blue). (**F**)
1147 Number of genes identified by variant-to-gene mapping at each time point. Black = shared in all
1148 stages, blue = shared in two stages, red = specific to one stage. (**G**) Set of implicated proxy-genes
1149 pairs that are both differentially expressed, display differential accessibility, and chromatin contact
1150 across T cell activation. (**H**) Example MS variant rs1077667 (PP=1.0) exhibits increased
1151 accessibility and contact with the promoters of the *GPR108* and *TRIP10* at 8 hours post activation,
1152 which are increased in expression at this time point. (**I**) Example allergy variant rs7380290
1153 interacts with the *SIK1* gene that is upregulated after activation.
1154 **Figure 5**: **Comparative predictive power of orthogonal V2G approaches**. (**A**) Percent overlap
1155 in predicted genes among chromatin-based and eQTL-based V2G approaches. (**B**) Physical
1156 variant-gene associations (HiC) are enriched for statistical variant-gene associations (eQTL) in
1157 activated CD4+ T cells. The histogram depicts the null distribution of shared variant-gene pairs
1158 expected at random (~5) while the red line indicates the observed number of variant-gene pairs

1159 (41) shared with the 127 eQTL identified by Soskic et al. in a similar CD4+ T cell activation system.
1160 (**C**) Expression of the shared genes. (**D**) Precision-Recall analysis of V2G gene predictions
1161 against the set of monogenic human inborn errors in immunity.

1162 **Figure 6: Functional validation of autoimmune V2G-implicated cRE at the *IL2* locus.** (**A**) The
1163 combination of evolutionary conservation (blue), open chromatin (red), and autoimmune disease-
1164 associated SNPs at the *IL2* locus identify putative cRE in quiescent vs. 8-hour activated naïve
1165 CD4+ T cells. (**B**) Activation-dependent TAD/sub-TAD structure (heatmaps), chromatin
1166 remodeling (grey) and promoter-OCR interactions (red) at the *IL2* locus. (**C**) Recombinant reporter
1167 assay showing transcriptional activity of the *IL2* URR (+35 to -500 from the TSS) in activated
1168 primary CD4+ T cells (N= 7 donors) compared to a basal promoter (pGL4, left panel). The right
1169 panel depicts transcriptional activity of the CNS regions indicated in **A** cloned upstream of the
1170 URR. All regions except the -80 CNS show statistically significant activity relative to the URR
1171 alone (N=7, P<0.05, line = median, box = 95/5% range). (**D**) Scheme of CRISPR/CAS9-based
1172 deletion of individual *IL2* CNS using flanking gRNAs. (**E**) Activation-induced secretion of IL-2
1173 protein by CRISPR-targeted primary CD4+ T cells (N=5 donors, left panel) or Jurkat cells (N=3
1174 replicates, right panel) relative to untargeted control (CAS9, no gRNA) cells (****P<0.0001,
1175 ***P<0.001, **P<0.01, *P<0.05, line = median, box = 95/5% range). In primary cells, *B2M* gRNAs
1176 served as an irrelevant targeted control. (**F**) Scheme of CRISPR/CAS9-based deletion of the 81.3
1177 kb region containing all distal *IL2* cRE using flanking gRNAs in Jurkat cells. (**G**) Activation-induced
1178 IL-2 protein (left panel) and mRNA (right panel) by control (black) vs. 81.3 kb deleted (red) Jurkat
1179 cells (N=3 separate clones).

1180 **Figure 6 – Figure Supplement 1: TF occupancy and stability of distal IL2 cRE in CD4+ T**
1181 **cell subsets**. Evolutionary conservation (blue), open chromatin (red), and autoimmune disease-
1182 associated SNPs at the IL2 locus in quiescent (A) and 8-hour activated (B) naïve CD4+ T cells.
1183 (C) Curated TF occupancy at the IL2 URR and distal cRE from the ReMap atlas of regulatory
1184 regions.

1185 **Figure 7: Distal *IL2* enhancers are evolutionarily conserved and impact *in vivo* T cell-**
1186 **mediated immunity in mice.** (**A**) Map of the human and mouse *IL2* locus depicting mammalian
1187 conservation (dark blue), ATAC-seq from activated human and mouse CD4+ T cells (red), and
1188 orthologous non-coding sequences conserved between mouse and human (light blue). (**B**) IL-2
1189 protein secretion (top left panel) and *Il2* mRNA expression (top right panel) by CD4+ T cells from
1190 wild-type or *Il2*-83 cRE ko mice in response to stimulation with plate-bound anti-CD3 and anti-
1191 CD28 Ab *in vitro*. Bottom panels show soluble anti-CD3-induced *in vitro* clonal expansion of wild-
1192 type or *Il2*-83 cRE ko CD4+ (left) and CD8+ (right) T cells measured by dye dilution[120]. (**C**) Foxp3
1193 expression by murine CD4+CD25- Tconv stimulated with anti-CD3 Ab and TGF-beta in the
1194 absence (top panels) vs. presence (bottom panels) of exogenous IL-2. Wild-type or *Il2*-83 cRE ko
1195 mice (N=6) were immunized intraperitoneally with chicken ovalbumin in incomplete Freund's
1196 adjuvant. The frequency of CD4+PD1+CXCR5^{hi} follicular helper T cells in the spleens of 3 animals
1197 was measured by flow cytometry at day 5 post-immunization (**D**), and levels of ovalbumin-specific
1198 IgG in the serum of 3 animals were measured at day 10 post-immunization (**E**). (**F**) In vivo IBD
1199 model. CD4+CD25- Tconv from wild-type (pink, red) or *Il2*-83 cRE ko (purple, blue) mice were
1200 transferred alone (purple, pink) or together with wild-type CD4+CD25+ Treg (red, blue) into
1201 RAG1-ko mice (N=5 per group). Animal weight was monitored for 40 days, and the number (in
1202 millions) of activated CD4+ Tconv in the spleens and mesenteric lymph nodes was measured by
1203 flow cytometry (inset). Statistical significance (*$p<0.05$) was measured by ANOVA or t-test.
1204 **Figure 8: Functional validation of autoimmune V2G-implicated genes.** (**A**) Recombinant
1205 reporter assay in primary activated CD4+ T cells (N= 7 donors) showing transcriptional activity of
1206 the reference vs. risk alleles of the *IL2* -128 cRE relative to the URR alone ($P<0.0001$). (**B**)
1207 Prominent TF motifs predicted to be disrupted (blue) or stabilized (red) by promoter-connected
1208 autoimmune SNPs. (**C**) 3D chromatin-based V2G genes are enriched for CRISPR-implicated
1209 genes that regulate CD4+ T cell activation. Observed enrichment of genes regulating multiple
1210 aspects of CD4+ T cell activation (IL-2, IL-2 receptor, CTLA-4, IFNg, TNFa or proliferation) from

1211 CRISPR screens by Freimer, Schmidt, and Shifrut among sets of 3D chromatin-implicated genes
1212 among individual diseases (green, FDR<0.05) or all diseases (purple, FDR<0.05). (**D**) Enrichment
1213 for 3D chromatin-based autoimmune V2G genes among genes with germline knock-out mouse
1214 immune (red) and other (black) phenotypes (adjP<0.05, IMPC database). (**E**) Autoimmune V2G-
1215 implicated genes with at least one pharmacologic modulator (rDGIdb). (**F**) Comparison of the
1216 number of manuscripts retrieved from PubMed related to autoimmune disease for each V2G gene
1217 with pharmaceutical agents available (x-axis) with an immune-specific expression score
1218 computed using the sum GTEX median expression values (v8) for whole blood or spleen divided
1219 by other tissues (y-axis). Genes highlighted in red were selected for functional validation in **G**. (**G**)
1220 Dose-dependent impact of the indicated pharmacologic agents targeting the V2G-implicated
1221 genes *KISS1R, CHRNB, OXER1, GPR18, GRK6, PTK6, MAP3K11, GPR183, GART,* and *SIK1*
1222 on proliferation of anti-CD3/28 activated murine (left panel) or human (right panel) CD4+ T cells
1223 *in vitro* (N=4).
1224 **Figure 8 – Figure Supplement 1**: Predicted impact of autoimmune disease-associated genetic
1225 variation at V2G-implicated loci. (A) Top 50 TF motifs impacted by autoimmune SNPs (see Table
1226 S15). HiC contact frequency matrix (heatmap), chromatin accessibility (grey), and loop calls (blue-
1227 IL19/ITGA4, red-FAIM3) at the T1D-rs3024505 locus (B) and the celiac- rs13010713 locus (C).
1228 Dashed box indicates the SNP affected by allelic variation. Genetic variation at rs3024505 is
1229 predicted to disrupt a MZF1 motif (D), and variation at rs13010713 is predicted to disrupt a SOX4
1230 binding motif (E). (F) Variation at MS SNP rs1077667 is predicted to disrupt TP53, TP63, and
1231 POU2F2 (OCT2) binding sites.
1232 **Figure 8 – Figure Supplement 2: Orthogonal validation of 3D chromatin-based V2G genes**.
1233 Expression (scaled TPM) of genes implicated by chromatin contacts in (A) prior CRISPRi/a
1234 screens and (C) genes with known immune phenotypes in the international mouse phenotyping
1235 consortium. Red = increased, blue = decreased.
1236

