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

Matthew C. Pahl*[1,2], Prabhat Sharma*[1,3], Rajan M. Thomas*[1,3], Zachary Thompson[1,3], Zachary Mount[1,3], James Pippin[1,2], Peter A. Morawski[4], Peng Sun[1,3,5], Chun Su[1,2], Daniel J. Campbell[4,6], Struan F.A. Grant[1,2,7,8,9], Andrew D. Wells[1,3,10,11]

[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 Hepatobilary and Pancreatic Surgery, The Affiliated Hospital of Qingdao University
[6]Department of Immunology, University of Washington School of Medicine, Seattle, WA
[7]Department of Genetics, Perelman School of Medicine, University of Pennsylvania
[8]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania
[9]Division of Endocrinology and Diabetes, The Children's Hospital of Philadelphia
[10]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania
[11]Institute for Immunology & Immune Health, Perelman School of Medicine, University of Pennsylvania

*Co-lead authors

Correspondence:
Andrew D. Wells, Ph.D.
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 shown by high-throughput CRISPR screens to control CD4+ T cell proliferation and function, and we pharmacologically validated 8 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.

54 **INTRODUCTION**

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

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

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

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

In this study, we measured the impact of TCR-CD28 activation on the autoimmune-associated *cis*-regulatory architecture of CD4+ helper T cells, identifying hundreds of effector genes not implicated by prior studies and leading to the discovery of a stretch of evolutionarily conserved enhancers required for normal expression of the canonical T cell activation gene *IL2* whose activity is influenced by autoimmune risk variants. The set of variant-connected effector genes defined by 3D physical proximity to autoimmune-associated cRE is enriched for genes that regulate T cell activation, as validated pharmacologically in this study and by CRISPR-based screens in orthogonal studies[21–23].

99 **RESULTS**

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

101 To explore the universe of genes and cREs that are affected by and may contribute to CD4+
102 activation, we characterized the dynamics of gene expression, chromatin accessibility, and 3D
103 chromatin conformation in human CD4+CD62L^{hi}CD44^{lo} naïve T cells purified directly *ex vivo* and
104 in response to *in vitro* activation through the T cell receptor (TCR) and CD28 for 8 or 24 hours
105 using RNA-seq (N=3 donors), ATAC-seq (N=3 donors), and HiC (N=2 donors, **Fig. 1A**). As
106 sequencing depth and restriction enzyme cutting frequency affects the resolution and reliability of
107 HiC experiments[24], we constructed libraries with two four-cutter restriction enzymes and
108 sequenced to a total of approximately 4 billion unique-valid reads per timepoint. We verified the
109 reproducibility of replicate ATAC-seq and RNA-seq libraries with principal component analysis
110 and HiC libraries using distance-controlled, stratum-adjusted correlation coefficient (SCC).
111 Replicate samples were highly correlated and clustered by activation stage (**Fig. S1A-C**). As
112 expected, quiescent naïve CD4+ T cells expressed high levels of *SELL, TCF7, CCR7,* and *IL7R*,
113 and rapidly upregulated *CD69, CD44, HLADR,* and *IL2RA* upon stimulation (**Fig 1B, Table S1**).
114 Genome-scale gene set variance analysis based on MsigDB hallmark pathways showed that
115 genes involved in KRAS and Hedgehog signaling are actively enriched in quiescent naive CD4+
116 T cells, while stimulated cell transcriptomes are enriched for genes involved in cell cycle,
117 metabolism, and TNF-, IL-2/STAT5-, IFNg-, Notch-, and MTORC1-mediated signaling pathways
118 (**Table S2**).

119 To define global gene expression dynamics during the course of CD4+ T cell activation, we
120 performed pairwise comparisons and k-means clustering, identifying 4390 differentially expressed
121 genes after 8 hours of stimulation (3289 upregulated and 1101 downregulated) and 3611
122 differentially expressed genes between 8 hours and 24 hours (3015 upregulated and 596
123 downregulated, **Fig. S1D**) that could be further separated into five clusters based on their distinct
124 trajectories (**Fig. 1C**, **Fig. S1E-G**, **Table S1**). Genes upregulated early upon activation (cluster 1;

125 n=1621 genes) are enriched for pathways involved in the unfolded protein response, cytokine
126 signaling, and translation (*e.g.*, *IL2, IFNG, TNF, IL3, IL8, IL2RA, ICOS, CD40LG, FASLG, MYC,*
127 *FOS, JUNB, REL, NFKB1, STAT5A,* **Table S3**). 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,* **Table S3**). Genes downregulated early
134 (cluster 4; n=1628 genes) are enriched for pathways involved in inositol phosphate biosynthesis,
135 neutrophil degranulation, and metabolism of nucleotides (*e.g.*, *KLF2, IL7R, RORA, IL10RA,* **Table**
136 **S3**), and genes upregulated late (cluster 5; n=2154 genes) are highly enriched for pathways
137 involved in cell cycle, DNA unwinding, DNA repair, chromosomal maintenance, beta-oxidation of
138 octanoyl-CoA, and cellular response to stress (*e.g.*, *CDK2, E2F1, CDK1, CCNE1, CCNA2, PCNA,*
139 *WEE1, CDC6, ORC1, MCM2*, **Table S3**). These patterns are consistent with known changes in
140 the cellular processes that operate during T cell activation, confirming that *in vitro* stimulation of
141 CD4+ T cells recapitulates gene expression programs known to be engaged during a T cell
142 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 (**Table S4**). Of
155 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 (**Fig. 2A**, **Table S5**). The change in accessibility over
158 the next 16 hours of stimulation showed the opposite dynamic, with 6,629 regions exhibiting
159 reduced accessibility, and only 4,417 DAR becoming more open (total of 11,046 DAR, **Fig. 2A**,
160 **Table S5**). These OCR represent the set of putative cRE with dynamic activity during T cell
161 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 (**Fig S2A; Table S6**), a trend observed across all cell states. Likewise, genes
172 located in A compartments show higher average expression than those located in B
173 compartments (**Fig S2B**). A quantitative comparison across cell states showed that 94% of the
174 CD4+ T cell genome remained stably compartmentalized into A (42%) and B (52%), indicating
175 that activation does not cause a major shift in the large-scale organization of the genome within
176 the nucleus.

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 (**Fig. 2C**, **Fig. S2C** and **D**, **Table S7**). TAD boundary dynamics were
184 categorized and 2198 boundaries exhibited a change in strength, 2030 boundaries shifted
185 position, and 4697 boundaries exhibited more complex changes such as loss of a boundary
186 resulting in merger of two neighboring TAD, addition of a boundary splitting one TAD into two, or
187 a combination of any of these changes. Genes nearby dynamic TAD boundaries are enriched for
188 pathways involved in RNA metabolism, cellular response to stress, and the activity of PTEN, p53,
189 JAK/STAT, Runx and Hedgehog (**Table S6**). We also detected chromatin stripes, which are
190 TAD-like structures that consist of a genomic anchor region that is brought into contact with a
191 larger loop domain via an active extrusion mechanism. Chromatin stripes are contained within
192 and/or overlap TAD regions, and are enriched for active enhancers and super-enhancers[28,29]. We
193 identified 1526 chromatin stripes in quiescent naïve CD4+ T cells, 1676 stripes in 8 hour
194 stimulated cells, and 2028 stripes at 24 hours post-stimulation (**Fig S3A, Table S8**). Consistent
195 with prior studies in other cell types, chromatin stripes were preferentially located in A
196 compartments, and genes and OCR within stripe regions showed increased expression and
197 chromatin accessibility (**Fig. S3B-D**).

198 Within active topological structures, transcriptional enhancers can regulate the expression of
199 distant genes by looping to physically interact with gene promoters[30,31]. To identify potential
200 regulatory cRE-gene interactions, we identified high confidence loop contacts across all cell states
201 using Fit-HiC (merged 1kb, 2kb, and 4kb resolutions, **Fig. 2D**). This approach detected 933,755
202 loop contacts in quiescent naïve CD4+ T cells, 900,267 loop contacts in 8-hour stimulated cells,

203 and 551,802 contacts in 24-hour stimulated cells (2,099,627 total unique loops). Approximately
204 23% of these loops involved a gene promoter at one end and an OCR at the other, and these
205 promoter-interacting OCR were enriched for enhancer signatures based on flanking histone
206 marks from CD4+ T cells in the epigenome roadmap database[32] (**Fig. S2E**). T cell activation
207 resulted in significant reorganization of the open chromatin-promoter interactome, as 907
208 promoter-OCR exhibited increased contact and 1333 showed decreased contact following 8
209 hours of stimulation (**Fig. 2E**). Continued stimulation over the next 16 hours was associated with
210 an increase in the contact frequency of 41 promoter-OCR pairs, while only 4 pairs exhibited
211 decreased contact (**Fig. 2E**). Activation-induced changes in chromatin architecture and gene
212 expression were highly correlated, as genomic regions exhibiting increased promoter connectivity
213 became more accessible at early stages of stimulation, which was associated with increased gene
214 transcription from connected promoters (**Fig. 2F, Table S9**). The accessibility of promoter-
215 connected OCR and the expression of their associated genes decreased globally from 8 to 24
216 hours of stimulation (**Fig. 2F**). We compared these loop calls to a prior chromatin capture analysis
217 in CD4+ T cells by Burren et al.[18] and found that roughly 40% of stable loops in both stimulated
218 and unstimulated cells were identical in both studies, despite differing in approach (HiC vs.
219 PCHiC), analysis (HiC vs. CHiCAGO), sample (naïve CD4+ vs. total CD4+), timepoint (8 vs. 4
220 hour), and donor individuals (**Fig. S2F**). As expected, unstimulated samples were more similar
221 than activated samples.

222 We next focused on the 5 sets of genes with the dynamic expression patterns defined in **Fig.**
223 **1D**, and identified 57,609 OCR that contact dynamic gene promoters in at least one stage. Most
224 dynamic genes contacted between 1 and ~35 OCR, with a median of 10 OCR per gene, but a
225 handful of dynamic genes were observed in contact with over 100 distinct open chromatin regions
226 (**Fig. S2G**). Similarly, most OCR were connected to a single dynamic gene, but many contacted
227 more than one gene (median 2 genes per OCR), suggesting that most dynamic genes have a
228 complex regulatory architecture. Increased gene expression upon activation correlated with an

229 increase in the accessibility and promoter contact frequency of distant cRE (**Fig**. **S2H**), as
230 exemplified by *GEM* and *IRF4* (**Fig. S4A**,**B**). Conversely, the 3D regulatory architecture of genes
231 like KLF2 and DPEP2, which were downregulated following stimulation, exhibited decreased
232 contact and accessibility (**Fig. S4C**,**D**).
233
234 **Transcription factor footprints enriched in dynamic open chromatin identify regulators of**
235 **T cell activation**
236 To explore what factors drive dynamic changes in the regulatory architecture of the CD4+ T
237 cell genome during activation, we conducted a quantitative footprinting analysis at 1,173,159
238 transcription factor (TF) motifs located in the consensus CD4+ T cell open chromatin landscape
239 using the average accessibility of the region surrounding each motif as a measure of regulatory
240 activity (**Table S10**). Activation of naïve CD4+ T cells resulted in increased chromatin accessibility
241 around bZIP motifs at tens of thousands of genomic regions by 8 hours post-stimulation (**Fig. 3A**
242 and **C**, **Table S10**), which was associated with increased expression of Fos and Jun family
243 members constituting the AP-1 complex, as well as the bZIP factors BATF, BACH1, and BACH2
244 (**Fig. 3A**). The activity of NFE2 and SMAD was increased without increased expression (**Fig. 3A**
245 and **C**), likely due to post-translational regulation of these factors by phosphorylation[33].
246 Conversely, the motifs for a number of TF exhibited significantly reduced accessibility early after
247 stimulation, including those for EGR2 and FOXP3 that are known to negatively regulate T cell
248 activation[34,35] (**Fig. 3A**). By 24 hours post-activation, bZIP activity remained largely unchanged
249 compared to 8 hours (**Fig. 3B**), but a number of factors showed decreased activity. These include
250 several members of the Sp family, the Myc cofactor MAZ that also cooperates with CTCF to
251 regulate chromatin looping[36], KLF2, which controls genes involved in naïve CD4+ T cell
252 quiescence and homing[2,37], NRF1, a factor implicated in age-associated T cell hypofunction[38],
253 and EGR2 and 3, which are known to oppose T cell activation and promote tolerance[34,39,40] (**Fig.**
254 **3B** and **C**).