1237 **Supplemental Files**

1238 Supplementary File 1: Expression and clustering of differentially expressed genes.

1239 Supplementary File 2: Pathway enrichment of differentially expressed RNA-seq genes

1240 Supplementary File 3: Pathway enrichment of differentially expressed RNA-seq genes by cluster

1241 Supplementary File 4: Accessible chromatin regions

1242 Supplementary File 5: Differential accessible open chromatin regions

1243 Supplementary File 6: A/B compartment calls

1244 Supplementary File 7: Differential TAD boundaries

1245 Supplementary File 8: Stripe calls

1246 Supplementary File 9: Differential contact frequency

1247 Supplementary File 10: Summary of TF footprinting

1248 Supplementary File 11: TF target gene pathway enrichment

1249 Supplementary File 12: List of all GWAS studies included

1250 Supplementary File 13: Variant to gene mapping across all timepoints

1251 Supplementary File 14: Partitioned LD score regression output

1252 Supplementary File 15: Autoimmune variants in open gene promoters

1253 Supplementary File 16: Comparison of V2G approaches

1254 Supplementary File 17: Motifs predicted to disrupt transcription factor binding sites.

1255 Supplementary File 18: V2G genes implicated by orthogonal data

1256 Supplementary File 19: V2G target gene drug repurposing results

1257

1258

1259

1260 **REFERENCES**

1261

1262 1. Zhang, F. & Lupski, J. R. Non-coding genetic variants in human disease. *Hum Mol Genet* **24**,
1263 R102-10 (2015).
1264 2. Weinreich, M. A. *et al.* KLF2 Transcription-Factor Deficiency in T Cells Results in Unrestrained
1265 Cytokine Production and Upregulation of Bystander Chemokine Receptors. *Immunity* **31**,
1266 122–130 (2009).
1267 3. Tzachanis, D. *et al.* Tob is a negative regulator of activation that is expressed in anergic and
1268 quiescent T cells. *Nat Immunol* **2**, 1174–82 (2001).
1269 4. Yusuf, I. & Fruman, D. A. Regulation of quiescence in lymphocytes. *Trends Immunol* **24**, 380–
1270 386 (2003).
1271 5. Zhu, J., Yamane, H. & Paul, W. E. Differentiation of Effector CD4 T Cell Populations. *Annu Rev*
1272 *Immunol* **28**, 445–489 (2010).
1273 6. Saravia, J., Chapman, N. M. & Chi, H. Helper T cell differentiation. *Cell Mol Immunol* **16**, 634–
1274 643 (2019).
1275 7. Farh, K. K.-H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease
1276 variants. *Nature* **518**, 337 343 (2014).
1277 8. Chen, L. *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune
1278 Cells. *Cell* **167**, 1398 1414.e24 (2016).
1279 9. Soskic, B. *et al.* Chromatin activity at GWAS loci identifies T cell states driving complex immune
1280 diseases. *Nat Genet* 1–8 (2019) doi:10.1038/s41588-019-0493-9.
1281 10. Robertson, C. C. *et al.* Fine-mapping, trans-ancestral and genomic analyses identify causal
1282 variants, cells, genes and drug targets for type 1 diabetes. *Nat Genet* **53**, 962–971 (2021).
1283 11. Lu, X. *et al.* Global discovery of lupus genetic risk variant allelic enhancer activity. *Nat*
1284 *Commun* **12**, 1611 (2021).
1285 12. Chandra, V. *et al.* Promoter-interacting expression quantitative trait loci are enriched for
1286 functional genetic variants. *Nat Genet* **53**, 110–119 (2021).
1287 13. Soskic, B. *et al.* Immune disease risk variants regulate gene expression dynamics during
1288 CD4+ T cell activation. *Nat Genet* 1–10 (2022) doi:10.1038/s41588-022-01066-3.
1289 14. Mouri, K. *et al.* Prioritization of autoimmune disease-associated genetic variants that perturb
1290 regulatory element activity in T cells. *Nat Genet* 1–10 (2022) doi:10.1038/s41588-022-
1291 01056-5.
1292 15. Rioux, J. D. & Abbas, A. K. Paths to understanding the genetic basis of autoimmune disease.
1293 *Nature* **435**, 584–589 (2005).
1294 16. Buckner, J. H. Mechanisms of impaired regulation by CD4(+)CD25(+)FOXP3(+) regulatory T
1295 cells in human autoimmune diseases. *Nat Rev Immunol* **10**, 849 859 (2010).
1296 17. Javierre, B. M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-coding
1297 Disease Variants to Target Gene Promoters. *Cell* **167**, 1369 1384.e19 (2016).
1298 18. Burren, O. S. *et al.* Chromosome contacts in activated T cells identify autoimmune disease
1299 candidate genes. *Genome Biol* **18**, 165 (2017).
1300 19. Yang, J. *et al.* Analysis of chromatin organization and gene expression in T cells identifies
1301 functional genes for rheumatoid arthritis. *Nat Commun* **11**, 4402 (2020).
1302 20. Bykova, M., Hou, Y., Eng, C. & Cheng, F. Quantitative trait locus (xQTL) approaches identify
1303 risk genes and drug targets from human non-coding genomes. *Hum Mol Genet* **31**, R105–
1304 R113 (2022).
1305 21. Shifrut, E. *et al.* Genome-wide CRISPR Screens in Primary Human T Cells Reveal Key
1306 Regulators of Immune Function. *Cell* **175**, 1958-1971.e15 (2018).
1307 22. Schmidt, R. *et al.* CRISPR activation and interference screens decode stimulation responses
1308 in primary human T cells. *Science* **375**, eabj4008 (2022).

1309 23. Freimer, J. W. *et al.* Systematic discovery and perturbation of regulatory genes in human T
1310 cells reveals the architecture of immune networks. *Nat Genet* **54**, 1133–1144 (2022).
1311 24. Su, C., Pahl, M. C., Grant, S. F. A. & Wells, A. D. Restriction enzyme selection dictates
1312 detection range sensitivity in chromatin conformation capture-based variant-to-gene
1313 mapping approaches. *Hum Genet* (2021) doi:10.1007/s00439-021-02326-8.
1314 25. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**,
1315 75 (2012).
1316 26. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals
1317 Folding Principles of the Human Genome. *Science* **326**, 289–293 (2009).
1318 27. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of
1319 chromatin interactions. *Nature* **485**, 376 380 (2012).
1320 28. Mirny, L. A., Imakaev, M. & Abdennur, N. Two major mechanisms of chromosome
1321 organization. *Curr Opin Cell Biol* **58**, 142–152 (2019).
1322 29. Yoon, S., Chandra, A. & Vahedi, G. Stripenn detects architectural stripes from chromatin
1323 conformation data using computer vision. *Nat Commun* **13**, 1602 (2022).
1324 30. Kim, T.-K. & Shiekhattar, R. Architectural and Functional Commonalities between Enhancers
1325 and Promoters. *Cell* **162**, 948 959 (2015).
1326 31. Hamamoto, K. & Fukaya, T. Molecular architecture of enhancer–promoter interaction. *Curr*
1327 *Opin Cell Biol* **74**, 62–70 (2022).
1328 32. Consortium, R. E. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**,
1329 317 330 (2015).
1330 33. Yoon, J.-H. *et al.* Phosphorylation status determines the opposing functions of Smad2/Smad3
1331 as STAT3 cofactors in TH17 differentiation. *Nat Commun* **6**, 7600 (2015).
1332 34. Safford, M. *et al.* Egr-2 and Egr-3 are negative regulators of T cell activation. *Nat Immunol* **6**,
1333 472–480 (2005).
1334 35. Schubert, L. A., Jeffery, E., Zhang, Y., Ramsdell, F. & Ziegler, S. F. Scurfin (FOXP3) Acts as
1335 a Repressor of Transcription and Regulates T Cell Activation. *J Biol Chem* **276**, 37672–
1336 37679 (2001).
1337 36. Xiao, T., Li, X. & Felsenfeld, G. The Myc-associated zinc finger protein (MAZ) works together
1338 with CTCF to control cohesin positioning and genome organization. *Proc National Acad*
1339 *Sci* **118**, e2023127118 (2021).
1340 37. Sebzda, E., Zou, Z., Lee, J. S., Wang, T. & Kahn, M. L. Transcription factor KLF2 regulates
1341 the migration of naive T cells by restricting chemokine receptor expression patterns. *Nat*
1342 *Immunol* **9**, 292–300 (2008).
1343 38. Moskowitz, D. M. *et al.* Epigenomics of human CD8 T cell differentiation and aging. *Sci*
1344 *Immunol* **2**, eaag0192 (2017).
1345 39. Harris, J. E. *et al.* Early Growth Response Gene-2, a Zinc-Finger Transcription Factor, Is
1346 Required for Full Induction of Clonal Anergy in CD4+ T Cells. *J Immunol* **173**, 7331–7338
1347 (2004).
1348 40. Williams, J. B. *et al.* The EGR2 targets LAG-3 and 4-1BB describe and regulate dysfunctional
1349 antigen-specific CD8+ T cells in the tumor microenvironment. *J Exp Med* **214**, 381–400
1350 (2017).
1351 41. Avitahl, N. *et al.* Ikaros Sets Thresholds for T Cell Activation and Regulates Chromosome
1352 Propagation. *Immunity* **10**, 333–343 (1999).
1353 42. Bandyopadhyay, S. *et al.* Interleukin 2 gene transcription is regulated by Ikaros-induced
1354 changes in histone acetylation in anergic T cells. *Blood* **109**, 2878–86 (2007).
1355 43. Thomas, R. M. *et al.* Ikaros Enforces the Costimulatory Requirement for IL2 Gene Expression
1356 and Is Required for Anergy Induction in CD4+ T Lymphocytes. *J Immunol* **179**, 7305–7315
1357 (2007).
1358 44. Thomas, R. M. *et al.* Ikaros Silences T-bet Expression and Interferon-γ Production during T
1359 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.

1410 67. Fulco, C. P. *et al.* Activity-by-contact model of enhancer–promoter regulation from thousands
1411 of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
1412 68. Cerosaletti, K. *et al.* Multiple autoimmune-associated variants confer decreased IL-2R
1413 signaling in CD4+ CD25(hi) T cells of type 1 diabetic and multiple sclerosis patients. *Plos*
1414 *One* **8**, e83811 (2013).
1415 69. Spolski, R., Li, P. & Leonard, W. J. Biology and regulation of IL-2: from molecular mechanisms
1416 to human therapy. *Nat Rev Immunol* **18**, 648 659 (2018).
1417 70. Abbas, A. K., Trotta, E., Simeonov, D. R., Marson, A. & Bluestone, J. A. Revisiting IL-2:
1418 Biology and therapeutic prospects. *Sci Immunol* **3**, eaat1482 (2018).
1419 71. Joosse, M. E. *et al.* Duplication of the IL2RA locus causes excessive IL-2 signaling and may
1420 predispose to very early onset colitis. *Mucosal Immunol* 1–11 (2021) doi:10.1038/s41385-
1421 021-00423-5.
1422 72. Fujita, T., Shibuya, H., Ohashi, T., Yamanishi, K. & Taniguchi, T. Regulation of human
1423 interleukin-2 gene: Functional DNA sequences in the 5′ flanking region for the gene
1424 expression in activated T lymphocytes. *Cell* **46**, 401–407 (1986).
1425 73. Durand, D. B., Bush, M. R., Morgan, J. G., Weiss, A. & Crabtree, G. R. A 275 basepair
1426 fragment at the 5' end of the interleukin 2 gene enhances expression from a heterologous
1427 promoter in response to signals from the T cell antigen receptor. *J Exp Medicine* **165**, 395
1428 407 (1987).
1429 74. Novak, T. J., White, P. M. & Rothenberg, E. V. Regulatory anatomy of the murine interleukin-
1430 2 gene. *Nucleic Acids Res* **18**, 4523–4533 (1990).
1431 75. Mehra, P. & Wells, A. D. Long-Range Transcriptional Control of the Il2 Gene by an Intergenic
1432 Enhancer. *Mol Cell Biol* **35**, 3880–3891 (2015).
1433 76. Consortium, T. E. P. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**,
1434 636–640 (2004).
1435 77. Lin, X. *et al.* Nested epistasis enhancer networks for robust genome regulation. *Science* **377**,
1436 1077–1085 (2022).
1437 78. Hedlund, G., Dohlsten, M., Ericsson, P. O. & Sjögren, H. O. Rapid response to Con A by
1438 CD4+CD45R− rat memory lymphocytes as compared to CD4+CD45R+ lymphocytes. *Cell*
1439 *Immunol* **119**, 317–326 (1989).
1440 79. Picker, L. *et al.* Direct demonstration of cytokine synthesis heterogeneity among human
1441 memory/effector T cells by flow cytometry. *Blood* **86**, 1408–1419 (1995).
1442 80. Sojka, D. K., Bruniquel, D., Schwartz, R. H. & Singh, N. J. IL-2 Secretion by CD4+ T Cells In
1443 Vivo Is Rapid, Transient, and Influenced by TCR-Specific Competition. *J Immunol* **172**,
1444 6136–6143 (2004).
1445 81. Powrie, F., Correa-Oliveira, R., Mauze, S. & Coffman, R. L. Regulatory interactions between
1446 CD45RBhigh and CD45RBlow CD4+ T cells are important for the balance between
1447 protective and pathogenic cell-mediated immunity. *J Exp Medicine* **179**, 589–600 (1994).
1448 82. Morawski, P. A., Mehra, P., Chen, C., Bhatti, T. & Wells, A. D. Foxp3 Protein Stability Is
1449 Regulated by Cyclin-dependent Kinase 2. *J Biol Chem* **288**, 24494–24502 (2013).
1450 83. Smigiel, K. S. *et al.* CCR7 provides localized access to IL-2 and defines homeostatically
1451 distinct regulatory T cell subsets. *J Exp Medicine* **211**, 121–136 (2014).
1452 84. Mostafavi, H., Spence, J. P., Naqvi, S. & Pritchard, J. K. Systematic differences in discovery
1453 of genetic effects on gene expression and complex traits. *Nat. Genet.* 1–10 (2023)
1454 doi:10.1038/s41588-023-01529-1.
1455 85. Kovanen, P. E. *et al.* T-cell development and function are modulated by dual specificity
1456 phosphatase DUSP5. *J Biological Chem* **283**, 17362–9 (2008).
1457 86. Chuang, H.-C. & Tan, T.-H. MAP4K Family Kinases and DUSP Family Phosphatases in T-
1458 Cell Signaling and Systemic Lupus Erythematosus. *Cells* **8**, 1433 (2019).
1459 87. Danileviciute, E. *et al.* PARK7/DJ-1 promotes pyruvate dehydrogenase activity and maintains
1460 Treg homeostasis during ageing. *Nat Metabolism* **4**, 589–607 (2022).

1461 88. McHugh, D. *et al.* N-arachidonoyl glycine, an abundant endogenous lipid, potently drives
1462 directed cellular migration through GPR18, the putative abnormal cannabidiol receptor.
1463 *Bmc Neurosci* **11**, 44 (2010).
1464 89. Kan, J. L., Jannatipour, M., Taylor, S. M. & Moran, R. G. Mouse cDNAs encoding a
1465 trifunctional protein of de novo purine synthesis and a related single-domain glycinamide
1466 ribonucleotide synthetase. *Gene* **137**, 195–202 (1993).
1467 90. Varney, M. J. *et al.* G protein–coupled receptor kinase 6 (GRK6) regulates insulin processing
1468 and secretion via effects on proinsulin conversion to insulin. *J Biol Chem* **298**, 102421
1469 (2022).
1470 91. Gilic, M. B. & Tyner, A. L. Targeting protein tyrosine kinase 6 in cancer. *Biochimica Et*
1471 *Biophysica Acta Bba - Rev Cancer* **1874**, 188432 (2020).
1472 92. Slattery, M. L., Lundgreen, A. & Wolff, R. K. MAP kinase genes and colon and rectal cancer.
1473 *Carcinogenesis* **33**, 2398–2408 (2012).
1474 93. Knackmuss, U. *et al.* MAP3K11 is a tumor suppressor targeted by the oncomiR miR-125b in
1475 early B cells. *Cell Death Differ* **23**, 242–252 (2016).
1476 94. Hammond, R. K. *et al.* Biological constraints on GWAS SNPs at suggestive significance
1477 thresholds reveal additional BMI loci. *Elife* **10**, e62206 (2021).
1478 95. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**,
1479 357–359 (2012).
1480 96. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
1481 97. Lun, A. T. L. & Smyth, G. K. csaw: a Bioconductor package for differential binding analysis of
1482 ChIP-seq data using sliding windows. *Nucleic Acids Res* **44**, e45–e45 (2016).
1483 98. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for
1484 differential expression analysis of digital gene expression data. *Bioinform Oxf Engl* **26**,
1485 139–40 (2009).
1486 99. Wingett, S. *et al.* HiCUP: pipeline for mapping and processing Hi-C data. *F1000research* **4**,
1487 1310 (2015).
1488 100. Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically
1489 labeled arrays. *Bioinform Oxf Engl* **36**, 311–316 (2019).
1490 101. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome
1491 organization. *Nat Methods* **9**, 999 1003 (2012).
1492 102. Nora, E. P. *et al.* Molecular basis of CTCF binding polarity in genome folding. *Nat Commun*
1493 **11**, 5612 (2020).
1494 103. Cresswell, K. G. & Dozmorov, M. G. TADCompare: An R Package for Differential and
1495 Temporal Analysis of Topologically Associated Domains. *Frontiers Genetics* **11**, 158
1496 (2020).
1497 104. Kaul, A., Bhattacharyya, S. & Ay, F. Identifying statistically significant chromatin contacts
1498 from Hi-C data with FitHiC2. *Nat Protoc* **15**, 991–1012 (2020).
1499 105. Stansfield, J. C., Cresswell, K. G., Vladimirov, V. I. & Dozmorov, M. G. HiCcompare: an R-
1500 package for joint normalization and comparison of HI-C datasets. *Bmc Bioinformatics* **19**,
1501 279 (2018).
1502 106. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput
1503 sequencing data. *Bioinform Oxf Engl* **31**, 166–9 (2014).
1504 107. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray
1505 and RNA-Seq data. *Bmc Bioinformatics* **14**, 7–7 (2013).
1506 108. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell*
1507 *Syst* **1**, 417–425 (2015).
1508 109. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and
1509 microarray studies. *Nucleic Acids Res* **43**, e47–e47 (2015).
1510 110. Li, Z. *et al.* Identification of transcription factor binding sites using ATAC-seq. *Genome Biol*
1511 **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

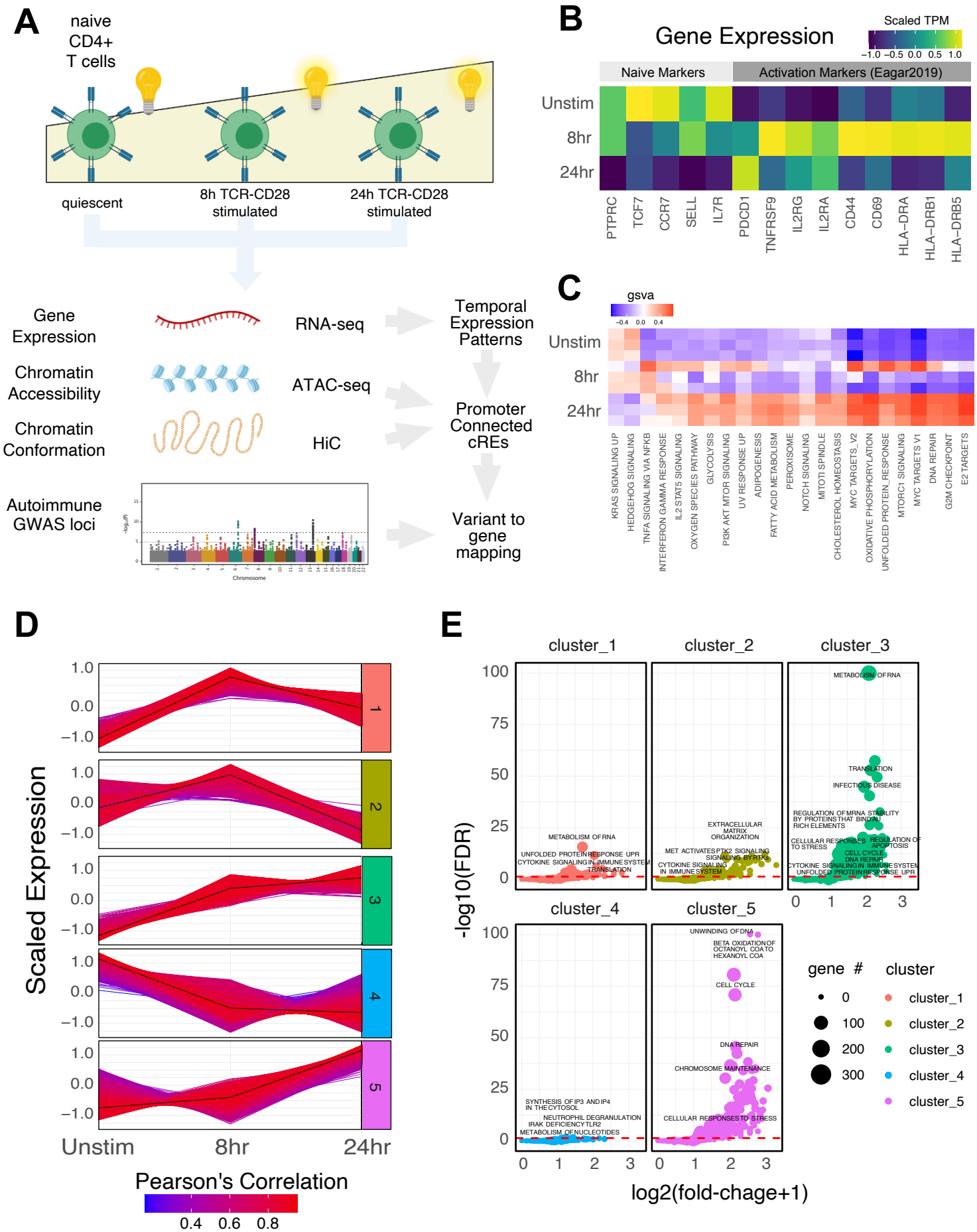
A
naive CD4+ T cells
quiescent
8h TCR-CD28 stimulated
24h TCR-CD28 stimulated
Gene Expression
RNA-seq
Chromatin Accessibility
ATAC-seq
Chromatin Conformation
HiC
Autoimmune GWAS loci
Chromosome
Temporal Expression Patterns
Promoter Connected cREs
Variant to gene mapping
B
Gene Expression
Scaled TPM
−1.0 −0.5 0.0 0.5 1.0
Naive Markers
Activation Markers (Eagar2019)
Unstim
8hr
24hr
PTPRC
TCF7
CCR7
SELL
IL7R
PDCD1
TNFRSF9
IL2RG
IL2RA
CD44
CD69
HLA−DRA
HLA−DRB1
HLA−DRB5
C
gsva
−0.4 0.0 0.4
Unstim
8hr
24hr
KRAS SIGNALING UP
HEDGEHOG SIGNALING
TNFA SIGNALING VIA NFKB
INTERFERON GAMMA RESPONSE
IL2 STAT5 SIGNALING
OXYGEN SPECIES PATHWAY
GLYCOLYSIS
PI3K AKT MTOR SIGNALING
UV RESPONSE UP
ADIPOGENESIS
FATTY ACID METABOLISM
PEROXISOME
NOTCH SIGNALING
MITOTI SPINDLE
CHOLESTEROL HOMEOSTASIS
MYC TARGETS_V2
OXIDATIVE PHOSPHORYLATION
UNFOLDED PROTEIN_RESPONSE
MTORC1 SIGNALING
MYC TARGETS V1
DNA REPAIR
G2M CHECKPOINT
E2 TARGETS
D
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
E
cluster_1
cluster_2
cluster_3
cluster_4
cluster_5
-log10(FDR)
100
75
50
25
0
METABOLISM OF RNA
UNFOLDED PROTEIN RESPONSE UPR
CYTOKINE SIGNALING IN IMMUNE SYSTEM
TRANSLATION
EXTRACELLULAR MATRIX ORGANIZATION
MET ACTIVATES PTK2 SIGNALING
CYTOKINE SIGNALING IN IMMUNE SYSTEM
TRANSLATION
INFECTIOUS DISEASE
REGULATION OF MRNA STABILITY BY PROTEINS THAT BIND AU RICH ELEMENTS
CELLULAR RESPONSES TO STRESS
REGULATION OF APOPTOSIS
CELL CYCLE
DNA REPAIR
CYTOKINE SIGNALING IN IMMUNE SYSTEM
UNFOLDED PROTEIN RESPONSE UPR
UNWINDING OF DNA
BETA OXIDATION OF OCTANOYL COA TO HEXANOYL COA
CELL CYCLE
DNA REPAIR
CHROMOSOME MAINTENANCE
CELLULAR RESPONSES TO STRESS
SYNTHESIS OF IP3 AND IP4 IN THE CYTOSOL
NEUTROPHIL DEGRANULATION
IRAK DEFICIENCY TLR2
METABOLISM OF NUCLEOTIDES
0 1 2 3
log2(fold-chage+1)
gene #
0
100
200
300
cluster
cluster_1
cluster_2
cluster_3
cluster_4
cluster_5

Figure 2

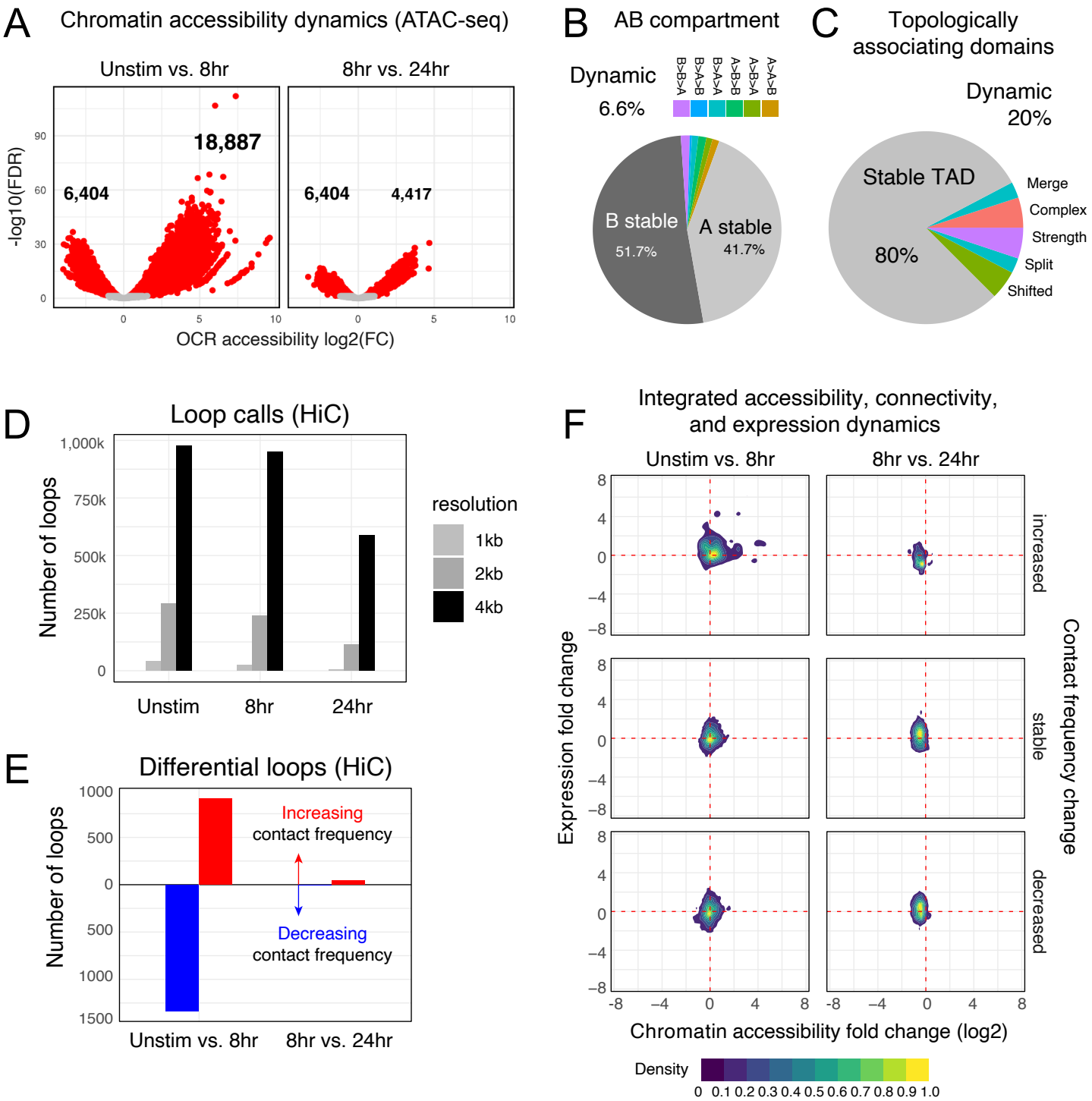
A
Chromatin accessibility dynamics (ATAC-seq)
Unstim vs. 8hr
8hr vs. 24hr
-log10(FDR)
18,887
6,404
6,404
4,417
OCR accessibility log2(FC)
B
AB compartment
Dynamic
6.6%
B>B>A
B<A>B
B>A>A
A>B>B
A<B>A
A>A>B
B stable
51.7%
A stable
41.7%
C
Topologically associating domains
Dynamic
20%
Stable TAD
80%
Merge
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
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