255 To explore how transcription factor activity may operate *via* the CD4+ T cell open chromatin
256 landscape to regulate distinct programs of dynamic gene expression during TCR/CD28-induced
257 activation, we focused on TF motifs enriched among those OCR specifically contacting promoters
258 of dynamic genes identified by our clustering analysis (**Fig. S5A**). The set of OCR contacting
259 dynamic gene promoters were enriched for the motifs of 89 expressed (TPM>5) transcription
260 factors, as compared to motifs present in the total open chromatin landscape (**Fig. 3D**). The
261 majority of this TF activity was enriched in OCR connected to genes highly upregulated at 24
262 hours post-activation (clusters 3 and 5, **Fig. 3D**), with the exception of CREB3, ELF1, ESRRA,
263 GABPA, RELA, XPB1, ZNF384, and the transcriptional repressor IKZF1 known as a strong
264 negative regulator of T cell activation and differentiation[41–44]. Conversely, motifs for IKZF1,
265 ZNF384, GABPA, ESRRA, and ELF1 were highly enriched in the set of OCR contacting genes
266 down-regulated early after activation (cluster 4, **Fig. 3D**). Motifs for KLF2 and the metabolic gene
267 regulator SREBF1 were likewise enriched in OCR connected to down-regulated genes. Open
268 chromatin regions interacting with genes in cluster 2 are negatively enriched for this set of TF
269 except for CTCF (**Fig. 3D**).
270 Finally, we integrated TF footprint, promoter connectome, and gene co-expression data to
271 construct TF-gene regulatory networks likely operating at each timepoint. The connections
272 between regulatory nodes are based on physical promoter-TF footprint interactions with
273 confidence weighted by gene co-expression (**Fig. 3E**, **Table S11**). Highly co-expressed genes at
274 the core of the unstimulated CD4+ T cell regulatory network encode transcription factors such as
275 KLF2, ETS1, IKZF1, and TCF7[45–47] that are known to be involved in T cell gene silencing,
276 quiescence, homeostasis, and homing. Genes connected to the top factors in this network were
277 enriched for pathways involved in immune signaling, DNA replication and repair, protein secretion,
278 and programmed cell death (**Fig S5B**). Costimulation through the TCR and CD28 induced a set
279 of core network genes active at both time points with known roles in T cell activation and
280 differentiation (*NFKB1, JUNB, MYC, IRF4, STAT5, STAT1, LEF1, ATF4*). Also part of this set is

281 *PLAGL2*, an oncogene in the Wnt pathway that regulates hypoxia-induced genes[48] with no prior
282 defined role in T cell activation. Additional nodes specifically implicated at 8 hours post-activation
283 are HIF1A, the major sensor of cellular hypoxia[49], and XBP1, a major transcriptional mediator of
284 the unfolded ER protein response with defined roles in T cell activation, differentiation, and
285 exhaustion[50,51]. Factors specifically implicated at 24 hours post-activation include E2F1, a
286 transcriptional regulator of both cell cycle and apoptosis in T cells[52,53], BHLHE40, a factor known
287 to control multiple aspects of T cell metabolism and differentiation[54], and the Myc cofactor MAZ
288 that has not been previously studied in the context of T cell function. Genes connected to factors
289 in the activated T cell networks were enriched for pathways involved in cytokine signaling, the
290 interferon response, transcription, cell cycle, DNA replication and repair, and programmed cell
291 death (**Fig S5B**). Together, these data indicate that concurrent but separable stimulation-induced
292 gene programs are the result of the activity of distinct sets of DNA binding factors mobilized by
293 antigen and costimulatory receptor signaling in naïve CD4+ T cells.
294

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

297 Following our established variant-to-gene (V2G) mapping approach to implicate functional
298 SNPs and their effector genes using the combination of GWAS and chromatin conformation
299 capture data[55–59], we intersected promoter-interacting OCR with autoimmune SNPs from the 95%
300 credible set derived from 15 autoimmune diseases (**Fig. 4A**, **Tables S12** and **S13**). Constraining
301 the GWAS SNPs in this way reduced the credible set size from an average of 14 variants per
302 sentinel to 3 variants per sentinel. To determine whether open chromatin in physical contact with
303 dynamically regulated genes in CD4+ T cells is enriched for autoimmune disease heritability, we
304 performed a partitioned LD score regression analysis. This landscape was enriched for variants
305 associated with susceptibility to inflammatory bowel disease (IBD), ulcerative colitis (UC), type I
306 diabetes (T1D), lupus (SLE), celiac disease (CEL), allergy (ALG), eczema (ECZ), and rheumatoid

307 arthritis (RA), but not for variants associated with psoriasis (PSO) or juvenile idiopathic arthritis
308 (JIA) (**Fig. 4B**, **Table S14**). The OCR connected to genes upregulated early and/or progressively
309 upon activation (clusters 1 and 3) were most strongly enriched for ALG, CEL, IBD/UC, RA and
310 T1D heritability, while SLE and ECZ heritability was most enriched in OCR connected to genes
311 upregulated later post-activation (clusters 3 and 5, **Fig. 4B**). SLE was also the only disease
312 (besides PSO and JIA) that was not enriched in open chromatin connected to down-regulated
313 genes (cluster 4, **Fig. 4B**).

314 The promoter-connected open chromatin landscape for all CD4+ T cell states in this study
315 contains 2606 putatively causal variants linked to ~half of the sentinel signals (423) for the 15
316 diseases analyzed, and are in contact with a total of 1836 genes (**Table S13**). A total of 1151
317 autoimmune variants localized to the promoters of 400 genes (-1500/+500bp from the TSS, **Table**
318 **S15**). These variants were on average 103 kb from the TSS of their connected gene (**Fig. 4C**),
319 and each variant contacted an average of 5 genes (**Fig. 4D**). The majority of linked SNPs interact
320 with genes in addition to the nearest gene, and ~half of linked SNPs 'skip' the nearest gene to
321 target only distant genes (**Fig. 4E**). Approximately 60% of connected genes were implicated
322 across all timepoints (**Fig. 4F**, **Table S13**), while ~40% (753) are dynamically regulated (clusters
323 1-5) in response to TCR/CD28 co-stimulation. Examples of SNP-genes pairs that exhibit dynamic
324 accessibility, chromosome contact, and expression in response to T cell activation are the *SIK1*,
325 *PARK7*, *DUSP5*, *CLEC2D*, *TRIP10*, *GPR108* and *IL2* loci (**Fig. 4G**, **Table S13**). The *TRIP10* and
326 *GPR108* promoters were each captured in contact with a high confidence variant rs1077667
327 (PP>0.99), which is located in an intron of *TNFSF14* and is associated with multiple sclerosis
328 (**Fig. 4H**). The accessibility of this SNP and it's contact with *TRIP10* and *GPR108* increased
329 following activation (**Fig. 4H**). Conversely, the allergy-associated SNP rs7380290 is accessible
330 and contacts the *SIK1* promoter in resting cells, but shows reduced accessibility and promoter
331 connectivity upon activation (**Fig. 4I**). *TRIP10* encodes a cytoskeletal binding protein involved in
332 endocytosis that suppresses glucose uptake in response to insulin signaling[60], *GPR108* encodes

333 an orphan G-protein coupled receptor, and *SIK1* encodes a salt-inducible kinase with roles in
334 cancer, epilepsy, and myeloid signaling[61]. None of these genes have been previously studied in
335 the context of T cell activation.
336
337 **Comparative predictive power against orthogonal variant-to-gene mapping approaches**
338 In an attempt to measure the predictive power of this approach relative to other V2G
339 approaches, we compared our chromatin capture-based autoimmune GWAS effector gene
340 predictions to the predictions of three other chromosome capture-based studies in human CD4+
341 T cells[17–19], two single-cell eQTL studies in human CD4+ T cells[13,62], and a set of 449 genes that
342 when mutated in humans cause inborn errors of immunity[63]. We integrated the chromatin-based
343 datasets with the same 95% credible set of autoimmune variants to create comparable physical
344 contact maps. Our HiC-ATAC-seq approach in naïve and activated human CD4+ implicated 1947
345 effector genes, 400 of which through proxies in open promoters and 1836 through contacts with
346 distal variants, of which 752 were dynamically regulated throughout activation. The capture-HiC
347 datasets from Burren et al., Yang et al., and Javierre et al. implicated 1408, 2572, and 2368
348 genes, respectively, and a total of 369 autoimmune GWAS effector genes were predicted in
349 common by all 3D chromatin-based approaches (**Fig. S6A**, **Table S16**). The scRNA-seq eQTL
350 studies by Soskic et al. and Schmiedel et al. implicated 171 and 221 genes, with 51 eGenes in
351 common (**Fig. S6A**). The Soskic variant-gene pairs were further co-localized with autoimmune
352 GWAS SNPs, and 41 of these eQTL were also predicted by our study (**Fig. S6B**). This
353 concordance is 10-fold higher than obtained by a random sampling of genes within 500 kb of
354 autoimmune variants (**Fig. S6C**). However, 1665 of the genes implicated in our study were not
355 identified in these statistical approaches. A total of 12 genes were implicated by all CD4+ T cell-
356 based approaches queried (APIP, ERAP2, FIGNL1, IL18R1, METTL21B, MRPL51, PPIF, PXK,
357 RMI2, RNASET2, SERPINB1, VAMP1). We find that 809 genes from our CD4 T cell V2G data

358 were not implicated by these other chromatin-based datasets, and altogether, 729 autoimmune
359 effector genes were uniquely predicted by our study.

360 To measure the predictive power and sensitivity of the sets of genes implicated by each
361 approach above, a precision-recall analysis was performed against the human inborn errors in
362 immunity (HIEI) 'gold standard' gene set. The set of genes from our study with disease-associated
363 SNPs located in their open promoters overlapped with 5% of the genes from the HIEI truth set
364 (**Fig. S6D**). However, including distal autoimmune variants captured interacting with gene
365 promoters from our HiC data increased the sensitivity of the approach 3-fold to 16% (**Fig. S6D**),
366 confirming the predictive value of long-range promoter-cRE/SNP contacts. This level of sensitivity
367 is comparable to that of other chromosome capture-based approaches for connecting SNPs to
368 genes using in CD4+ T cells (**Fig. S6D**). Conversely, the eQTL-based approaches were 3- to 4-
369 fold less sensitive than the chromatin-constrained approaches in predicting HIEI genes (**Fig.**
370 **S6D**).

371 The fraction of HIEI genes among the set of genes implicated by our CD4 V2G study is ~3.5%,
372 a predictive power greater than that of the Yang (2.5%) and Javierre (3%) approaches, but less
373 than that of Burren (**Fig. S6D**). While inclusion of variants in distal cRE improves sensitivity over
374 proximal cRE alone, it reduces precision, as the predictive power of only genes from our study
375 with accessible disease-associated SNPs in their promoters is 6%. The predictive power of our
376 activated CD4+ T cell V2G approach is comparable to that of the DICE eQTL approach (3.5%),
377 but less precise than the Soskic eQTL (**Fig. S6D**). The proportion of total HIEI genes predicted
378 by our dynamically regulated CD4 V2G gene set drops from 16% to 10%, but the precision of this
379 set doubled to ~6.2% (**Fig. S6D**), indicating that focusing the V2G set on those SNP-gene pairs
380 that are dynamically regulated by T cell activation increases predictive power. These analyses
381 indicate that chromosome capture-based V2G is a more sensitive approach for identifying true
382 effector genes than eQTL approaches, but comes with additional predicted genes that represent

383 either false positives or true effector genes for which monogenic LOF/GOF mutations have not
384 yet been characterized in humans.
385
386 **Functional validation of novel autoimmune V2G-predicted enhancers at the *IL2* locus**
387 This analysis specifically implicated potential dynamic, disease-associated regulatory
388 elements in intergenic space at the *IL2* locus. The *IL2* gene encodes a cytokine with crucial,
389 pleotropic roles in the immune system, and dysregulation of IL-2 and IL-2 receptor signaling leads
390 to immunodeficiency and autoimmune disorders in mice and humans[64–67]. Activation-induced
391 transcription of *IL2* involves an upstream regulatory region (URR) ~375 bp from the TSS that has
392 served as a paradigm of tissue-specific, inducible gene transcription for nearly four decades[68–70].
393 However, the presence of evolutionarily conserved non-coding sequences (CNS) in the ~150 kb
394 of intergenic space 46, 51, 80, 83, 85, 96, 122, and 128 kb upstream of the TSS suggest that
395 additional regulatory elements may have evolved to control *IL2*[71] (**Fig. 5A**). The -51 kb CNS
396 contains a SNP linked to T1D, IBD, PSO, CEL and allergy (rs72669153), while the -85 kb CNS
397 contains a SNP linked to RA (rs6232750) and the -128 kb CNS contains two SNPs linked to T1D,
398 JIA, and SLE (rs1512973 and rs12504008, **Fig. 5A**). In TCR/CD28-stimulated naïve CD4+ T cells,
399 these CNS are remodeled to become highly accessible (**Fig. 5A**), and they loop to interact
400 physically with the *IL2* URR (**Fig. 5B**) at both time points. ChIP-seq analyses in human T cells[72]
401 show that the URR and all distal CNS except -85 are occupied by TF such as Jun/Fos (AP-1),
402 NFAT, and NFkB that are known regulators of *IL2* (**Fig. S7**), and the -85, -122, and -128 CNS are
403 occupied by additional TF not previously thought to be direct regulators of *IL2*, such as MYC,
404 BCL6, and STAT5 (**Fig. S7**). Recombinant reporter assays in primary activated human CD4+ T
405 cells showed that the -46, -51, -83, and -128 CNS/OCR can enhance transcription from the URR
406 (**Fig. 5C**). To determine whether the native elements contribute to the expression of *IL2*, we
407 targeted each CNS/OCR individually in primary human CD4+ T cells or Jurkat T cells using
408 CRISPR/CAS9 (**Fig. 5D**) and measured IL-2 secretion following TCR/CD28 stimulation (**Fig. 5E**).

409 Deletion of the -46, -51, -83, -85, -122, and -128 kb elements in primary human CD4+ T cells each
410 resulted in a ~50% reduction in IL-2 production, while deletion of the -80 kb element had little
411 effect. A very similar pattern of impact was observed when these elements were deleted
412 individually in Jurkat T cells (**Fig. 5E**). The URR has a stronger contribution to IL-2 production
413 than any individual intergenic element, as deletion of the URR almost completely abrogated
414 activation-induced IL-2 production by both primary CD4+ or Jurkat T cells (**Fig. 5E**). To determine
415 whether these intergenic enhancers exist in synergistic epistasis[73] necessary for *IL2*
416 transactivation, we generated Jurkat T cell clones in which the stretch of all 7 elements located -
417 46 kb to -128 kb upstream of *IL2* was deleted using CRISPR/CAS9 (**Fig. 5F**). Despite the URR
418 and 46 kb of upstream sequence being intact in these clones, loss of the 81.3 kb stretch of
419 intergenic enhancers renders these cells incapable of expressing *IL2* at both the mRNA and
420 protein level in response to stimulation (**Fig. 5G**). These results show that the URR is not sufficient
421 for activation-induced expression of *IL2*, and that *IL2* has a previously unappreciated, complex,
422 and autoimmune disease-associated regulatory architecture that was accurately predicted by our
423 3D epigenomic V2G approach. Importantly, we find that the distal *IL2* cRE are highly accessible
424 in quiescent memory T cell subsets (Th1, Th2, Th17, Th1-17, Th22) isolated directly *ex vivo* from
425 human blood, whereas naïve CD4+ T cells and non-IL-2-producing Treg showed little accessibility
426 at these elements (**Fig. S7**). This suggests that stable remodeling of distal *IL2* cRE can persist *in*
427 *vivo* after TCR signals cease, and that this epigenetic imprinting contributes to the immediate,
428 activation-induced production of IL-2 exhibited by memory, but not naïve or regulatory, CD4+ T
429 cells[74–76].
430

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

433 The intergenic *IL2* enhancers defined above are conserved at the sequence level between
434 human and mouse (**Fig. 6A**). To test whether enhancer function is likewise evolutionarily

435 conserved, we used zygotic CRISPR/CAS9 targeting to generate mice with a 583 bp deletion of
436 the ortholog of the -128 kb human enhancer CNS/OCR situated 83 kb upstream from mouse *Il2*
437 (*Il2*-83-cRE-ko, C57BL6 background). Deletion of this genomic region did not discernably affect
438 T lymphocyte development in *Il2*-83-cRE-ko mice. However, peripheral $CD4^+$ T cells from *Il2*-83-
439 cRE-ko mice produced substantially less IL-2 at both the protein and mRNA level, and exhibited
440 reduced proliferation, in response to *in vitro* stimulation (**Fig. 6B**), confirming that the enhancer
441 function of the orthologous -128 and -83 kb elements is conserved between human and mouse.
442 The *in vitro* induction of Foxp3+ regulatory T cells (Treg) from conventional CD4+ T cell precursors
443 in response to antigenic stimulation in the present of TGF-beta, a differentiation process highly
444 dependent upon IL-2, was reduced 4-fold in *Il2*-83-cRE-ko mice compared to wild-type mice (**Fig.**
445 **6C**), but could be rescued by addition of exogenous IL-2 to the culture (**Fig. 6C**).

446 To test whether the distal -83 kb *Il2* enhancer contributes to physiologic immune responses
447 *in vivo*, we immunized wild-type and *Il2*-83-cRE-ko mice with the model antigen chicken
448 ovalbumin. *Il2*-83-cRE-ko animals showed a nominal increase in the differentiation of CD4+ T
449 cells into CXCR5+PD-1^{hi} follicular helper T cells (Tfh) in the spleen compared to wild-type animals
450 (**Fig. 6D**), a process known to be antagonized by IL-2, and generated significantly elevated levels
451 of ovalbumin-specific IgG antibody following immunization (**Fig. 6E**). To determine whether the
452 83 kb *Il2* enhancer contributes to auto-inflammatory disease susceptibility *in vivo*, we used an
453 adoptive transfer model of IBD in which the development of colitis is determined by the balance
454 between conventional and regulatory CD4+ T cell activities[77,78]. Upon transfer into lymphocyte-
455 deficient Rag1-ko recipients, wild-type CD4+ T cells respond to microbiota in the gut by inducing
456 inflammatory colitis that leads to weight loss (**Fig. 6F**, WT Tconv), while co-transfer of wild-type
457 Treg effectively controls inflammation (**Fig. 6F**, WT Tconv + WT Treg). Transfer of *Il2*-83-cRE-ko
458 Tconv led to significantly less colitis (**Fig. 6F**, *Il2*-83-cRE-ko Tconv), and was associated with a
459 nominal decrease in *Il2*-83-cRE-ko Tconv in the spleen, and a significant decrease in the
460 accumulation of *Il2*-83-cRE-ko Tconv in the mesenteric lymph nodes (MLN) that drain the

461 intestines (**Fig. 6F** inset). Treg require IL-2 from Tconv for their function and homeostasis[79], and
462 despite their reduced numbers, *Il2*-83-cRE-ko Tconv were able to support the homeostasis and
463 function of co-transferred wild-type Treg (**Fig. 6F**, *Il2*-83-cRE-ko Tconv + WT Treg). That the -83
464 kb ortholog of the human -128 kb *IL2* enhancer contributes to physiologic immune responses and
465 inflammatory disease susceptibility *in vivo.*
466
467 **Impact of autoimmune risk-associated genetic variation on *cis*-regulatory element activity**
468 The chromatin conformation approach used here employs GWAS variants as 'signposts' to
469 identify disease-relevant regulatory elements and connect them to their target genes, but does
470 not *per se* determine the effect of disease-associated genetic variation on enhancer activity or
471 gene expression. We experimentally measured variant effects at the *IL2* locus, where the -128
472 enhancer defined above contains two SNPs linked to T1D, JIA, and SLE (rs1512973 and
473 rs12504008). Using a recombinant reporter assay in primary activated CD4+ T cells, we
474 confirmed that disease-associated genetic variation influences intergenic *IL2* enhancer activity at
475 the -128 kb element, in that the risk allele contributes significantly less transcriptional activity than
476 the reference allele (**Fig. 7A**).
477 Autoimmune variants are likely to influence disease risk by altering the activity of *cis*-
478 regulatory elements in T cells. At genome scale, we identified over 1000 cRE likely impacted by
479 autoimmune disease-associated genetic variation, in that they contain autoimmune risk SNPs
480 that are predicted to decrease or increase transcription factor binding affinity. Overall, 1370
481 autoimmune risk variants in open chromatin were predicted to influence the activity of 495 DNA
482 binding factors (**Table S17**), including PLAG1, PRDM1, BACH2, MYC, TBX21, BHLHE40, LEF1,
483 TCF7, BCL6, IRF, p53, STAT (**Fig. S8A**), and hundreds of NFkB, EGR, KLF, FKH/FOX, and FOS-
484 JUN sites (**Fig. 7B**). For example, the T1D SNP rs3024505 (PP = 0.200) connected to the
485 promoters of *IL9* and *FAIM3* (**Fig. S8B**) and the celiac SNP rs13010713 (PP = 0.154) contacting
486 the *ITGA4* promoter (**Fig. S8C**) are predicted to disrupt binding sites for *MZF1* (**Fig. S8D**) and

487 SOX4 (**Fig. S8E**). Similarly, the MS SNP rs1077667 contacting the promoters of *GPR108* and
488 *TRIP10* (**Fig. 4H**) is predicted to reduce affinity for TP53, TP63, and OCT2/POU2F2 (**Fig. S8F**).
489

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

491 To determine whether genes identified *via* their physical interaction with autoimmune variants
492 in CD4+ T cell chromatin contact maps tend to be directly involved in T cell activation and function,
493 we compared the set of autoimmune genes implicated by chromatin contacts in this study to sets
494 of genes identified in CRISPR-based screens that control aspects of CD4+ T cell activation like
495 proliferation and expression of the inflammatory genes *IFNG, CTLA4, IL2, IL2RA,* and *TNF*[21–23].
496 The set of all V2G-implicated genes was highly enriched for genes shown to regulate IL-2, IL-2
497 receptor, CTLA-4, and proliferation (**Fig. 7C, Fig. S9A, Table S18**). For example, 202 genes
498 shown to regulate IL-2 production and 166 genes shown to regulate proliferation were also
499 implicated in our autoimmune V2G set (**Fig. S9A, Table S16**). Genes implicated by V2G in
500 activated CD4+ T cells were moderately enriched for genes known to control the production of
501 IFN-G, but at the individual disease level, only genes connected to CRO- and ATD-associated
502 variants were enriched for IFNG regulatory genes (**Fig. 7C, Fig. S9A**). Genes contacting CRO,
503 PSO, RA, SLE, T1D and VIT variants were moderately enriched for TNF regulatory genes, but
504 the set of all V2G genes was not enriched for TNF genes. We also queried the orthologs of our
505 V2G-implicated genes from the international mouse phenotype consortium (IMPC) database and
506 identified 97 genes that when knocked out give an immune phenotype and 126 V2G genes that
507 result in a hematopoietic phenotype (**Fig. 7D**, **Fig. S9B**, **Table S18**). This frequency of observed
508 immune/hematopoietic phenotypes represents a significant (adjP<0.05) ~30% enrichment over
509 expected. This gene set was also enriched for mortality, homeostasis/metabolism, growth/body
510 size, skeleton, and embryonic phenotypes (**Fig. 7D**, **Table S18**).

511 An important application of this V2G approach is the identification of novel regulators of T cell
512 activation and their potential as drug targets, as nearly 20% of implicated genes have at least one

513 chemical modulator currently available (**Fig 7E**, **Table S19**). As shown above, this approach
514 validates genes that are well-studied regulators of T cell function, however, a significant portion
515 of implicated genes are not well-studied and are not currently known to regulate T cell activation
516 (**Fig. 7F**). We observed a trend that genes expressed more highly in immune tissues (GTEX)
517 have in general been better investigated, and identified several less-studied genes that could be
518 novel targets, including the kinases GRK6, PTK6, SIK1, and MAP3K11, the G protein-coupled
519 receptors OXER1, GPR183, GPR18, and KISS1R, the acetylcholine receptor CHRNB1, and the
520 *de novo* purine pathway enzyme GART. To determine whether these V2G-implicated genes are
521 novel regulators of T cell activation, we used commercially available pharmacologic modulators
522 in dose-response assays of activation-induced T cell proliferation. Stimulation of T cells in the
523 presence of ligands for CHRNB1, KISS1R and OXER1 did not significantly affect T cell
524 proliferation, however, small molecules targeting GRK6, PTK6, MAP3K11, GPR183, GART, and
525 SIK1 inhibited T cell activation in the nanomolar to micromolar range (**Fig. 7G**). Together, these
526 data show that maps of dynamic, 3D variant-gene chromatin contacts in stimulated CD4+ T cells
527 are able to identify genes with *bona fide* roles in T cell activation.
528

529 **DISCUSSION**

530 By measuring dynamic changes in chromosome folding, chromatin accessibility, and gene
531 expression in naïve CD4+ T cells as a function of TCR-CD28 costimulation, we identified the
532 putative *cis*-regulatory landscape of autoimmune disease-associated genetic variation, and
533 physically connected these elements to their putative effector genes. We conducted
534 pharmacologic targeting experiments and compared our results with orthogonal eQTL and
535 CRISPR-based studies to validate sets of effector genes with a confluence of evidence supporting
536 their role in CD4+ T cell activation. This and prior chromosome capture-based studies show that
537 most cRE and their variants interact with, and therefore have the potential to regulate, more than
538 one gene (median 5 in this study), supporting a scenario in which multiple effector genes are

539 operative at a GWAS locus. We also observe that the vast majority (87-95%) of GWAS effector
540 genes predicted by chromosome capture-based approaches are not the genes nearest to the 614
541 sentinel SNPs queried in this study, while roughly one-third of eGenes predicted by eQTL
542 approaches are the nearest to a sentinel. Of the roughly 2000 total genes implicated by both
543 eQTL and our chromatin/GWAS-based V2G, only 41 were predicted by both approaches, which
544 is consistent with the view that eQTL-based and GWAS-based approaches inherently implicate
545 different types of genes. While eQTLs cluster strongly near transcription start sites of genes with
546 simple regulatory structures and low enrichment for functional annotations, GWAS variants are
547 generally far from the TSS of genes with complex regulatory landscapes[80]. Moreover, eGene
548 discovery by eQTL studies using scRNA-seq approaches are significantly limited by the low gene
549 detection power inherent to these methods, particularly in rare cell types. Chromatin conformation
550 mapping in disease-relevant cell types therefore brings a crucial added dimension to 1D GWAS
551 data and eQTL findings for understanding the complex genetic and epigenetic mechanisms that
552 regulate gene expression and for predicting novel therapeutic targets.