Figure 3

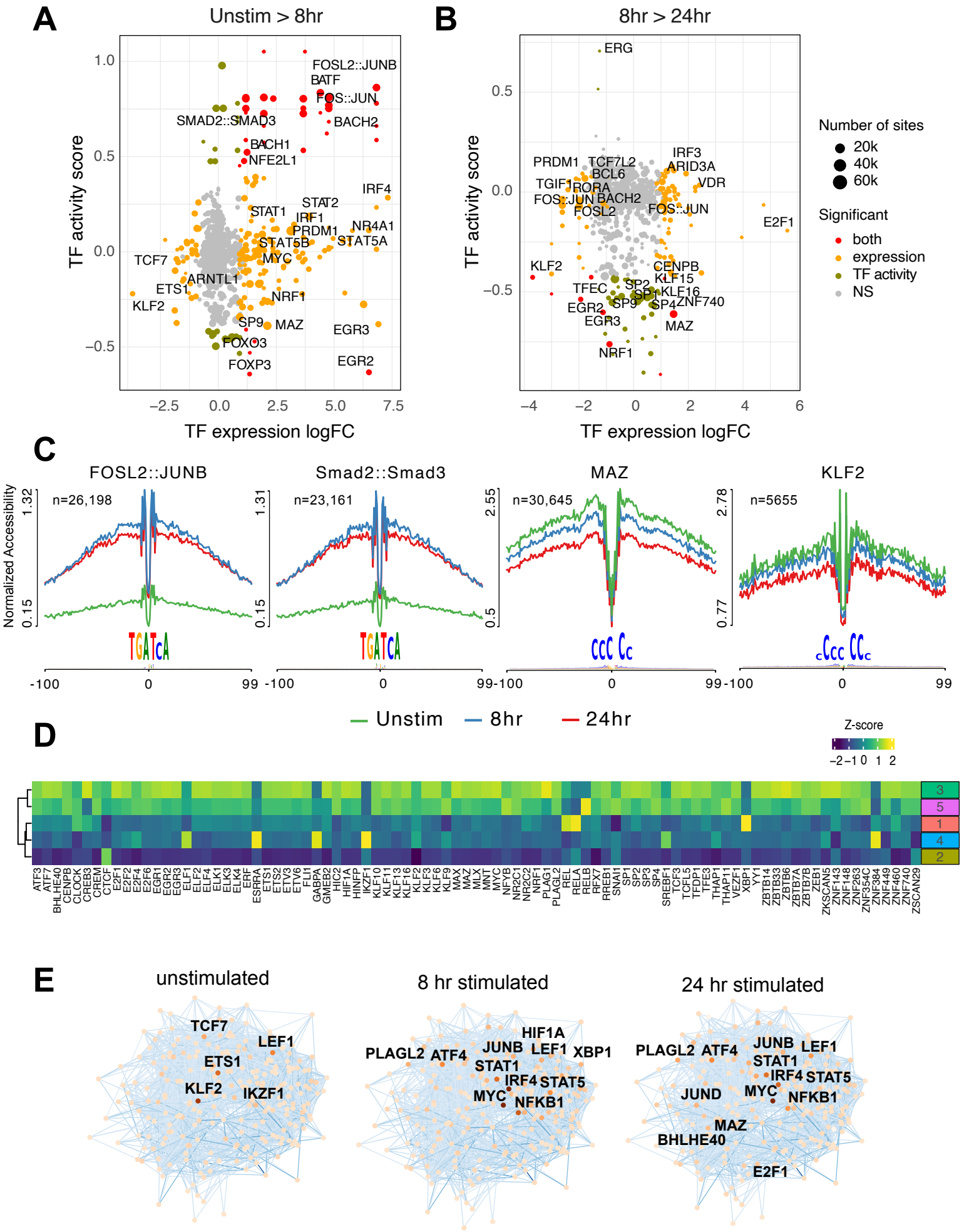
A
Unstim > 8hr
TF activity score
TF expression logFC
FOSL2::JUNB
BATF
FOS::JUN
BACH2
SMAD2::SMAD3
BACH1
NFE2L1
IRF4
STAT2
STAT1
IRF1
PRDM1
NR4A1
STAT5B
STAT5A
TCF7
MYC
ARNTL1
ETS1
NRF1
KLF2
SP9
MAZ
EGR3
FOXO3
FOXP3
EGR2
B
8hr > 24hr
ERG
PRDM1
TCF7L2
BCL6
IRF3
ARID3A
VDR
TGIF1
RORA
FOS::JUN
BACH2
FOSL2
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

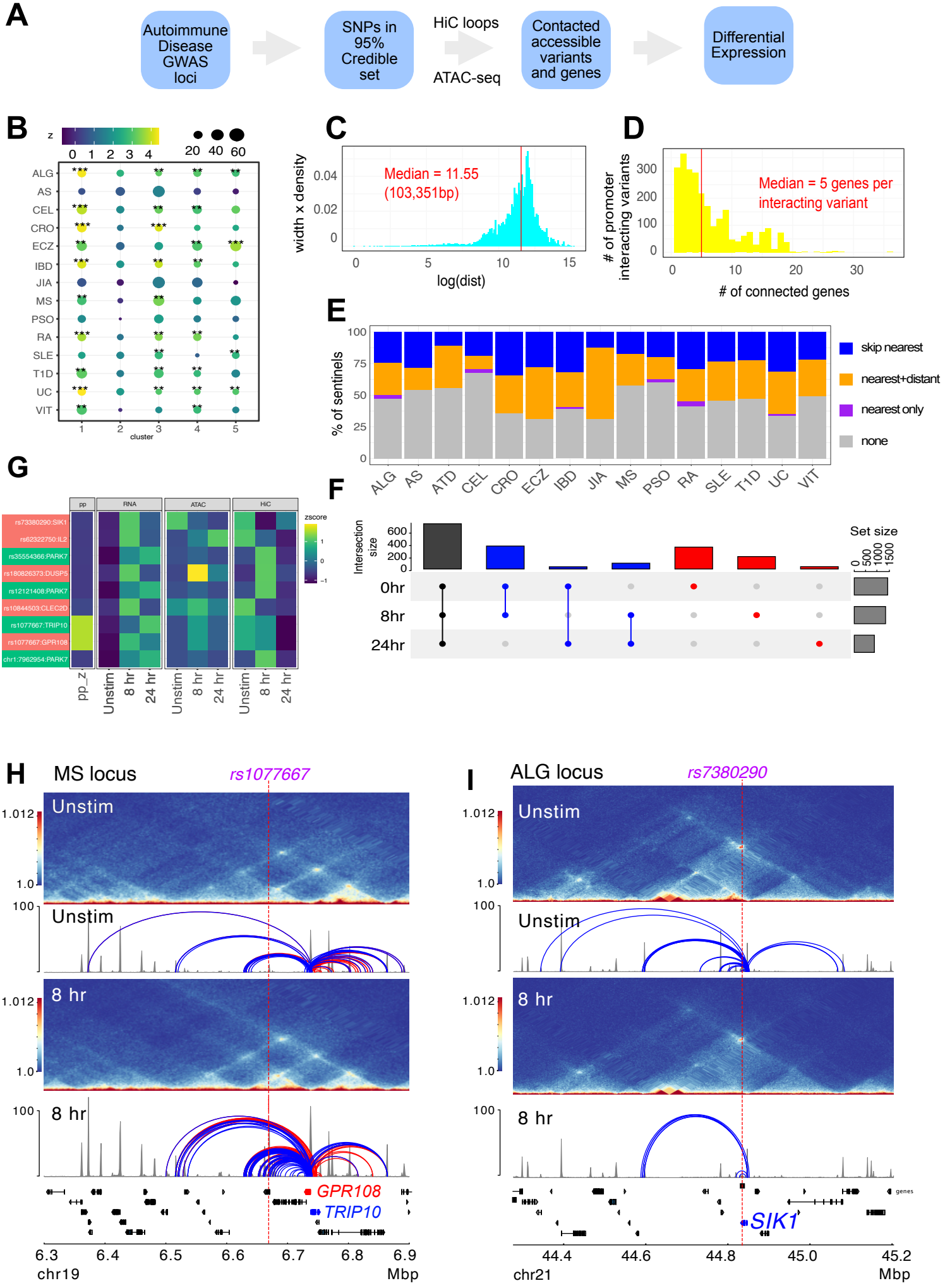
A
Autoimmune Disease GWAS loci
SNPs in 95% Credible set
HiC loops
ATAC-seq
Contacted accessible variants and genes
Differential Expression
B
z
0 1 2 3 4
20 40 60
ALG
AS
CEL
CRO
ECZ
IBD
JIA
MS
PSO
RA
SLE
T1D
UC
VIT
1 2 3 4 5
cluster
C
width x density
Median = 11.55 (103,351bp)
log(dist)
D
# of promoter 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
pp
RNA
ATAC
HiC
zscore
rs73380290:SIK1
rs62322750:IL2
rs35554366:PARK7
rs180826373:DUSP5
rs12121408:PARK7
rs10844503:CLEC2D
rs1077667:TRIP10
rs1077667:GPR108
chr1:7962954:PARK7
pp_z
Unstim
8 hr
24 hr
H
MS locus
rs1077667
Unstim
8 hr
GPR108
TRIP10
chr19
Mbp
I
ALG locus
rs7380290
Unstim
8 hr
SIK1
chr21
Mbp