553 Why did our V2G approach not capture a larger proportion of genes from the human inborn
554 errors in immunity 'truth set'? Low recall/sensitivity could result from the fact that a substantial
555 portion of the mutations in the full 449 HIEI gene set contains disease genes that do not result in
556 an autoimmune phenotype. However, restricting the HIEI truth set to only those disease gene
557 mutations that result in an autoimmune phenotype (143 genes) only increased recall from 16% to
558 17.5%, suggesting that this is not a major factor reducing sensitivity. Also, our method uses
559 GWAS signals as an input, and unlike GWAS for height, BMI, blood traits, etc., most autoimmune
560 GWAS signals are likely not saturated, which limits discoverability. Another reason for reduced
561 sensitivity is the likelihood that many of the genes in the HIEI truth set operate in cells other than
562 CD4+ T cells, and consistent with this, when the Javierre pcHiC V2G is extended to all immune
563 cell types tested, recall increases from 17% to 27%). This observation emphasizes how limited
564 inclusion cell types and states in a study can significantly limit the power to detect autoimmune

565 effector genes. Why are so few of our V2G genes present in the HIEI truth set? Low precision or
566 predictive power could be due to false positives in the V2G gene set; *i.e.*, contacts between
567 disease-associated cRE and genes that are not in fact involved in disease susceptibility. This
568 does not argue *per se* against the biological relevance of the cRE or the gene, only that the linkage
569 to disease susceptibility is not true. Alternatively, the absence of chromatin/GWAS-implicated
570 genes in the truth set could be a false negative from the point of view of the truth set; *i.e.*,
571 monogenic, disease-causing mutations in these genes may exist in the human population, but
572 have not yet been discovered. For example, in the year 2000 approximately 100 monogenic
573 mutations causing human inborn errors of immunity were known, but this increased by ~11
574 disease mutations per year until 300 human inborn errors of immunity had been identified by
575 2017. The advent of next-generation exome/genome sequencing and the COVID-19 pandemic
576 resulted in an increase in the rate of HIEI discovery between 2018-2021 to ~40 per year to the
577 current recognized HIEI of ~450. If we assume a conservative continued rate of HIEI discovery of
578 25 disease genes per year, the next 10 years will show us an additional ~250 disease genes that
579 are not currently contained in the truth set.

580 Many of the genes identified in this 3D epigenomic V2G screen have known roles in T cell
581 activation and function. An example is *IL2*, and we used the resulting maps to identify and validate
582 a stretch of previously unknown distal enhancers whose activity is required for *IL2* expression and
583 is influenced by autoimmune genetic variation. Another example is the phosphatase DUSP5 that
584 regulates MAPK signaling during T cell activation[81,82]. However, roles for many of the genes
585 implicated here in T cell activation are not known. For example, one of the top implicated genes,
586 PARK7, is a deglycase studied in the context of Parkinson's disease, but has a recently been
587 shown to modulate regulatory T cell function[83]. The orphan G protein-coupled receptor GPR108
588 is another top gene uniquely implicated by our chromatin-based V2G that has not been studied
589 in T cells, but was identified in a recent CRISPR screen for genes affecting IL-2 levels[23]. Also co-
590 implicated by our study and recent CRISPR screens are the cannabidiol receptor GPR18[84] and

591 the purine biosynthetic enzyme GART[85]. The GPR18 agonist arachidonyl glycine inhibited CD4+
592 T cell activation above 10 uM, while the GPR18 antagonist O-1918 slightly enhanced T cell
593 activation. This GPR was implicated by both chromatin- and eQTL-based approaches.
594 Antagonism of GART, an enzyme we previous identified as a V2G effector gene in COVID19
595 severity[58], with the FDA-approved drug lometrexol inhibited T cell activation in the 10 nM range.
596 Antagonism of GRK6, a member of the G-coupled receptor kinase family associated with insulin
597 secretion and T2D susceptibility[86], and PTK6, an oncogenic kinase studied in the context of
598 cancer[87], led to inhibition of T cell activation in the nM to uM range. These targets were implicated
599 by the other chromatin-based approaches but not by eQTL. Inhibition of MAP3K11, a kinase that
600 facilitates signaling through JNK1 and IkappaB kinase in cancer[88,89] inhibited stimulation-induced
601 CD4+ T cell proliferation in the 100 nM range, as did 25-hydroxycholesterol, a ligand of the G
602 protein-coupled receptor GPR183 that was implicated by both chromatin- and eQTL-based
603 approaches. SIK1 is a member of the salt-inducible kinase family uniquely implicated in our study
604 that negatively regulates TORC activity[61], and a small molecule SIK1 inhibitor potently
605 antagonized stimulation-induced CD4+ T cell activation in the pM range.
606 Our integration of high-resolution Hi-C, ATAC-seq, RNA-seq and GWAS data in a single
607 immune cell type across multiple activation states identified hundreds of autoimmune variant-
608 gene pairs at ~half of all GWAS loci studied, and application of this technique to additional immune
609 cell types will likely identify effector genes at many of the remaining loci. This study highlights the
610 value of chromosome conformation data as a powerful biological constraint for focusing variant-
611 to-gene mapping efforts[90], and shows that dynamic changes in the spatial conformation of the
612 genome that accompany cell state transitions alter gene expression by exposing promoters to a
613 varying array of *cis*-regulatory elements, transcription factors, and genetic variants.
614
615
616

617 **MATERIALS AND METHODS**

618 T cell isolation and *in vitro* stimulation

619 Human primary CD4+ T cells were purified from the apheresis products obtained from healthy,
620 screened human donors through University of Pennsylvania Human Immunology Core (HIC).
621 Naïve CD4+ T cells were purified using EasySepTM human naïve CD4+ T cell isolation kit II
622 (STEM cells Technologies, cat#17555) by immunomagnetic negative selection as per
623 manufacturer's protocol. Isolated untouched, highly purified (93-98%) naïve human CD4 T cells
624 were activated using anti-CD3+anti-CD28 Dynabeads (1:1, Thermofisher scientific, cat # 11161D)
625 for 8-24 hours. Cells were then used to prepare sequencing libraries. The human leukemic T cell
626 line Jurkat was obtained from ATCC, cloned by limiting dilution, and clones with high activation-
627 induced secretion of IL-2 were selected for further study. Single-cell mouse lymphocyte
628 suspensions were prepared from spleen and lymph nodes isolated from 6-week-old female wild-
629 type or *Il2*-83cRE-ko mice (C57BL6 background). Mouse CD4+ T cells were purified by negative
630 selection using a CD4+ T cell isolation kit (Miltenyi Biotec, cat. # 130-104-454). For CD8 depletion,
631 CD8+ T cells were positively stained using CD8a (Ly-2) microbeads (Miltenyi Biotec, cat. # 130-
632 117-044) and the negative flow-through fraction was collected. Tregs were purified from the
633 mouse lymphocytes using a CD4+CD25+ regulatory T cell isolation kit (Miltenyi Biotec, cat. # 130-
634 091-041). The purity of the isolated cells was checked by flow cytometry, showing approximately
635 95% cell purity. For iTreg induction, CD4+ CD25- T cells (1x 10^6) were activated in a 24-well plate
636 pre-coated with anti-CD3 (1 µg/mL) and the cells were cultured for 72 hours in RPMI culture
637 medium with soluble anti-CD3 (0.5 µg/mL), IL-2 (25 units/mL), TGF-beta (3 ng/mL), anti-IFN-
638 gamma (5 µg/mL), and anti-IL-4 (5 µg/mL). iTreg induction cultures were also set up without
639 adding IL-2. iTreg cultures were harvested after 72 hours, and cells were stained for Foxp3
640 expression.

641

642 *Il2*-83-cRE ko Mice

643 The CRISPR/CAS method was employed to delete the intergenic -83 CNS sequence between *Il2*
644 and *Il21*. CRISPR guide RNAs were designed (two guide RNAs upstream of -83CNS) and one
645 guide RNA (downstream of -83CNS) using guide RNA design tools (http://crispr.mit.edu.guides).
646 The pX335 plasmid (Addgene #42335) was used to generate a DNA template for sgRNA *in vitro*
647 transcription. To *in vitro* transcribe the sgRNA, a T7 sequence was incorporated at the 5' end of
648 the guide RNA sequence, and a part of the sgRNA scaffold sequence from px335 was added at
649 the 3' end of the guide RNA. Oligonucleotide sequences containing the guide RNA (underlined)
650 along with these added sequences were synthesized by IDT. The oligonucleotide sequences
651 were as follows:
652 T7_guideRNA#1_scaffold#:
653 TTAATACGACTCACTATAGG<u>TTTTCCACGGATCTGCTCGG</u>GTTTTAGAGCTAGAAATAGC
654 T7_guideRNA#2_scaffold#:
655 TTAATACGACTCACTATAGG<u>TGCTTTCTAGGTGAAGCCCC</u>GTTTTAGAGCTAGAAATAGC
656 T7_guideRNA#3_scaffold#:
657 TTAATACGACTCACTATAGGT<u>CATTTGAGCCTAACTACTC</u>GTTTTAGAGCTAGAAATAGC
658 Along with the above-cited sequences, a reverse primer sequence
659 (AGCACCGACTCGGTGCCACT) from PX335 was used to amplify a PCR-generated template
660 for sgRNA *in vitro* transcription. The PCR product (~ 117 bp) was verified on a 2% agarose gel
661 and then gel-purified using the QIAQuick gel extraction kit. The T7 sgRNA PCR product (500 ng)
662 was used as a template for *in vitro* transcription with the T7 High Yield RNA synthesis kit (NEB,
663 cat # E2040S). The reaction was incubated at 37°C for 4 hours, and then the sgRNA was purified
664 using the MEGAclear kit (Life Technologies, cat # AM1908) according to the kit protocol. The
665 sgRNA was eluted with elution buffer preheated to 95°C. The eluted sgRNA was centrifuged at
666 13,000 rpm for 20 minutes at 4°C, and the suspension was transferred into a new RNase-free
667 tube. The sgRNA quality was checked by a bioanalyzer using an RNA nano Chip. The sgRNA
668 was diluted to 500 ng/μl in injection buffer (sterile 10 mM Tris/0.1 mM EDTA, pH 7.5) and stored

669 in a -80°C freezer until use. An injection mixture (final volume 30 µl) was prepared in injection
670 buffer by mixing 500 ng/µl of each of the sgRNAs (left and right) with Cas9 mRNA (1 µg/µl, TriLink,
671 cat # L-6125). Fertilized eggs collected from B6/129 mice were microinjected at the CHOP
672 transgenic core and transferred into pseudo-pregnant B6 females. The pups were genotyped
673 using primers flanking the targeted sequence (Forward primer: TTAGGACCTCACCCATCACAA
674 and reverse primer: CATGCCCAGCTACTCTGACAT). The PCR product was cloned into the
675 Promega PGEM T Easy TA cloning vector, and plasmid DNA was Sanger sequenced to
676 determine the size of the deletion. The targeting resulted in mice with a 500 bp and 583 bp deletion
677 at the targeted -83 CNS site and these mutant mice showed same phenotype. The *Il2*-83-cRE
678 mutant heterozygous male mice were backcrossed with B6 females for 10 successive
679 generations.
680
681 ELISA
682 Cell culture supernatants collected at various time intervals *from in vitro* stimulated T cells of mice
683 or humans were analyzed for IL-2 by ELISA using kits purchased from ThermoFisher Scientific
684 (mouse IL-2 ELISA kit: cat # 88-7024-88) and human IL-2 ELISA kit, cat # 88-7025-76). IL-2
685 ELISA was performed following the protocol provided by the vendor.
686
687 *In vivo* ovalbumin immunization
688 Eight week-old female WT and Il2-CNS-83 KO mice were injected intraperitoneally with 50 µg of
689 ovalbumin (Sigma cat # A5503-5G) mixed with IFA (Sigma cat # F5506). Blood was collected
690 from the immunized mice after 10 days, and serum was separated. The level of ovalbumin-
691 specific IgG in the serum was determined by ELISA using an ovalbumin-specific IgG ELISA kit
692 purchased from MyBiosource (cat # MBS763696). The immunized mice were sacrificed on day
693 10 of immunization, and spleen and lymph nodes were collected. Splenocytes and lymph node
694 cells were stained with CD4 BV785, CD25 BV650, CD44 Percp Cy5.5, CD62L APC eFL780,

695 CXCR5 BV421, PD-1 APC, Bcl-6 PE antibodies. TFH frequency was determined by flow
696 cytometry analysis.
697
698 *In vivo* inflammatory colitis model
699 Conventional CD4+CD25-ve T cells and CD4+CD25+ Tregs were purified from spleens and
700 lymph nodes of male wild-type or Il2-CNS-83 ko mice (6-8 weeks of age). T cells were transferred
701 into Rag1-ko male mice (5 per group) by retro-orbital injection of 1 million wild-type or Il2-CNS-83
702 ko Tconv cells alone or along with (0.25×10^6) wild-type Tregs. Experimental Rag1-ko recipients
703 were weighed 3 times per week and scored for IBD-induced clinical symptoms. Mice were
704 sacrificed 72 days post-transfer, and spleen, mesenteric lymph nodes, and colon were collected.
705 Single-cell suspensions were prepared, cells were stained for CD4, CD8, CD25, CD44, and
706 Foxp3, and analyzed by flow cytometry. The absolute T cell count was estimated using cell count
707 and cell frequency derived from the flow cytometry analysis.
708
709 RNA-seq library generation and sequencing
710 RNA was isolated from ~ 1 million of each cell stage using Trizol Re- agent (Invitrogen), purified
711 using the Directzol RNA Miniprep Kit (Zymo Research), and depleted of contaminating genomic
712 DNA using DNAse I. Purified RNA was checked for quality on a Bioanlayzer 2100 using the Nano
713 RNA Chip and samples with RIN>7 were used for RNA-seq library preparation. RNA samples
714 were depleted of rRNA using QIAseq Fastselect RNA removal kit (Qiagen). Samples were then
715 processed for the preparation of libraries using the SMARTer Stranded Total RNA Sample Prep
716 Kit (Takara Bio USA) according to the manufacturer's instructions. Briefly, the purified first-strand
717 cDNA is amplified into RNA-seq libraries using SeqAmp DNA Polymerase and the Forward and
718 the Reverse PCR Primers from the Illumina Indexing Primer Set HT for Illumina. Quality and
719 quantity of the libraries was assessed using the Agilent 2100 Bioanalyzer system and Qubit

720 fluorometer (Life Technologies). Sequencing was performed on the NovaSeq 6000 platform at
721 the CHOP Center for Spatial and Functional Genomics.
722
723 ATAC-seq library generation and sequencing
724 A total of 50,000 to 100,000 sorted cells were centrifuged at 550 g for 5 min at 4 °C. The cell pellet
725 was washed with cold PBS and resuspended in 50 µL cold lysis buffer (10 mM Tris-HCl, pH 7.4,
726 10 mM NaCl, 3 mM MgCl2, 0.1% NP-40/IGEPAL CA-630) and immediately centrifuged at 550 g
727 for 10 min at 4 °C. Nuclei were resuspended in the Nextera transposition reaction mix (25 µL 2x
728 TD Buffer, 2.5 µL Nextera Tn5 transposase (Illumina Cat #FC-121-1030), and 22.5 µL nuclease
729 free H2O) on ice, then incubated for 45 min at 37 °C. The tagmented DNA was then purified using
730 the Qiagen MinElute kit eluted with 10.5 µL Elution Buffer (EB). Ten microliters of purified
731 tagmented DNA was PCR amplified using Nextera primers for 12 cycles to generate each library.
732 PCR reaction was subsequently cleaned up using 1.5x AMPureXP beads (Agencourt), and
733 concentrations were measured by Qubit. Libraries were paired-end sequenced on the Illumina
734 HiSeq 4000 platform (100 bp read length).
735
736 Hi-C library preparation
737 Hi-C library preparation on FACS-sorted CD4+ T cells was performed using the Arima-HiC kit
738 (Arima Genomics Inc), according to the manufacturer's protocols. Briefly, cells were crosslinked
739 using formaldehyde. Crosslinked cells were then subject to the Arima-HiC protocol, which utilizes
740 multiple restriction enzymes to digest chromatin. Arima-HiC sequencing libraries were prepared
741 by first shearing purified proximally-ligated DNA and then size-selecting 200-600 bp DNA
742 fragments using AmpureXP beads (Beckman Coulter). The size-selected fragments were then
743 enriched using Enrichment Beads (provided in the Arima-HiC kit), and then converted into
744 Illumina-compatible sequencing libraries with the Swift Accel-NGS 2SPlus DNA Library Kit (Swift,
745 21024) and Swift 2S Indexing Kit (Swift, 26148). The purified, PCR-amplified DNA underwent

746 standard QC (qPCR, Bioanalyzer, and KAPA Library Quantification [Roche, KK4824]) and was
747 sequenced with unique single indexes on the Illumina NovaSeq 6000 Sequencing System using
748 200 bp reads.
749
750 ATAC-seq data analysis
751 ATAC-seq peaks from libraries unstimulated and stimulated Naïve CD4+ T cells were called using
752 the ENCODE ATAC-seq pipeline (https://www.encodeproject.org/atac-seq/). Briefly, pair-end
753 reads from three biological replicates for each cell type were aligned to hg19 genome using
754 bowtie2[91], and duplicate reads were removed from the alignment. Narrow peaks were called
755 independently for each replicate using macs2[92] (-p 0.01 --nomodel --shift -75 --extsize 150 -B --
756 SPMR --keep-dup all --call-summits) and ENCODE blacklist regions (ENCSR636HFF) were
757 removed from peaks in individual replicates. Reproducible peaks, peaks called in at least two
758 replicates, were used to generate a consensus peakset. Signal peaks were normalized using
759 csaw[93] in 10kb bins background regions and low abundance peaks (CPM>1) were excluded from
760 the analysis. Tests for differential accessibility was conducted with the glmQLFit approach
761 implemented in edgeR[94] using the normalization factors calculated by csaw. OCRs with FDR <
762 0.05 and abs($\log_2$FC > 1) between stages were considered differentially accessible.
763
764 Hi-C data analysis
765 Paired-end reads from two replicates were pre-processed using the HICUP pipeline v0.7.4[95], with
766 bowtie as aligner and hg19 as the reference genome. The alignment .bam file were parsed to
767 .pairs format using pairtools v0.3.0 (https://github.com/open2c/pairtools) and pairix v0.3.7
768 (https://github.com/4dn-dcic/pairix), and eventually converted to pre-binned Hi-C matrix in .cool
769 format by cooler v0.8.10[96] with multiple resolutions (500bp, 1kbp, 2kbp, 2.5kbp, 4kbp, 5kbp,
770 10kbp, 25kbp, 40kbp, 50kbp, 100kbp, 250kbp, 500kbp, 1Mbp and 2.5Mbp) and normalized with
771 ICE method[97]. Replicate similarity was determined by HiCRep v1.12.2[96] at 10K resolution. For

772 each sample, eigenvectors were determined from an ICE balanced Hi-C matrix with 40kb
773 resolution using cooltools v0.3.2[98] and first principal components were used to determine A/B
774 compartments with GC% of genome region as reference track to determine the sign. Differential
775 TAD comparison was performed using TADcompare with the default settings for each
776 chromosome (v1.4.0)[99]. Finally, for each cell type, significant intra-chromosomal interaction loops
777 were determined under multiple resolutions (1kb, 2kb and 4kb) using the Hi-C loop caller Fit-Hi-
778 C2 v2.0.7[100] (FDR<1e-6) on merged replicates matrix. The consensus chromatin loops within
779 resolution were identified by combining all three stages. These sets of loops were used as
780 consensus for quantitative differential analysis explained below. The final consensus interaction
781 loops for visualization were collected by merging loops from all the resolutions with preference to
782 keep the highest resolution. Quantitative loop differential analysis across cell types was performed
783 on fast lasso normalized interaction frequency (IF) implemented in multiCompareHiC v1.8.0[101] for
784 each chromosome at resolution 1kb, 2kb and 4kb independently. The contacts with zero
785 interaction frequency (IF) among more than 80% of the samples and average IF less than 5 were
786 excluded from differential analysis. The QLF test based on a generalized linear model was
787 performed in cell type-pairwise comparisons, and p-values were corrected with FDR. The final
788 differential loops were identified by overlapping differential IF contacts with consensus interaction
789 loops.
790
791 Bulk RNA-seq data analysis
792 Bulk RNA-seq libraries were sequenced on an Illumina Novaseq 6000 instrument. The pair-end
793 fastq files were mapped to the genome assembly hg19 by STAR (v2.6.0c) independently for each
794 replicate. The GencodeV19 annotation was used for gene feature annotation and the raw read
795 count for gene feature was calculated by htseq-count (v0.6.1)[102] with parameter settings -f bam -
796 r pos -s yes -t exon -m union. The gene features localized on chrM or annotated as rRNAs, small
797 coding RNA, or pseudo genes were removed from the final sample-by-gene read count matrix.

798 Gene set variation analysis was performed using GSVA 1.42.0[103] with the MSigDB hallmark
799 geneset[104], with resulting scores analyzed using limma (limma_3.50.3)[105]. Low abundance peaks
800 (CPM>1) were excluded from the analysis. Testing for differential expression was conducted with
801 the glmQLFit approach implemented in edgeR[94]. Genes with FDR<0.05 and abs(log_2FC>1)
802 between stages were considered differentially expressed. Differential genes were then clustered
803 using k-means clustering. The number of clusters was determined using the elbow method on
804 the weighted sum of squares, where was set to k=5. Score for how similar each gene followed
805 the clusters expression pattern was determined by calculating pearson correlation coefficients
806 between each gene in the cluster and the cluster centroid.
807
808 Transcription factor footprinting and motif analysis
809 Transcription factor footprints were called using Regulatory Analysis Toolbox HINT-ATAC
810 (v0.13.0) with pooled ATAC-seq data for each stage and consensus peak calls[106]. The rgt-hint
811 footprinting was run with parameters –atac-seq, --paired-end, and organism=hg19 set. The output
812 footprint coordinates were subsequently matched using rgt-motifanalysis matching with
813 parameters --organism hg19 and –pseudocount 0.8 set. The JASPAR2020 position weight matrix
814 database was used to match footprints[107]. Differential analysis of TF binding across conditions
815 was performed using rgt-hint differential with parameters –organism hg19, --bc, --nc 24 using the
816 motif matched transcription factor footprints. An activity score is then calculated based on the
817 accessibility surrounding the footprint.
818
819 Partitioned heritability LD score regression enrichment analysis
820
821 Partitioned heritability LD score regression[108] (v1.0.0) was used to identify heritability enrichment
822 with GWAS summary statistics and open chromatin regions annotated to genes. The baseline
823 analysis was performed using LDSCORE data
824 (https://data.broadinstitute.org/alkesgroup/LDSCORE) with LD scores, regression weights, and

825 allele frequencies from the 1000G phase 1 data. The summary statistics were obtained from
826 studies as described in Table S14 and harmonized with the munge_sumstats.py script.
827 Annotations for Partitioned LD score regression were generated using the coordinates of open
828 chromatin regions that contact gene promoters through Hi-C loops for each cell type. Finally,
829 partitioned LD scores were compared to baseline LD scores to measure enrichment fold change
830 and enrichment p-values, which were adjusted with FDR across all comparisons.
831
832 Variant-to-gene mapping using HiC-derived promoter contacts
833 95% credible sets were determined as previously described[109]. Briefly, P values from GWAS
834 summary statistics were converted to Bayes factors. Posterior probabilities were then calculated
835 for each variant. The variants were ordered from highest to lowest posterior probabilities added
836 to the credible set until the cumulative sum of posterior probabilities reached 95%. Variants in the
837 95% credible set were then intersected with the CD4+ T cell promoter interacting region OCRs
838 from the three timepoints using the R package GenomicRanges (v1.46.1)[110].
839
840 Genomic reference and visualizations
841 All analyses were performed using the hg19 reference genome using gencodeV19 as the gene
842 reference. Genomic tracks were visualized with pyGenomeTracks v3.5[111]. HiC matrices depict
843 the log1p(balanced count) from the cooler count matrix. ATAC-seq tracks were generated from
844 bigwig files that were normalized using deeptools[112].
845
846 Precision-recall analyses of HIEI genes
847 To benchmark our autoimmune effector gene predictions against prior eQTL and chromatin
848 capture studies, we curated a list of 449 expert curated genes from Human Errors in Inborn
849 Immunity 2022, where mutations have been reported to cause immune phenotypes in humans[63].
850 Most of these genes were identified through rare loss of function mutations in the coding region.