Figure 5

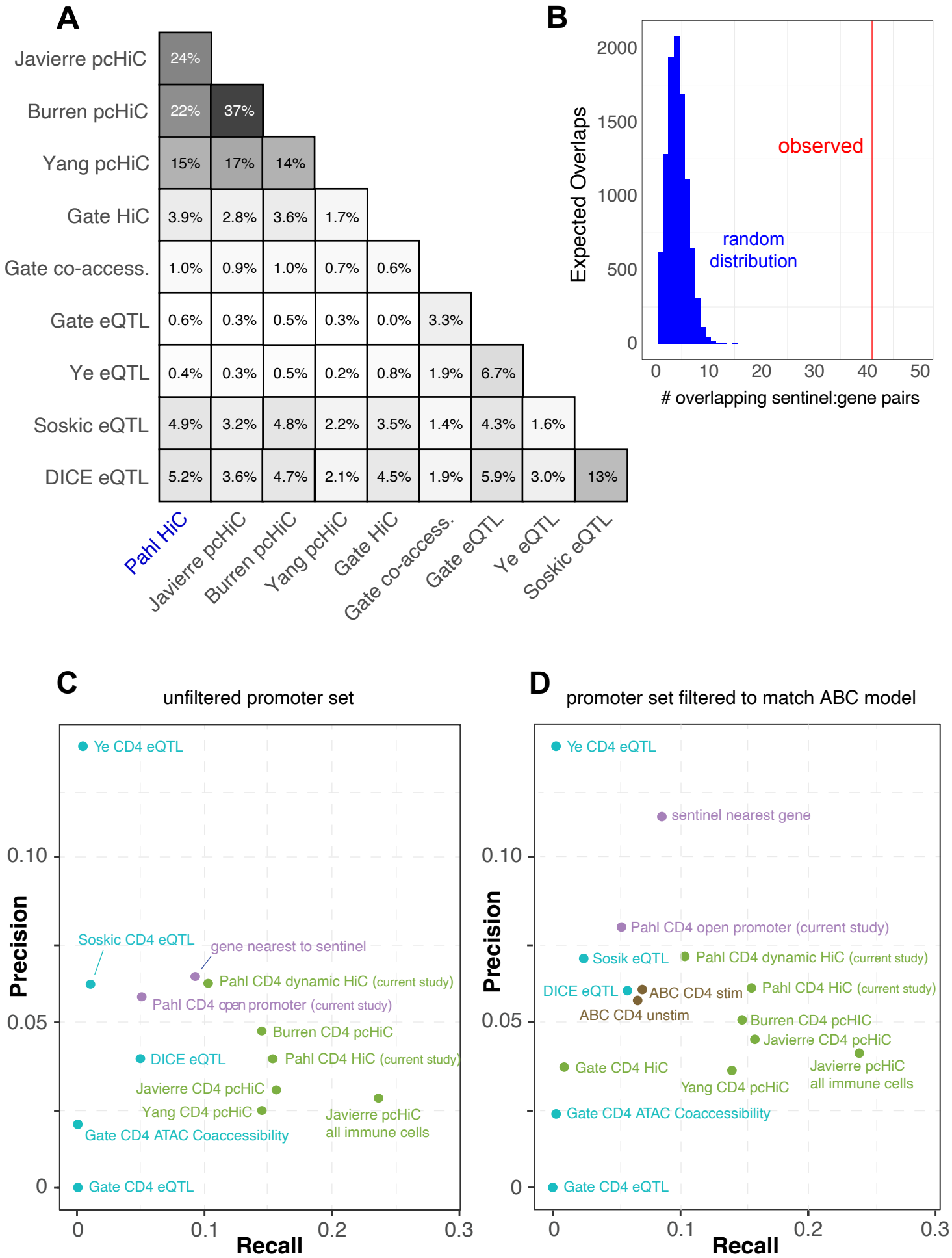
A
Javierre pcHiC
Burren pcHiC
Yang pcHiC
Gate HiC
Gate co-access.
Gate eQTL
Ye eQTL
Soskic eQTL
DICE eQTL
Pahl HiC
24%
22% 37%
15% 17% 14%
3.9% 2.8% 3.6% 1.7%
1.0% 0.9% 1.0% 0.7% 0.6%
0.6% 0.3% 0.5% 0.3% 0.0% 3.3%
0.4% 0.3% 0.5% 0.2% 0.8% 1.9% 6.7%
4.9% 3.2% 4.8% 2.2% 3.5% 1.4% 4.3% 1.6%
5.2% 3.6% 4.7% 2.1% 4.5% 1.9% 5.9% 3.0% 13%
B
Expected Overlaps
2000
1500
1000
500
0
observed
random distribution
0 10 20 30 40 50
overlapping sentinel:gene pairs
C
unfiltered promoter set
Precision
Recall
0.10
0.05
0
0 0.1 0.2 0.3
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
0.10
0.05
0
0 0.1 0.2 0.3
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

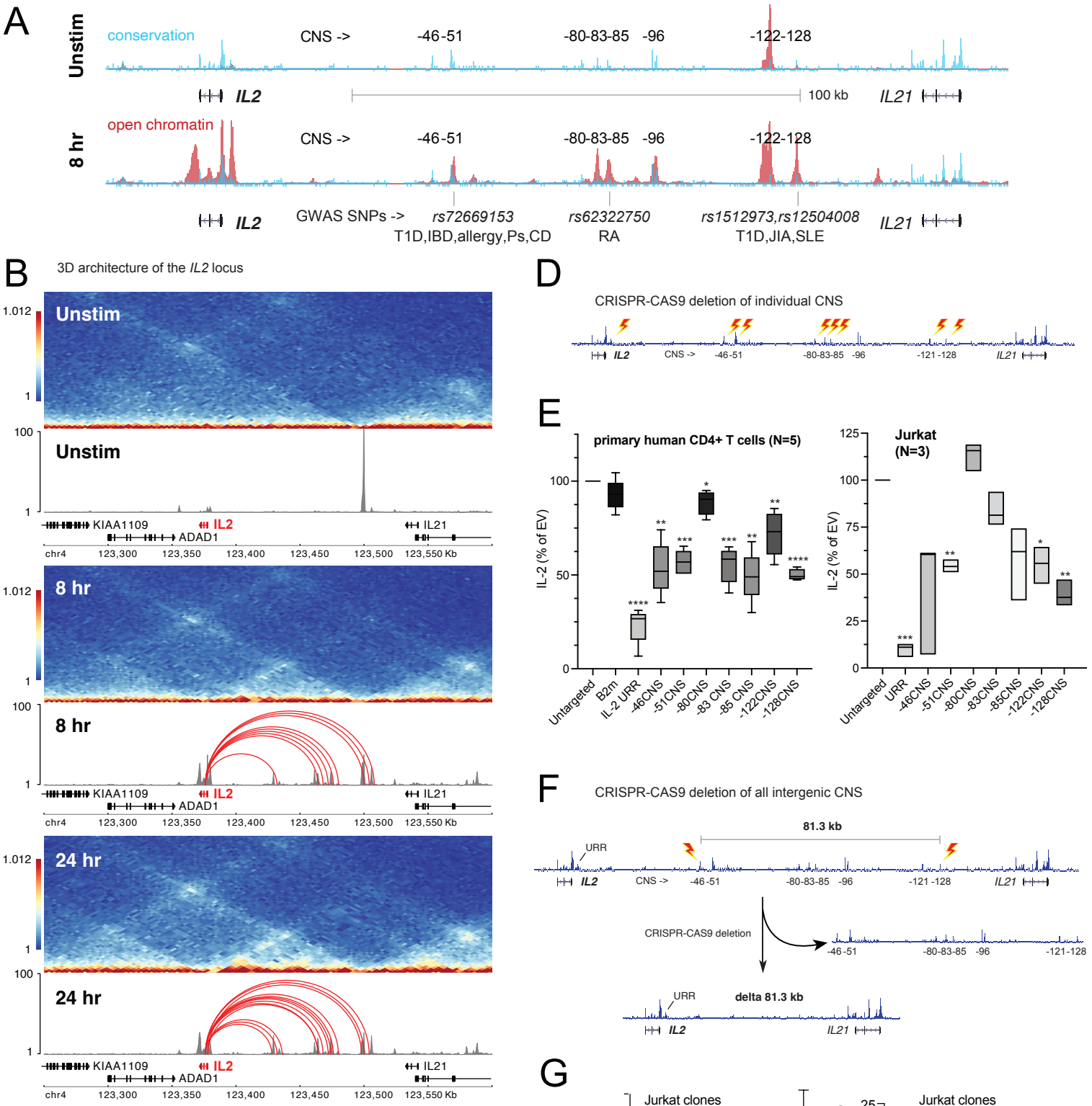
A
Unstim
conservation
CNS ->
-46-51
-80-83-85
-96
-122-128
IL2
100 kb
IL21
8 hr
open chromatin
GWAS SNPs ->
rs72669153
T1D,IBD,allergy,Ps,CD
rs62322750
RA
rs1512973,rs12504008
T1D,JIA,SLE
B
3D architecture of the IL2 locus
Unstim
8 hr
24 hr
KIAA1109
ADAD1
IL21
chr4
123,300
123,350
123,400
123,450
123,500
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

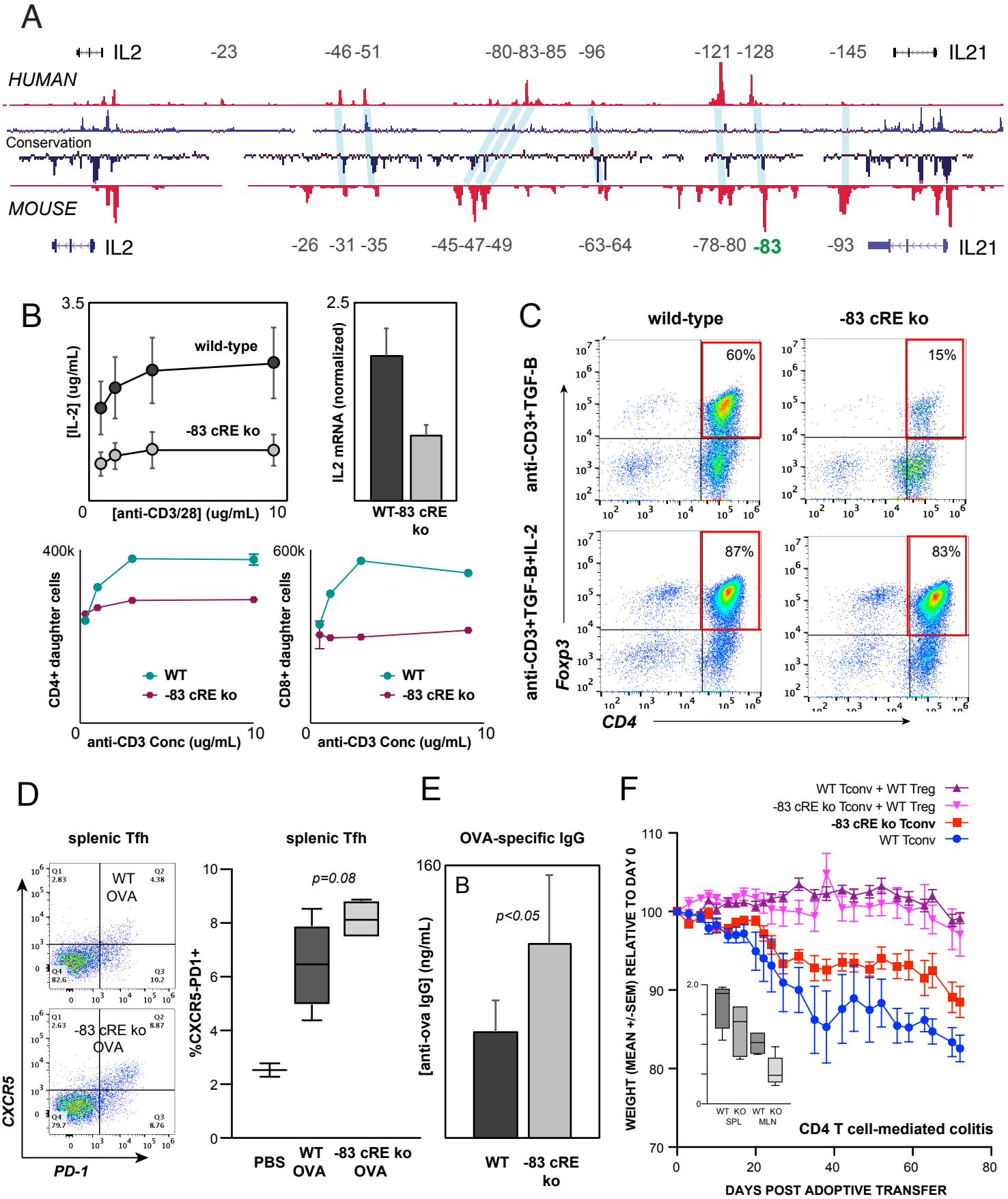
A
IL2
-23
-46 -51
-80-83-85 -96
-121 -128
-145
IL21
HUMAN
Conservation
MOUSE
IL2
-26 -31 -35
-45-47-49
-63-64
-78-80 -83
-93
IL21
B
[IL-2] (ug/mL)
wild-type
-83 cRE ko
[anti-CD3/28] (ug/mL)
IL2 mRNA (normalized)
WT-83 cRE ko
CD4+ daughter cells
CD8+ daughter cells
WT
-83 cRE ko
anti-CD3 Conc (ug/mL)
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
B
p<0.05
[anti-ova IgG] (ng/mL)
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