851 In addition to this we also took a subset of 145 genes reported to specifically result in autoimmune
852 phenotypes. These two sets of genes were treated as “gold standard” sets of genes to compare
853 different approaches to predict GWAS effector genes. We curated the results of several other
854 chromatin conformation or eQTL based methods of assigning variants to genes to compare with
855 our study. For the chromatin confirmation studies, we obtained loop calls from the associated
856 publications and intersected with the list variants in the 95% credible set from 15 GWAS used to
857 predict effector genes using GenomicRanges (v1.46.1)[110]. For eQTL studies, we obtained the
858 summary stats files and identified eGenes linked to any member of the credible set reported as
859 an eQTL in the transcriptome wide association analysis (TWAS; p-value < 1e-4). Precision was
860 calculated as the ratio of the number predicted genes in the truth set to the total number of
861 predicted genes, while recall was calculated as the ratio of the number predicted genes in the
862 truth set to the total number of genes in truth set.
863
864 Colocalized eQTL comparisons
865 In addition to comparisons with the eGenes identified by TWAS in the Soskic et al. study, we also
866 compared overlap of gene nominations in our study with the eQTLs that colocalized with an
867 autoimmune GWAS[13]. Enrichment of sentinel-gene assignments was conducted similarly as
868 described previously[56]. Briefly, a null distribution was constructed by randomly selecting genes
869 within 1 Mb of the sentinel compared to the set of colocalized cis-eQTL[13] found in through 10,000
870 iterations. The observed overlap reports the set of gene identified by both our HiC based approach
871 with the set of colocalized eQTLs. We report the empirical P value of the observed value relative
872 to null distribution.
873
874 International mouse phenotyping consortium comparisons
875 The set of HiC implicated genes were compared to the mouse international phenotyping
876 consortium set of genes with reported phenotypes[113]. We converted the list of V2G implicated

877 genes to mouse homologs using homologene. We tested for enrichment for each phenotype using
878 a one-sided proportion test implicated in R prop.test with type set to “upper”.
879
880 Identification of pharmacological agents
881 We queried the Drug-Gene Interaction Database with the set of V2G implicated genes for
882 chemicals using rDGIdb (v1.20.0)[114]. To identify the number of papers for each gene with at least
883 one drug annotated to target it, we queried pubmed titles and abstracts using the R package
884 RISmed (v2.3.0) with each gene’s name and either “autoimmune” and the list of autoimmune
885 diseases (**Table S17**). A score to approximate expression specificity was computed using the
886 sum GTEX median expression values (v8) for whole blood or spleen divided by other tissues[115].
887
888 Lentiviral-based CRISPR/CAS9 targeting in Jurkat cells
889 LentiCRISPRv2-mCherry vectors encoding gRNA-CAS9 and the fluorescent reporter mCherry
890 were used for Jurkat targeting. CRISPR guide RNAs (sgRNA) targeting human IL-2-21 intergenic
891 -46, -51, -80, -83, -85, -122, -128 CNS regions were designed using http://crispr.tefor.net and
892 cloned into lentiCRISPRv2-mCherry. Empty vector without gRNA insert was used as a control.
893 Below is the list of CRISPR gRNA for causing deletion of human IL2-21 intergenic regions:
894 CNS region CRISPR gRNA
895 -46 CNS gRNA #1 AGGATGCCTACCTCCAAATG
896 gRNA # 2 AGGTGACAACATTTAGTCAG
897 -51 CNS gRNA #1 GGCAACGAAATTCACTGTGA
898 gRNA # 2 ATTCTAACAGGAATCATTCG
899 -80 CNS gRNA # 1 GTTCTACCTATGCCGCATTG
900 gRNA # 2 GAGATTTACTCAGTCCAATG
901 -83 CNS gRNA # 1 GTGACAAGCATGACTCTACA
902 gRNA # 2 GTGATGGTGAATTAAGCTGA

903 -85 CNS gRNA #1 AGGGTTTTCTAGTTACGAGA

904 gRNA # 2 ATGGTTAGTTAGCTCCCAAG

905 -96 CNS gRNA#1 TGGGAAAAACATCTTACCTG

906 gRNA # 2 TGGCCCATGAACCATCAAAG

907 -122 CNS gRNA# 1 GTTATTAATCTAAGCGGAGA

908 gRNA# 2 GGAAGTTAGGCAGTCAATCG

909 -128 CNS gRNA#1 CTTCAATCATTGCATTCCAC

910 gRNA# 2 TGACACCACCCCTGCTTGAG

911 HEK 293T cells were grown in RPMI1640 complete medium (RPMI1640 + 1X P/S, 1x L-Glu, 10%
912 FBS), 37C, 7%CO2. 293T cells were transfected with 10 ug of lenti-CRISPR-V2-CRE construct
913 along with packaging plasmid 6 ug of PsPAX2 (Addgene, Cat #12260) and 3.5 ug of PmD2.G
914 (Addgene, Cat #12259) using Lipofectamine 2000 transfection reagent (Invitrogen cat #
915 11668019). After 6 hours, the transfection medium was replaced with complete culture medium.
916 Transfected cells were incubated at 37C for 48-72 hours in a cell culture incubator. and then the
917 Lentiviral supernatants were harvested and spun at 300g for 5 minutes to remove cellular debris.
918 Lentiviral supernatants were concentrated using Lenti-X^{TM} Concentrator (Takara Bio, Cat #
919 631232) and then centrifuged at 1500g for 30 minutes at 4°C and supernatant was discarded.
920 The lentiviral pellet was resuspended at a ratio of 1:20 of the original volume using RPMI media
921 and concentrated virus supernatant aliquots were prepared and stored until use at -80°C. To
922 achieve high transduction efficiency, the viral supernatant was titrated in Jurkat cells through
923 transduction using various dilutions of the viral supernatants and transduction efficiency was
924 determined by mcherry expression analyzed through flow cytometry. Jurkat cells were seeded in
925 a 24 well plate at 0.5 x106/well in culture media, viral supernatant with 8 ug/mL of polybrene was
926 added to each well. Spinfection was performed for 90 min. at 2500 rpm, and transduced cells
927 were equilibrated at 37C for ~6 hrs, followed by incubation at 37C 5%CO2 for ~72 hours culture.
928 Transduced Jurkat Cells were then harvested and stimulated using PMA (15 ng/mL) + Ionomycin

929 (1 uM) + human anti-CD28 (2 ug/ml), BioXcell cat # BE0248 in 96 well culture plates (TPP, cat #
930 92097) in triplicates. Cell culture supernatants were collected at the end of culture and analyzed
931 for IL-2 by ELISA using a kit (cat # 88-7025-76) purchased from Thermofisher Scientific.
932
933 gRNA-CAS9 RNP-based targeting in primary human CD4+ T cells
934 Primary Human CD4 T cells derived from 5 normal healthy donors were obtained from Human
935 Immunology core (University of Pennsylvania). Alt-R S.p. HiFi Cas9 Nuclease V3 cat # 1081061
936 CAS9 and following list of Alt-R® CRISPR-Cas9 sgRNA targeting IL-2-21 CRE were purchased
937 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

938

939 Primary Human CD4 T (5-10e6) were incubated with sgRNA and CAS9 protein complex and
940 electroporation was done using P3 Primary Cell 4D-Nucleofector™ X Kit L (Lonza, cat # V4XP-
941 3024) and Lonza 4D nucleofector system. B2M gene was CRISPR targeted as a positive control.
942 As per manufacturer's protocol cells were electroporated pulse code Fl-115 in 100ul cuvette
943 format. After nucleofection, cells were allowed to rest in the complete media for ~2 days. Cells
944 were then harvested, washed with PBS and aliquots of cells were used for further experimentation
945 such as flow staining and cell activation. Primary human CD4 T Cells (0.1e6/well) were seeded
946 in triplicates (for each experimental condition) in 96 well plate format in RPMI complete medium
947 and stimulated using ImmunoCult™ Human CD3/CD28 T Cell Activator (STEM cells
948 Technologies, cat # 10971). Cell culture supernatants were collected at 4h of stimulation and
949 stored at -80C until assayed. IL-2 ELISA was performed using Thermofisher Scientific IL-2 Human
950 Uncoated ELISA Kit with Plates, cat # 88-7025-76.

951

952 Recombinant reporter assays

953 The *IL2* URR was cloned into Xho I and Hind III sites of PGL4 Luc vector (Promega) and then
954 ~500 bp individual sequence representing *IL2-IL21* intergenic CNS at -46, -51, -80, -83, -85, -128
955 were cloned upstream to IL-2 URR at the Xho I site of the pGL4 vector. Primary human CD4 T
956 cells obtained from 5 normal healthy donors were transfected with PGL4-cRE-URR constructs.
957 Briefly, primary human CD4 T cells were activated with anti-CD3+ anti-CD28 Dynabeads (1:1)
958 (Thermofisher scientific, cat # 11161D) + IL-2 overnight and then 1 million aliquots of cells
959 (triplicates) were electroporated using Nucleofector 2b, human T cell Nuclefactor kit (Lonza VPA
960 # 1002, program # T-020) with 2ug of PGL4 firefly vector constructs along with 0.2 ug of PGL4
961 Renilla vector; cells were allowed to rest ON in RPMI + IL-2 and then re-stimulated with plate-
962 bound anti-CD3+ anti-CD28 (2 ug/ml each) for 5 hrs. Dual luciferase assay was performed with
963 cell lysate prepared using Promega dual luciferase assay kit. Cell lysate was prepared in PLB as
964 per manufacturer's protocols and then firefly and renilla luciferase activities were analyzed by
965 spectramax ID5 (Molecular Devices). Firefly luciferase activity was normalized against the internal
966 control renilla luciferase activity.

967

968 Pharmacologic validation of novel T cell activation regulatory genes

969 The drugs lometrexol (GART antagonist), 25-hydroxycholesterol (GPR183 ligand), epinephrine &
970 norepinephrine (CHRNB agonists), and arachidonoyl glycine (GPR18 agonist) were purchased
971 from Sigma. The GPR18 antagonist O-1918 was purchased from ChemCruz. CEP1347
972 (MAP3K11 antagonist), kisspeptin 234 (KISSIR ligand) and the PTK6 antagonist tilfrinib were
973 purchased from Tocris. The GRK6 antagonist GRK-IN-1 was purchased from DC Chemicals. The
974 SIK1 antagonist HG-91-9-1 was purchased from Selleckchem. The OXER1 agonist 5-Oxo-ete
975 was purchased from Cayman chemicals. The drugs were dissolved in DMSO or ethanol as
976 suggested by the vendor and working stocking concentrations were prepared in RPMI1640
977 medium or PBS. The drug effect on T cell proliferation was assayed using murine lymphocytes

978 cultured under TCR and CD28 activation conditions. Spleen and lymph nodes were collected from
979 6-8 weeks old female C57BL/7 mice and single cell lymphocyte cell suspensions were prepared
980 in RPMI 1640 complete medium. 20 million lymphocytes were labeled with CFSE and re-
981 suspended in RPMI 1640 medium. Cells (0.5 million labeled cells/well) were loaded into 48 well
982 culture plate and activated with mouse anti-CD3 and anti-CD28 agonistic antibodies (1μg mL
983 each). Drugs at the indicated concentrations were added in the culture medium and both
984 untreated and drug treated cell cultures were incubated at 37°C for 72 hours in a cell culture
985 incubator. Cells were harvested after 3 days of culture, washed with PBS and then stained with
986 live-dead aqua dye. After washing with FACS buffer (PBS containing 2% FBS), cells were stained
987 with fluorochrome conjugated antibodies CD4-APC, CD8-PB, CD44-Percp Cy5.5 and CD25-
988 BV650. Stained cells were analyzed on a Beckman Coulter Cytoflex S flow cytometer. The
989 division profile of CD4+ CFSE+ T cells were gated on live populations. The flow data was
990 analyzed using Flowjo10 software and the number of divided CD4+ T cells were determined as
991 described previously[116].
992
993

994 **FIGURE LEGENDS**

995 **Figure 1: Defined gene expression dynamics throughout activation of naïve CD4+ T cells.**
996 (**A**) Diagram of study design: RNA-seq, ATAC-seq, and HiC libraries were prepared from sorted
997 CD4+CD62L^{hi}CD44^{lo} naïve T cells (donor N=3) prior to or following 8hr or 24hr stimulation with
998 anti-CD3/28 beads. cREs are classified to genes allowing for interrogation of the pattern of
999 expression, accessibility, and chromatin structure changes of autoimmune-associated variants.
1000 (**B**) Heatmap showing normalized expression of known markers of T cell activation. (**C**) K-means
1001 clustering of differentially expressed genes into five groups using the elbow and within-cluster-
1002 sum of squares methods to select the number of clusters. The centroid of the cluster is depicted
1003 as a black line, and members of the cluster are depicted as colored lines. Color indicates the
1004 Pearson's correlation coefficient between the gene and the cluster centroid, with red indicating
1005 higher correlation, and blue lower correlation.

1006

1007 **Figure 2: Chromatin accessibility and architecture changes across three stages of T cell**
1008 **activation.** (**A**) Volcano plots depicting the $\log_2$ change in accessibility in all reproducible OCRs
1009 (present in 2 of the 3 replicates) compared to the $-\log_{10}$(FDR), significant points are indicated in
1010 red (FDR < 0.05; abs($\log_2$FC) > 1). (**B**) Genomic regions were classified by membership in A vs.
1011 B compartments at each time point. The pie chart depicts regions with differential A/B
1012 compartment assignment in response to activation. (**C**) TAD structure was determined at each
1013 time point, and regions exhibiting a shift, split/merge, change in strength, or more complex
1014 rearrangements are depicted in the pie chart. (**D**) Loop calls identified from HiC data for each
1015 timepoint called at three resolutions (1kb, 2kb, 4kb bins). (**E**) Differential loop-calls called across
1016 all resolutions. (**F**) Density plots of cRE accessibility and gene expression separated by whether
1017 contact frequency increased, decreased, or remained stable from the transition from unstimulated
1018 to 8 hour activation and 8 hour to 24 hour activation.

1019

1020 **Figure 3: Transcription factors prediction potential regulators of chromatin status and**
1021 **expression changes**. Footprints were annotated to each TF motif by sequence matching to the
1022 JASPAR motif database, and the average accessibility of the region surrounding each motif was
1023 used as a measure of activity at each time point. Scatterplots depict the change in accessibility
1024 for each TF motif (activity score) and $\log_2$FC of TF gene expression between unstimulated and 8
1025 hour activation (**A**) or 8 hour and 24 hour activation (**B**). TF were classified as differentially
1026 expressed (orange), differentially active (brown), both (red), or neither (grey). Dot size indicates
1027 the number of predicted footprints occupied by each motif. (**C**) Average accessibility (normalized
1028 by depth and motif number) in a 200bp window surrounding motif footprints for FOSL2::JUNB,
1029 Smad2::Smad3, MAZ, and KLF2 from three timepoints (unstim: green; 8hr: blue; 24hr: red). (**D**)
1030 Z-score of TF motif enrichment for cRE connected to the five clusters compared to all OCR.
1031 Lighter color indicates higher specificity of enrichment for that TF cluster. (**E**) Connections
1032 between different TFs based on physical interactions and predicted regulation-based co-
1033 expression determined by GENIE3 for the three timepoints. Node color indicates expression
1034 (TPM; darker = higher expression, ligher = lower expression), edge color reflects confidence in
1035 the interaction called by GENIE3 (darker higher confidence).
1036
1037 **Figure 4: Variant-to-gene mapping of autoimmune-associated loci implicates genetic**
1038 **variants in the control of T cell activation**. (**A**) Identification of genes in contact with open
1039 chromatin regions harboring autoimmune-associated SNPs. Sentinel SNPs from 15 autoimmune
1040 GWAS (**Table S11**) were used to identify the 95% credible set of proxy SNPs for each lead SNP.
1041 SNP locations were integrated with ATAC-seq and HiC data to identify the 95% credible set of
1042 accessible SNPs in physical contact with a gene promoter. Genes are further refined based on
1043 expression dynamics over the time course of T cell activation. (**B**) Partitioned LD score regression
1044 for autoimmune GWAS studies using the OCR in contact with the genes in the five clusters
1045 defined by RNA-seq. Circle size = enrichment, color = significance, **FDR<0.01, ***FDR<0.001.

1046 (**C**) Log distribution of the 1D distance between each proxy SNP and its interacting gene based
1047 on 3D chromatin conformation (median=103 kbp). (**D**) Distribution of the number of genes
1048 contacted by each accessible variant (median=5). (**E**) Fraction of open disease-associated
1049 variants that interact with no gene promoters (grey), only with the nearest gene promoter (purple),
1050 with the nearest gene and a distant gene(s) (orange), and with only a distant gene(s) (blue). (**F**)
1051 Number of genes identified by variant-to-gene mapping at each time point. Black = shared in all
1052 stages, blue = shared in two stages, red = specific to one stage. (**G**) Set of implicated proxy-genes
1053 pairs that are both differentially expressed, display differential accessibility, and chromatin contact
1054 across T cell activation. (**H**) Example MS variant rs1077667 (PP=1.0) exhibits increased
1055 accessibility and contact with the promoters of the *GPR108* and *TRIP10* at 8 hours post activation,
1056 which are increased in expression at this time point. (**I**) Example allergy variant rs7380290
1057 interacts with the *SIK1* gene that is upregulated after activation.
1058
1059 **Figure 5: Functional validation of autoimmune V2G-implicated cRE at the *IL2* locus.** (**A**) The
1060 combination of evolutionary conservation (blue), open chromatin (red), and autoimmune disease-
1061 associated SNPs at the *IL2* locus identify putative cRE in quiescent vs. 8-hour activated naïve
1062 CD4+ T cells. (**B**) Activation-dependent TAD/sub-TAD structure (heatmaps), chromatin
1063 remodeling (grey) and promoter-OCR interactions (red) at the *IL2* locus. (**C**) Recombinant reporter
1064 assay showing transcriptional activity of the *IL2* URR (+35 to -500 from the TSS) in activated
1065 primary CD4+ T cells (N= 7 donors) compared to a basal promoter (pGL4, left panel). The right
1066 panel depicts transcriptional activity of the CNS regions indicated in **A** cloned upstream of the
1067 URR. All regions except the -80 CNS show statistically significant activity relative to the URR
1068 alone (N=7, P<0.05, line = median, box = 95/5% range). (**D**) Scheme of CRISPR/CAS9-based
1069 deletion of individual *IL2* CNS using flanking gRNAs. (**E**) Activation-induced secretion of IL-2
1070 protein by CRISPR-targeted primary CD4+ T cells (N=5 donors, left panel) or Jurkat cells (N=3
1071 replicates, right panel) relative to untargeted control (CAS9, no gRNA) cells (****P<0.0001,

1072 ***P<0.001, **P<0.01, *P<0.05, line = median, box = 95/5% range). In primary cells, *B2M* gRNAs
1073 served as an irrelevant targeted control. (**F**) Scheme of CRISPR/CAS9-based deletion of the 81.3
1074 kb region containing all distal *IL2* cRE using flanking gRNAs in Jurkat cells. (**G**) Activation-induced
1075 IL-2 protein (left panel) and mRNA (right panel) by control (black) vs. 81.3 kb deleted (red) Jurkat
1076 cells (N=3 separate clones).
1077
1078 **Figure 6: Distal *IL2* enhancers are evolutionarily conserved and impact *in vivo* T cell-**
1079 **mediated immunity in mice.** (**A**) Map of the human and mouse *IL2* locus depicting mammalian
1080 conservation (dark blue), ATAC-seq from activated human and mouse CD4+ T cells (red), and
1081 orthologous non-coding sequences conserved between mouse and human (light blue). (**B**) IL-2
1082 protein secretion (top left panel) and *Il2* mRNA expression (top right panel) by CD4+ T cells from
1083 wild-type or *Il2*-83 cRE ko mice in response to stimulation with plate-bound anti-CD3 and anti-
1084 CD28 Ab *in vitro*. Bottom panels show soluble anti-CD3-induced *in vitro* clonal expansion of wild-
1085 type or *Il2*-83 cRE ko CD4+ (left) and CD8+ (right) T cells measured by dye dilution[116]. (**C**) Foxp3
1086 expression by murine CD4+CD25- Tconv stimulated with anti-CD3 Ab and TGF-beta in the
1087 absence (top panels) vs. presence (bottom panels) of exogenous IL-2. Wild-type or *Il2*-83 cRE ko
1088 mice (N=6) were immunized intraperitoneally with chicken ovalbumin in incomplete Freund's
1089 adjuvant. The frequency of CD4+PD1+CXCR5^{hi} follicular helper T cells in the spleens of 3 animals
1090 was measured by flow cytometry at day 5 post-immunization (**D**), and levels of ovalbumin-specific
1091 IgG in the serum of 3 animals were measured at day 10 post-immunization (**E**). (**F**) In vivo IBD
1092 model. CD4+CD25- Tconv from wild-type (pink, red) or *Il2*-83 cRE ko (purple, blue) mice were
1093 transferred alone (purple, pink) or together with wild-type CD4+CD25+ Treg (red, blue) into
1094 RAG1-ko mice (N=5 per group). Animal weight was monitored for 40 days, and the number (in
1095 millions) of activated CD4+ Tconv in the spleens and mesenteric lymph nodes was measured by
1096 flow cytometry (inset). Statistical significance (*p<0.05) was measured by ANOVA or t-test.
1097

1098 **Figure 7: Functional validation of autoimmune V2G-implicated genes.** (**A**) Recombinant
1099 reporter assay in primary activated CD4+ T cells (N= 7 donors) showing transcriptional activity of
1100 the reference vs. risk alleles of the *IL2* -128 cRE relative to the URR alone (P<0.0001). (**B**)
1101 Prominent TF motifs predicted to be disrupted (blue) or stabilized (red) by promoter-connected
1102 autoimmune SNPs. (**C**) 3D chromatin-based V2G genes are enriched for CRISPR-implicated
1103 genes that regulate CD4+ T cell activation. Observed enrichment of genes regulating multiple
1104 aspects of CD4+ T cell activation (IL-2, IL-2 receptor, CTLA-4, IFNg, TNFa or proliferation) from
1105 CRISPR screens by Freimer, Schmidt, and Shifrut among sets of 3D chromatin-implicated genes
1106 among individual diseases (green, FDR<0.05) or all diseases (purple, FDR<0.05). (**D**) Enrichment
1107 for 3D chromatin-based autoimmune V2G genes among genes with germline knock-out mouse
1108 immune (red) and other (black) phenotypes (adjP<0.05, IMPC database). (**E**) Autoimmune V2G-
1109 implicated genes with at least one pharmacologic modulator (rDGIdb). (**F**) Comparison of the
1110 number of manuscripts retrieved from PubMed related to autoimmune disease for each V2G gene
1111 with pharmaceutical agents available (x-axis) with an immune-specific expression score
1112 computed using the sum GTEX median expression values (v8) for whole blood or spleen divided
1113 by other tissues (y-axis). Genes highlighted in red were selected for functional validation in **G**. (**G**)
1114 Dose-dependent impact of the indicated pharmacologic agents targeting the V2G-implicated
1115 genes *KISS1R, CHRNB, OXER1, GPR18, GRK6, PTK6, MAP3K11, GPR183, GART,* and *SIK1*
1116 on proliferation of anti-CD3/28 activated murine CD4+ T cells *in vitro* (N=4).
1117
1118