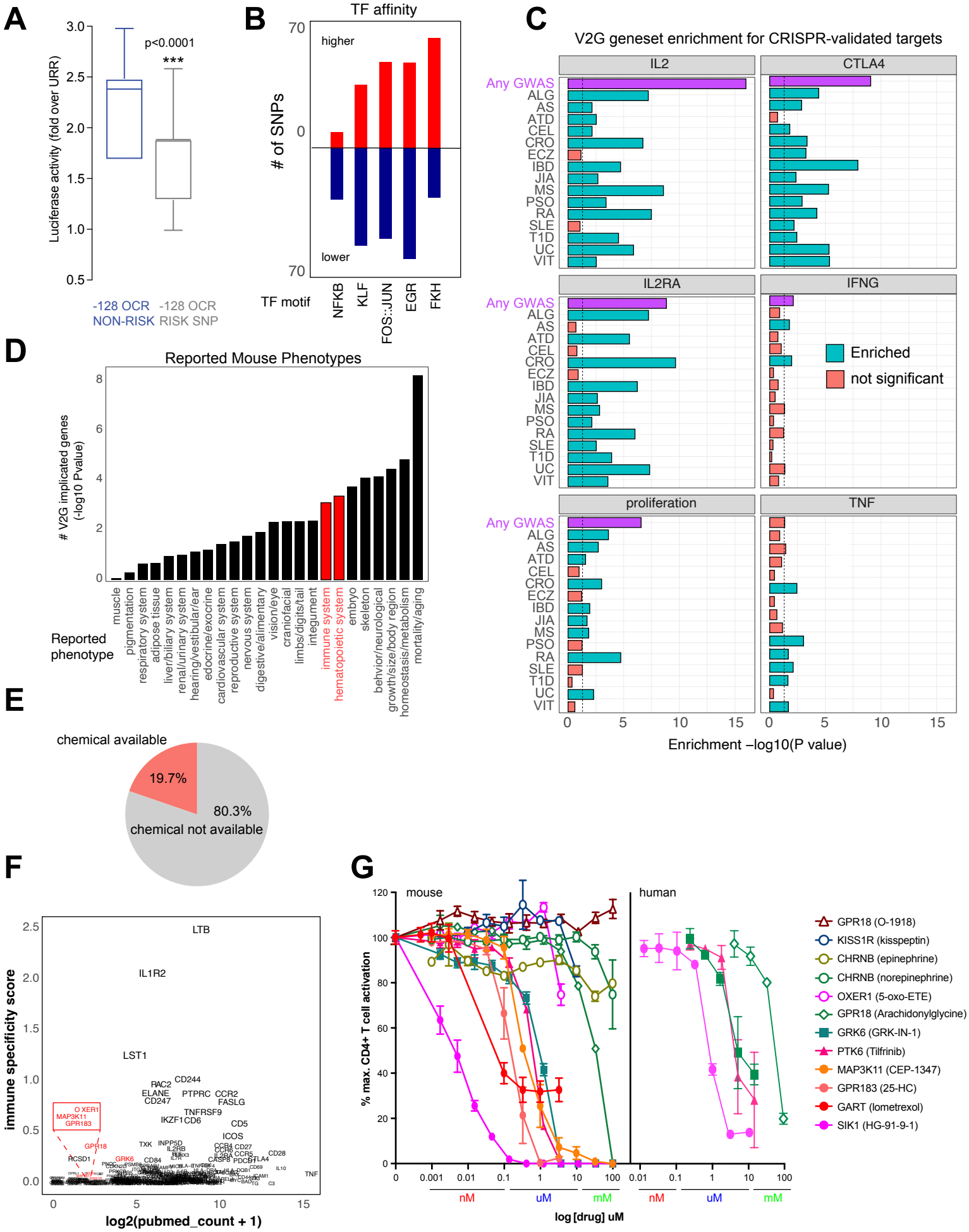
A
Luciferase activity (fold over URR)
p<0.0001
***
-128 OCR NON-RISK
-128 OCR RISK SNP
B
TF affinity
higher
lower
# of SNPs
TF motif
NFKB
KLF
FOS::JUN
EGR
FKH
C
V2G geneset enrichment for CRISPR-validated targets
IL2
CTLA4
IL2RA
IFNG
proliferation
TNF
Any GWAS
ALG
AS
ATD
CEL
CRO
ECZ
IBD
JIA
MS
PSO
RA
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
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 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.

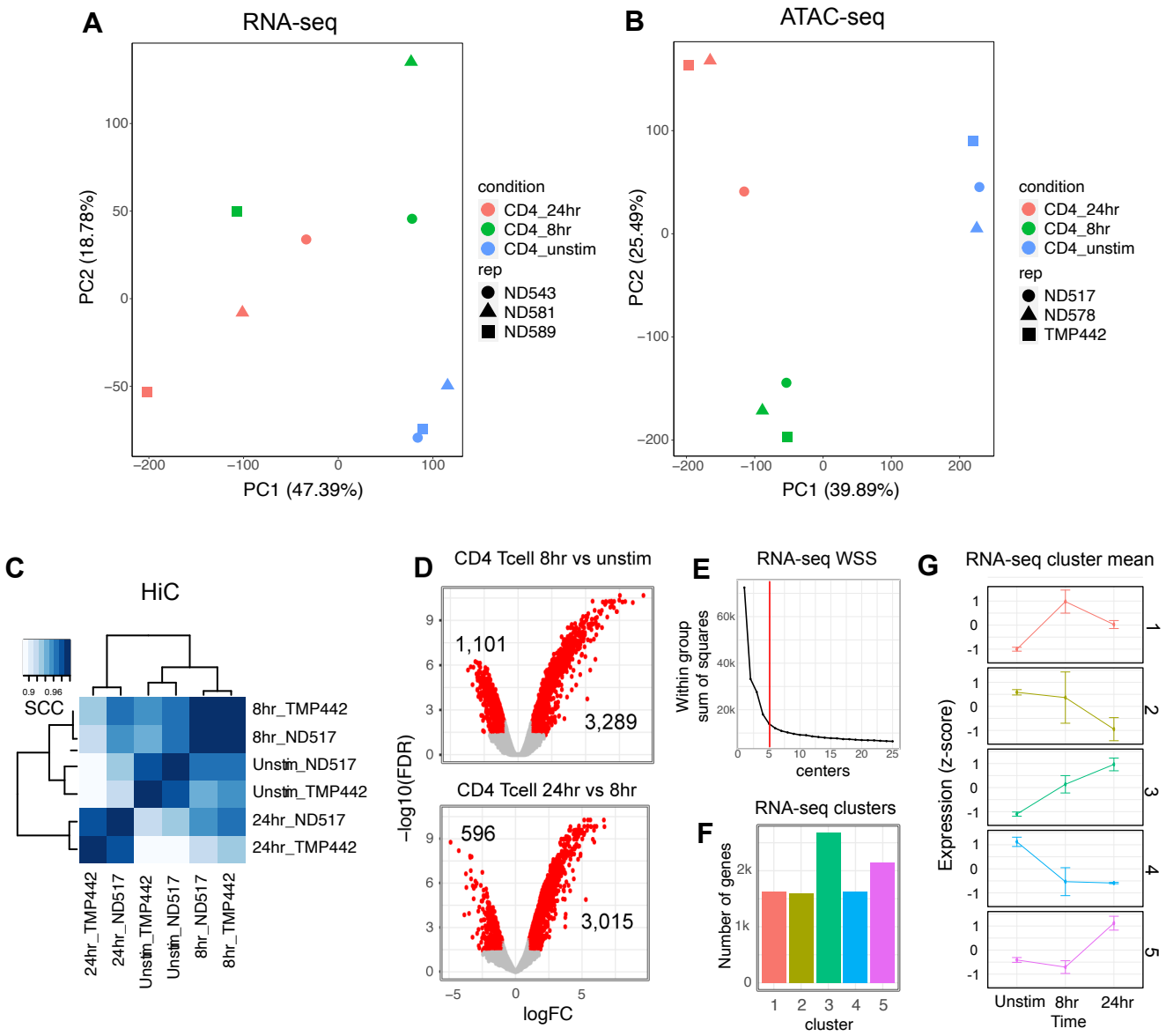

Figure S2: Characterization of CD4+ T cell epigenomic data. (A) Accessibility of OCR and (B) expression of genes located in A vs. B compartments (log2 fkpm). Differential accessibility (C) and expression (D) of cRE-gene pairs in regions exhibiting differential TAD structure. (E) Enrichment of promoter 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).

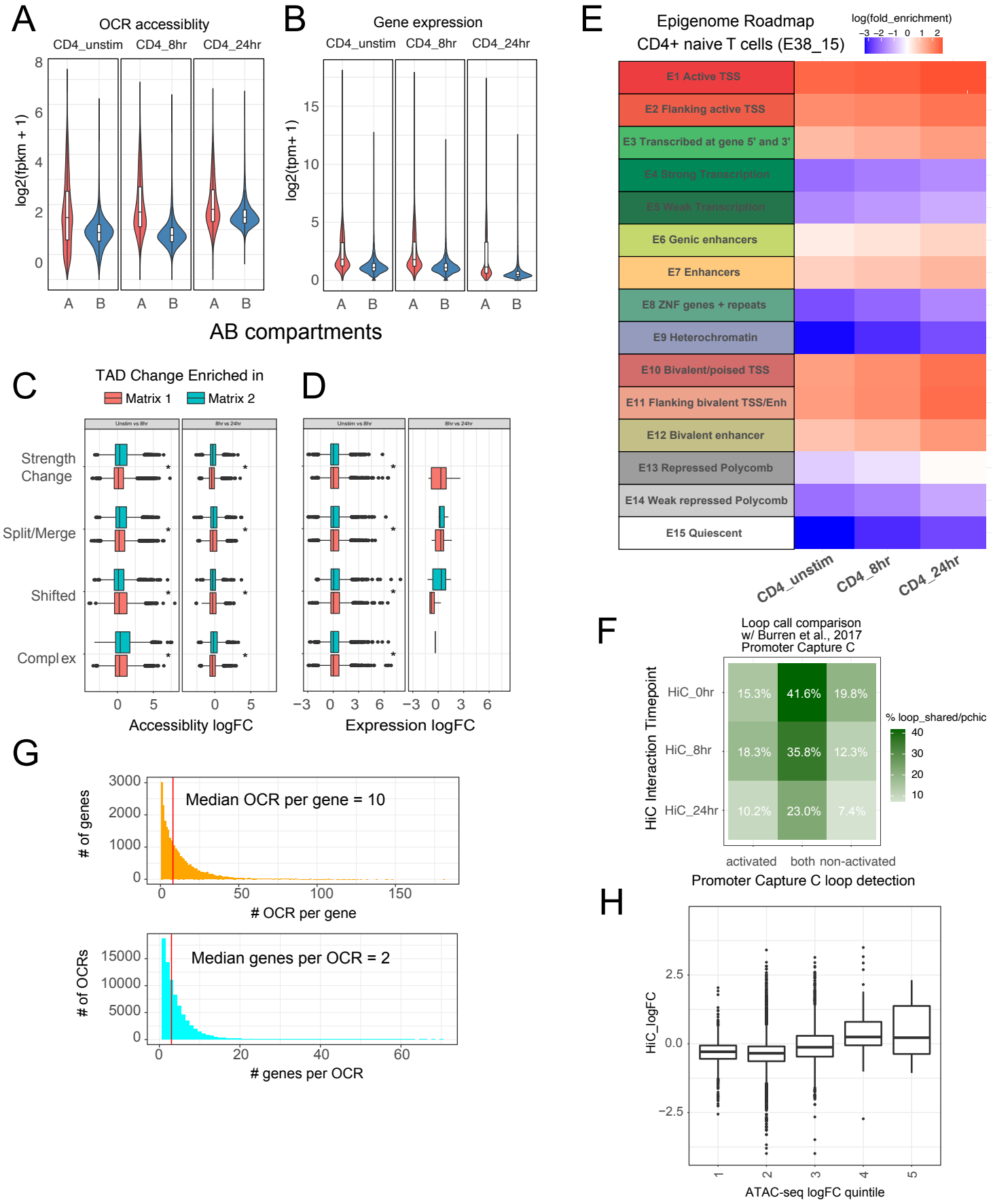

Figure S3: Characterization of chromatin stripes during CD4+ T cell activation. (A) Stripes called per timepoint. (B) Proportion of SNPs annotated in A vs B compartments. Accessibility of OCR (C) and expression of genes (D) located in anchor (small more interactive region of stripe), stripe excluding the anchor (stripe), or outside stripes.

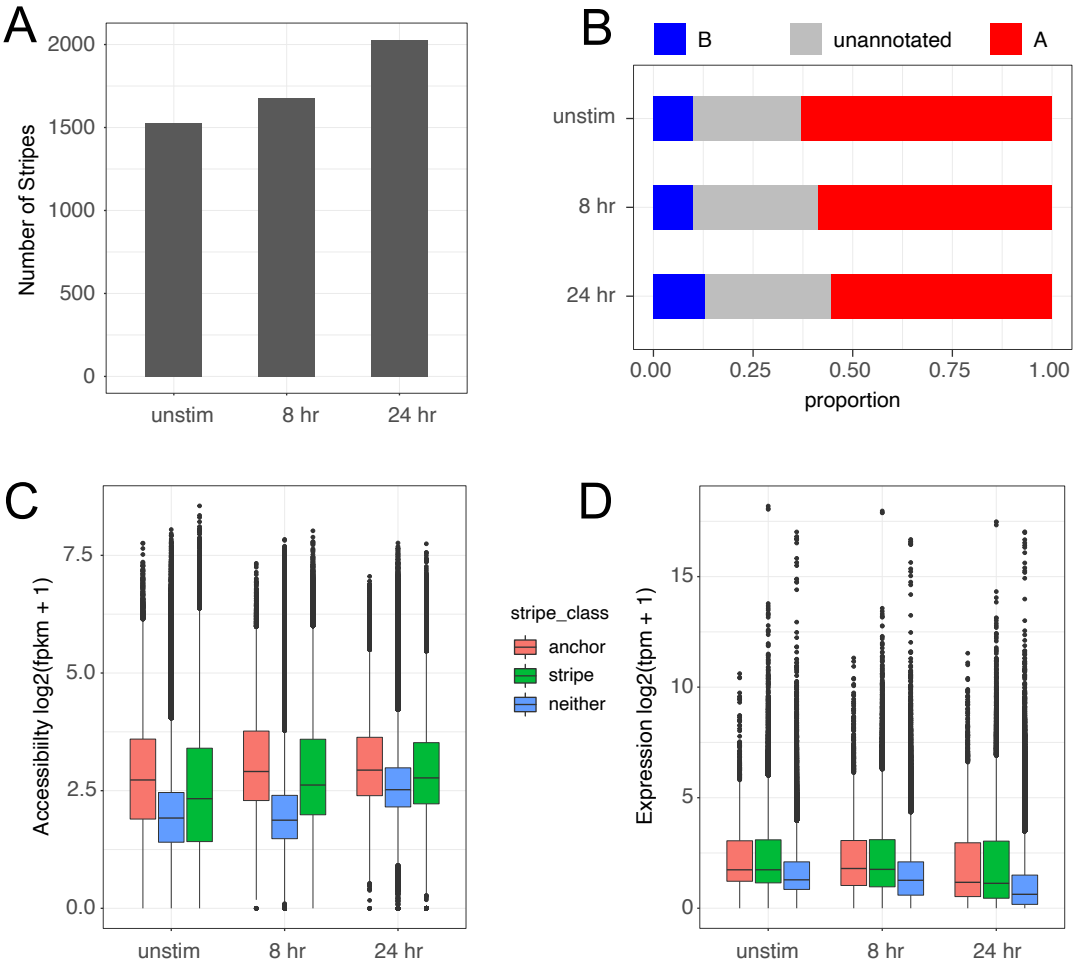

Figure S4: Large-scale epigenetic changes at dynamically expressed genes during CD4+ T cell activation. HiC contact frequency matrix (heatmap), chromatin accessibility (grey), and loop calls (red) for (A) GEM, (B) IRF4, (C) KLF2, and (D) DPEP2. The IRF4 matrix is truncated at by end of the chromosome annotation, and inset shows the expression (TPM) at each time point for each gene.

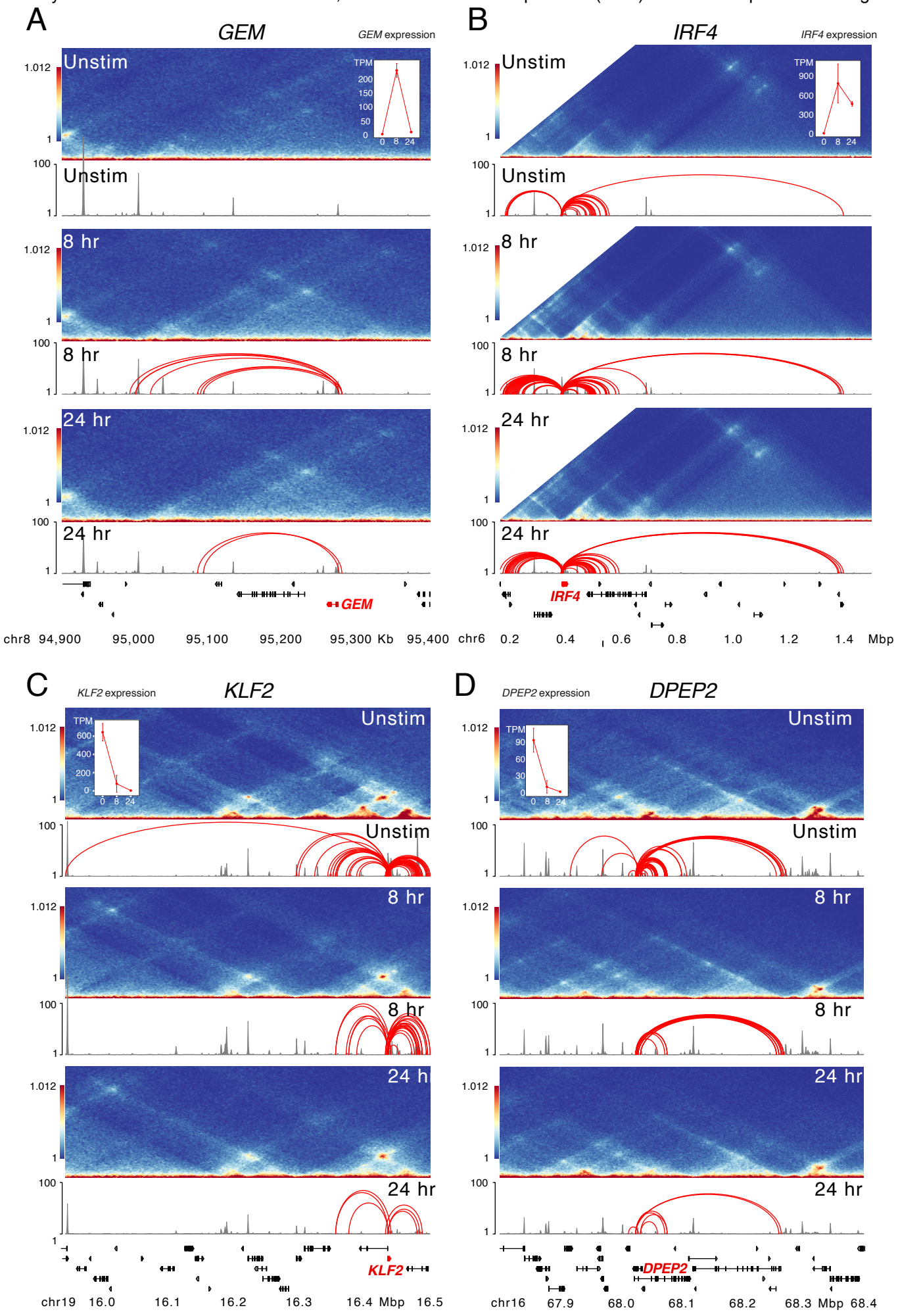

Figure S5: Enrichment of transcription family members and gene regulatory network construction. (A) -log10(P) enrichment of transcription factors annotated by family for each expression cluster. (B) Top five pathways enriched for a subset of TF-regulated gene expression.

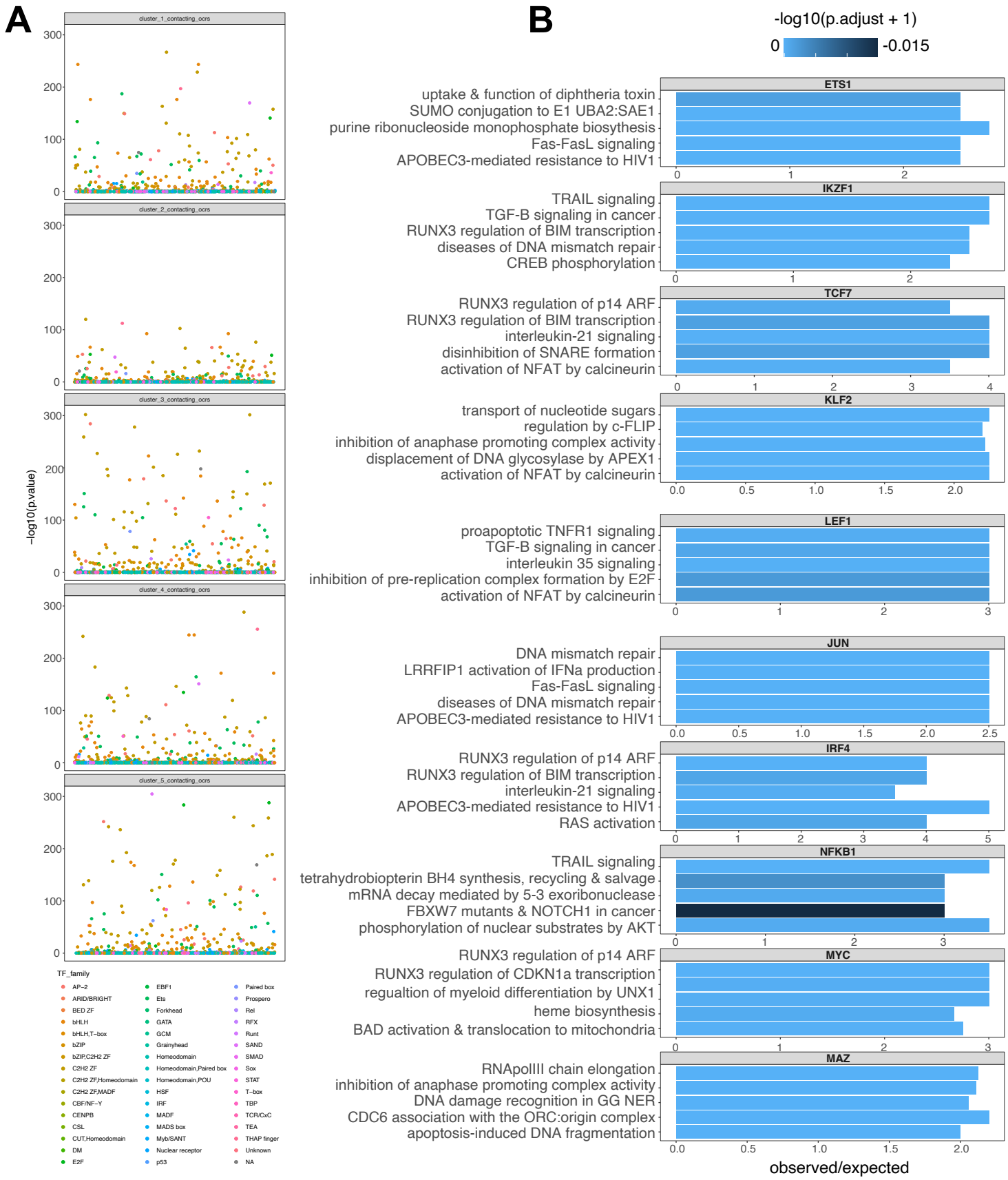

Figure S6: TF occupancy and stability of distal IL2 cRE in CD4+ T cell subsets. Evolutionary conservation (blue), open chromatin (red), and autoimmune disease-associated SNPs at the IL2 locus in quiescent (A) and 8-hour activated (B) naïve CD4+ T cells. (C) Curated TF occupancy at the IL2 URR and distal cRE from the ReMap atlas of regulatory regions.

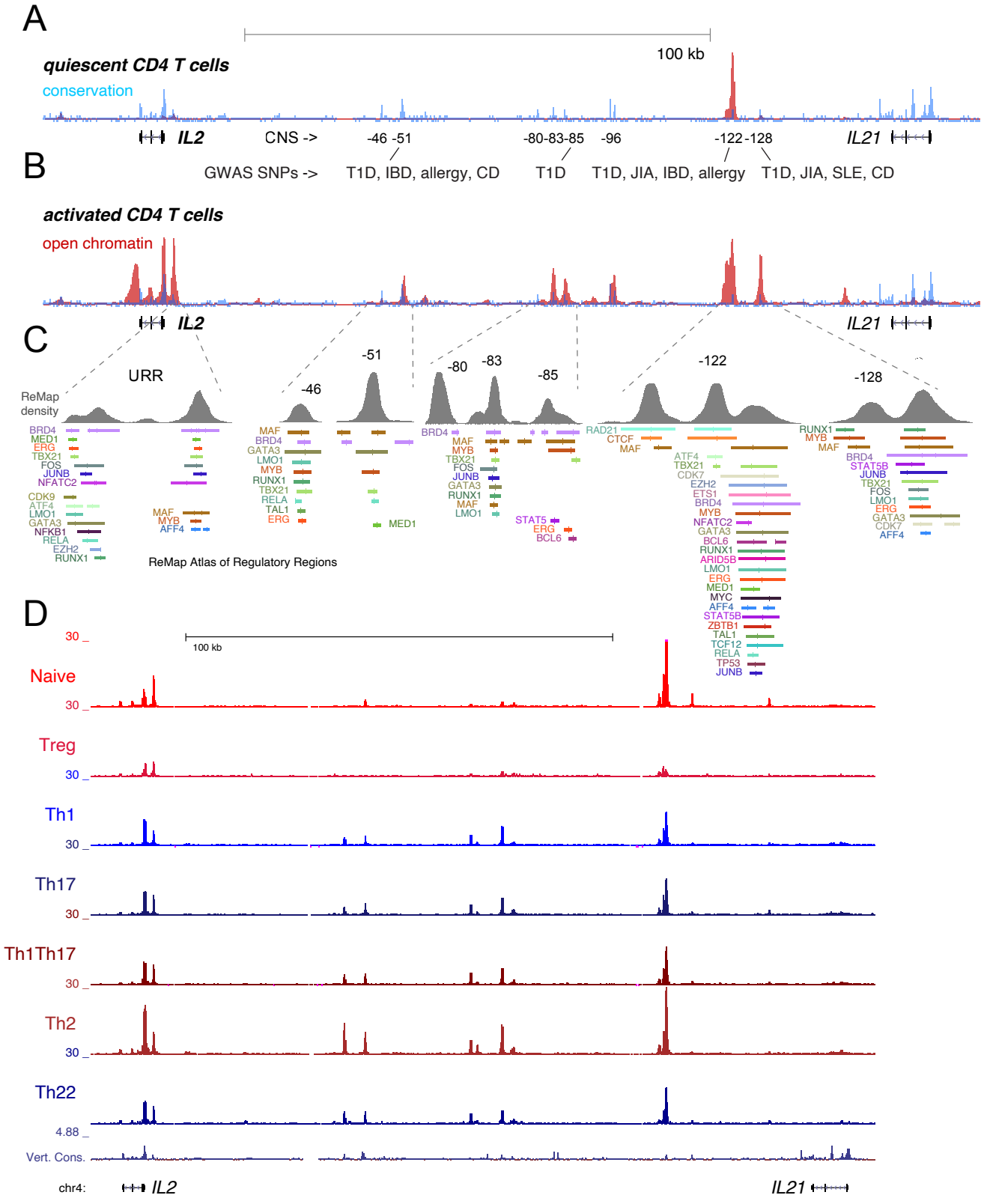

Figure S7: Orthogonal validation of 3D chromatin-based V2G genes. Expression (scaled TPM) of genes implicated by chromatin contacts in (A) prior colocalized CD4+ T cell single-cell eQTL27, (B) prior CRISPRi/a screens, and (C) genes with known immune phenotypes in the international mouse phenotyping consortium. Red = increased, blue = decreased.

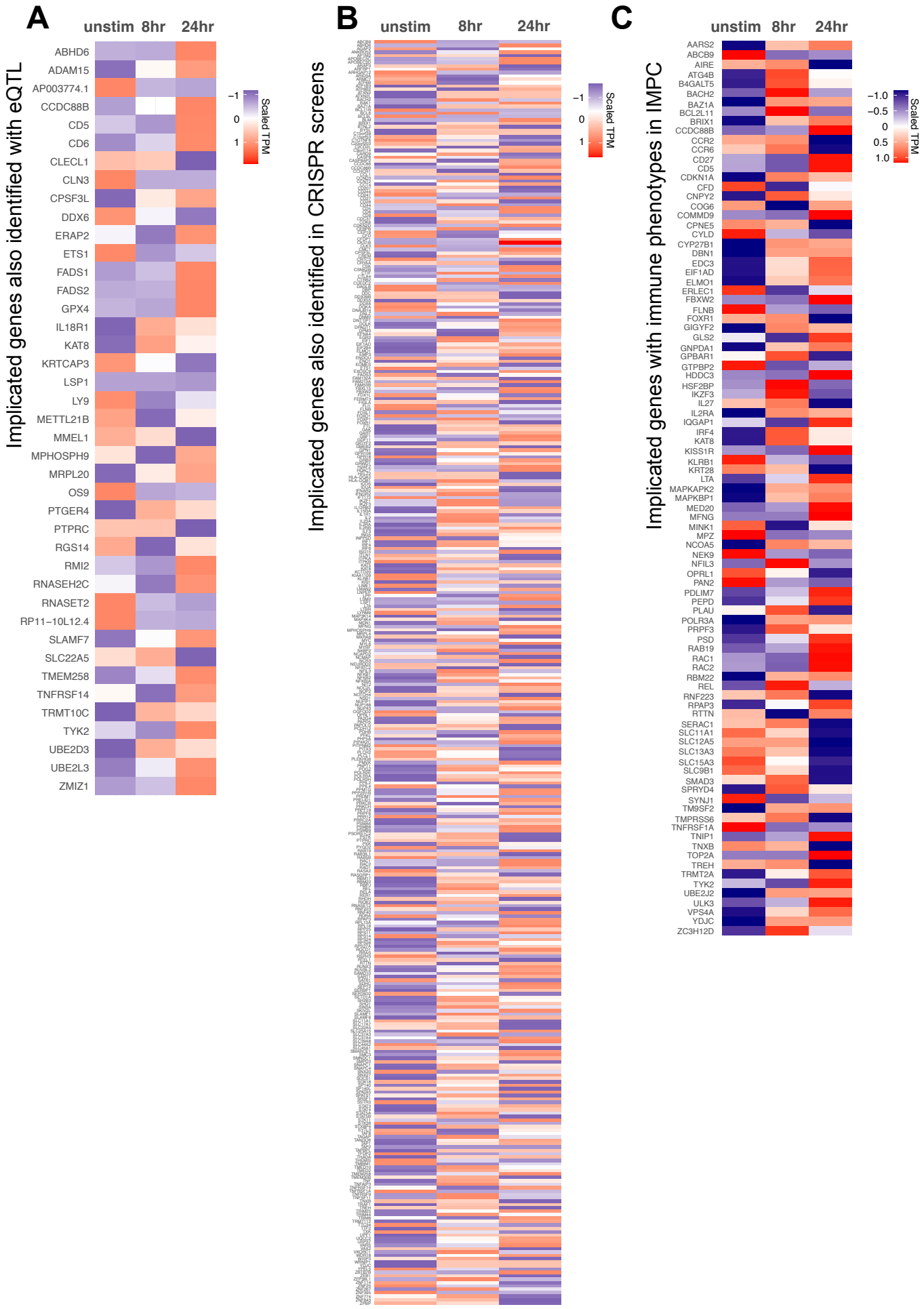
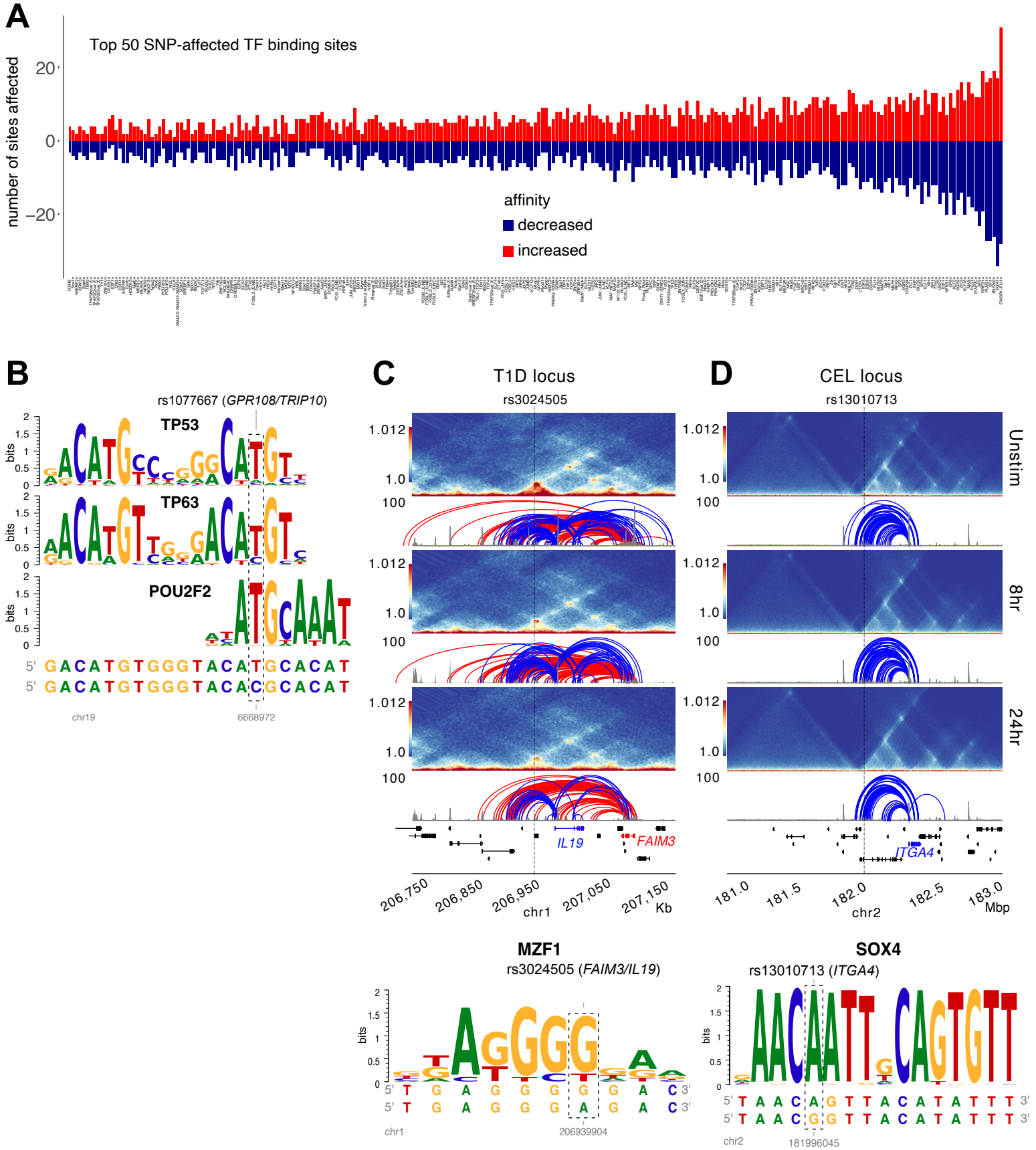
Figure S8: Predicted impact of autoimmune disease-associated genetic variation at V2G-implicated loci. (A) Top 50 TF motifs impacted by autoimmune SNPs (excerpted from Table S15). (B) Genetic variation at the MS rs1077667 SNP is predicted to disrupt TP53, TP63, and POU2F2 (OCT2) binding sites. Dashed box indicates the SNP affected by allelic variation. HiC contact frequency matrix (heatmap), chromatin accessibility (grey), loop calls (blue-IL19/ITGA4, red-FAIM3) and predicted TF motif disruption for the rs3024505-IL19-FAIM3 T1D locus (C) and the rs1301071-ITGA4 celiac locus (D). SNP rs3024505 is predicted to disrupt a MZF1 motif and rs13010713 is predicted to disrupt a SOX4 binding motif.