1119 **Supplemental Tables**

1120 Table S1: Expression and clustering of differentially expressed genes.

1121 Table S2: Pathway enrichment of differentially expressed RNA-seq genes

1122 Table S3: Pathway enrichment of differentially expressed RNA-seq genes by cluster

1123 Table S4: Accessible chromatin regions

1124 Table S5: Differential accessible open chromatin regions

1125 Table S6: A/B compartment calls

1126 Table S7: Differential TAD boundaries

1127 Table S8: Stripe calls

1128 Table S9: Differential contact frequency

1129 Table S10: Summary of TF footprinting

1130 Table S11: TF target gene pathway enrichment

1131 Table S12: List of all GWAS studies included

1132 Table S13: Variant to gene mapping across all timepoints

1133 Table S14: Partitioned LD score regression output

1134 Table S15: Autoimmune variants in open gene promoters

1135 Table S16: Comparison of V2G approaches

1136 Table S17: Motifs predicted to disrupt transcription factor binding sites.

1137 Table S18: V2G genes implicated by orthogonal data

1138 Table S19: V2G target gene drug repurposing results

1139

1140

1141

1142 **REFERENCES**

1143

1144 1. Zhang, F. & Lupski, J. R. Non-coding genetic variants in human disease. *Hum Mol Genet* **24**,
1145 R102-10 (2015).
1146 2. Weinreich, M. A. *et al.* KLF2 Transcription-Factor Deficiency in T Cells Results in Unrestrained
1147 Cytokine Production and Upregulation of Bystander Chemokine Receptors. *Immunity* **31**,
1148 122–130 (2009).
1149 3. Tzachanis, D. *et al.* Tob is a negative regulator of activation that is expressed in anergic and
1150 quiescent T cells. *Nat Immunol* **2**, 1174–82 (2001).
1151 4. Yusuf, I. & Fruman, D. A. Regulation of quiescence in lymphocytes. *Trends Immunol* **24**, 380–
1152 386 (2003).
1153 5. Zhu, J., Yamane, H. & Paul, W. E. Differentiation of Effector CD4 T Cell Populations. *Annu Rev*
1154 *Immunol* **28**, 445–489 (2010).
1155 6. Saravia, J., Chapman, N. M. & Chi, H. Helper T cell differentiation. *Cell Mol Immunol* **16**, 634–
1156 643 (2019).
1157 7. Farh, K. K.-H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease
1158 variants. *Nature* **518**, 337 343 (2014).
1159 8. Chen, L. *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune
1160 Cells. *Cell* **167**, 1398 1414.e24 (2016).
1161 9. Soskic, B. *et al.* Chromatin activity at GWAS loci identifies T cell states driving complex immune
1162 diseases. *Nat Genet* 1–8 (2019) doi:10.1038/s41588-019-0493-9.
1163 10. Robertson, C. C. *et al.* Fine-mapping, trans-ancestral and genomic analyses identify causal
1164 variants, cells, genes and drug targets for type 1 diabetes. *Nat Genet* **53**, 962–971 (2021).
1165 11. Lu, X. *et al.* Global discovery of lupus genetic risk variant allelic enhancer activity. *Nat*
1166 *Commun* **12**, 1611 (2021).
1167 12. Chandra, V. *et al.* Promoter-interacting expression quantitative trait loci are enriched for
1168 functional genetic variants. *Nat Genet* **53**, 110–119 (2021).
1169 13. Soskic, B. *et al.* Immune disease risk variants regulate gene expression dynamics during
1170 CD4+ T cell activation. *Nat Genet* 1–10 (2022) doi:10.1038/s41588-022-01066-3.
1171 14. Mouri, K. *et al.* Prioritization of autoimmune disease-associated genetic variants that perturb
1172 regulatory element activity in T cells. *Nat Genet* 1–10 (2022) doi:10.1038/s41588-022-
1173 01056-5.
1174 15. Rioux, J. D. & Abbas, A. K. Paths to understanding the genetic basis of autoimmune disease.
1175 *Nature* **435**, 584–589 (2005).
1176 16. Buckner, J. H. Mechanisms of impaired regulation by CD4(+)CD25(+)FOXP3(+) regulatory T
1177 cells in human autoimmune diseases. *Nat Rev Immunol* **10**, 849 859 (2010).
1178 17. Javierre, B. M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-coding
1179 Disease Variants to Target Gene Promoters. *Cell* **167**, 1369 1384.e19 (2016).
1180 18. Burren, O. S. *et al.* Chromosome contacts in activated T cells identify autoimmune disease
1181 candidate genes. *Genome Biol* **18**, 165 (2017).
1182 19. Yang, J. *et al.* Analysis of chromatin organization and gene expression in T cells identifies
1183 functional genes for rheumatoid arthritis. *Nat Commun* **11**, 4402 (2020).
1184 20. Bykova, M., Hou, Y., Eng, C. & Cheng, F. Quantitative trait locus (xQTL) approaches identify
1185 risk genes and drug targets from human non-coding genomes. *Hum Mol Genet* **31**, R105–
1186 R113 (2022).
1187 21. Shifrut, E. *et al.* Genome-wide CRISPR Screens in Primary Human T Cells Reveal Key
1188 Regulators of Immune Function. *Cell* **175**, 1958-1971.e15 (2018).
1189 22. Schmidt, R. *et al.* CRISPR activation and interference screens decode stimulation responses
1190 in primary human T cells. *Science* **375**, eabj4008 (2022).

1191 23. Freimer, J. W. *et al.* Systematic discovery and perturbation of regulatory genes in human T
1192 cells reveals the architecture of immune networks. *Nat Genet* **54**, 1133–1144 (2022).
1193 24. Su, C., Pahl, M. C., Grant, S. F. A. & Wells, A. D. Restriction enzyme selection dictates
1194 detection range sensitivity in chromatin conformation capture-based variant-to-gene
1195 mapping approaches. *Hum Genet* (2021) doi:10.1007/s00439-021-02326-8.
1196 25. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**,
1197 75 (2012).
1198 26. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals
1199 Folding Principles of the Human Genome. *Science* **326**, 289–293 (2009).
1200 27. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of
1201 chromatin interactions. *Nature* **485**, 376 380 (2012).
1202 28. Mirny, L. A., Imakaev, M. & Abdennur, N. Two major mechanisms of chromosome
1203 organization. *Curr Opin Cell Biol* **58**, 142–152 (2019).
1204 29. Yoon, S., Chandra, A. & Vahedi, G. Stripenn detects architectural stripes from chromatin
1205 conformation data using computer vision. *Nat Commun* **13**, 1602 (2022).
1206 30. Kim, T.-K. & Shiekhattar, R. Architectural and Functional Commonalities between Enhancers
1207 and Promoters. *Cell* **162**, 948 959 (2015).
1208 31. Hamamoto, K. & Fukaya, T. Molecular architecture of enhancer–promoter interaction. *Curr*
1209 *Opin Cell Biol* **74**, 62–70 (2022).
1210 32. Consortium, R. E. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**,
1211 317 330 (2015).
1212 33. Yoon, J.-H. *et al.* Phosphorylation status determines the opposing functions of Smad2/Smad3
1213 as STAT3 cofactors in TH17 differentiation. *Nat Commun* **6**, 7600 (2015).
1214 34. Safford, M. *et al.* Egr-2 and Egr-3 are negative regulators of T cell activation. *Nat Immunol* **6**,
1215 472–480 (2005).
1216 35. Schubert, L. A., Jeffery, E., Zhang, Y., Ramsdell, F. & Ziegler, S. F. Scurfin (FOXP3) Acts as
1217 a Repressor of Transcription and Regulates T Cell Activation. *J Biol Chem* **276**, 37672–
1218 37679 (2001).
1219 36. Xiao, T., Li, X. & Felsenfeld, G. The Myc-associated zinc finger protein (MAZ) works together
1220 with CTCF to control cohesin positioning and genome organization. *Proc National Acad*
1221 *Sci* **118**, e2023127118 (2021).
1222 37. Sebzda, E., Zou, Z., Lee, J. S., Wang, T. & Kahn, M. L. Transcription factor KLF2 regulates
1223 the migration of naive T cells by restricting chemokine receptor expression patterns. *Nat*
1224 *Immunol* **9**, 292–300 (2008).
1225 38. Moskowitz, D. M. *et al.* Epigenomics of human CD8 T cell differentiation and aging. *Sci*
1226 *Immunol* **2**, eaag0192 (2017).
1227 39. Harris, J. E. *et al.* Early Growth Response Gene-2, a Zinc-Finger Transcription Factor, Is
1228 Required for Full Induction of Clonal Anergy in CD4+ T Cells. *J Immunol* **173**, 7331–7338
1229 (2004).
1230 40. Williams, J. B. *et al.* The EGR2 targets LAG-3 and 4-1BB describe and regulate dysfunctional
1231 antigen-specific CD8+ T cells in the tumor microenvironment. *J Exp Med* **214**, 381–400
1232 (2017).
1233 41. Avitahl, N. *et al.* Ikaros Sets Thresholds for T Cell Activation and Regulates Chromosome
1234 Propagation. *Immunity* **10**, 333–343 (1999).
1235 42. Bandyopadhyay, S. *et al.* Interleukin 2 gene transcription is regulated by Ikaros-induced
1236 changes in histone acetylation in anergic T cells. *Blood* **109**, 2878–86 (2007).
1237 43. Thomas, R. M. *et al.* Ikaros Enforces the Costimulatory Requirement for IL2 Gene Expression
1238 and Is Required for Anergy Induction in CD4+ T Lymphocytes. *J Immunol* **179**, 7305–7315
1239 (2007).
1240 44. Thomas, R. M. *et al.* Ikaros Silences T-bet Expression and Interferon-γ Production during T
1241 Helper 2 Differentiation. *J Biol Chem* **285**, 2545–2553 (2010).

1242 45. Weber, B. N. *et al.* A critical role for TCF-1 in T-lineage specification and differentiation. *Nature*
1243 **476**, 63–68 (2011).
1244 46. Johnson, J. L. *et al.* Lineage-Determining Transcription Factor TCF-1 Initiates the Epigenetic
1245 Identity of T Cells. *Immunity* **48**, 243-257.e10 (2018).
1246 47. Wang, W. *et al.* TCF-1 promotes chromatin interactions across topologically associating
1247 domains in T cell progenitors. *Nat Immunol* 1–11 (2022) doi:10.1038/s41590-022-01232-
1248 z.
1249 48. Furukawa, T. *et al.* Involvement of PLAGL2 in activation of iron deficient- and hypoxia-induced
1250 gene expression in mouse cell lines. *Oncogene* **20**, 4718–4727 (2001).
1251 49. Wu, Q. *et al.* Hypoxia-inducible factors: master regulators of hypoxic tumor immune escape.
1252 *J Hematol Oncol* **15**, 77 (2022).
1253 50. Fu, W. *et al.* A multiply redundant genetic switch "locks in" the transcriptional signature of
1254 regulatory T cells. *Nat Immunol* **13**, 972–980 (2012).
1255 51. Pramanik, J. *et al.* Genome-wide analyses reveal the IRE1a-XBP1 pathway promotes T helper
1256 cell differentiation by resolving secretory stress and accelerating proliferation. *Genome*
1257 *Med* **10**, 76 (2018).
1258 52. Gao, X., Tewari, K., Svaren, J. & Suresh, M. Role of cell cycle regulator E2F1 in regulating
1259 CD8 T cell responses during acute and chronic viral infection. *Virology* **324**, 567–576
1260 (2004).
1261 53. DeRyckere, D. & DeGregori, J. E2F1 and E2F2 Are Differentially Required for Homeostasis-
1262 Driven and Antigen-Induced T Cell Proliferation In Vivo. *J Immunol* **175**, 647–655 (2005).
1263 54. Cook, M. E., Jarjour, N. N., Lin, C.-C. & Edelson, B. T. Transcription Factor Bhlhe40 in
1264 Immunity and Autoimmunity. *Trends Immunol* **41**, 1023–1036 (2020).
1265 55. Chesi, A. *et al.* Genome-scale Capture C promoter interactions implicate effector genes at
1266 GWAS loci for bone mineral density. *Nat Commun* **10**, 1260 (2019).
1267 56. Su, C. *et al.* Mapping effector genes at lupus GWAS loci using promoter Capture-C in follicular
1268 helper T cells. *Nat Commun* **11**, 3294 (2020).
1269 57. Pahl, M. C. *et al.* Cis-regulatory architecture of human ESC-derived hypothalamic neuron
1270 differentiation aids in variant-to-gene mapping of relevant complex traits. *Nat Commun* **12**,
1271 6749 (2021).
1272 58. Pahl, M. C. *et al.* Implicating effector genes at COVID-19 GWAS loci using promoter-focused
1273 Capture-C in disease-relevant immune cell types. *Genome Biol* **23**, 125 (2022).
1274 59. Su, C. *et al.* 3D chromatin maps of the human pancreas reveal lineage-specific regulatory
1275 architecture of T2D risk. *Cell Metab* **34**, 1394-1409.e4 (2022).
1276 60. Feng, Y. *et al.* The Cdc42-interacting Protein-4 (CIP4) Gene Knock-out Mouse Reveals
1277 Delayed and Decreased Endocytosis*. *J Biol Chem* **285**, 4348–4354 (2010).
1278 61. Darling, N. J. & Cohen, P. Nuts and bolts of the salt-inducible kinases (SIKs). *Biochem J* **478**,
1279 1377–1397 (2021).
1280 62. Schmiedel, B. J. *et al.* Single-cell eQTL analysis of activated T cell subsets reveals activation
1281 and cell type–dependent effects of disease-risk variants. *Sci Immunol* **7**, eabm2508
1282 (2022).
1283 63. Tangye, S. G. *et al.* Human Inborn Errors of Immunity: 2022 Update on the Classification from
1284 the International Union of Immunological Societies Expert Committee. *J. Clin. Immunol.*
1285 **42**, 1473–1507 (2022).
1286 64. Cerosaletti, K. *et al.* Multiple autoimmune-associated variants confer decreased IL-2R
1287 signaling in CD4+ CD25(hi) T cells of type 1 diabetic and multiple sclerosis patients. *Plos*
1288 *One* **8**, e83811 (2013).
1289 65. Spolski, R., Li, P. & Leonard, W. J. Biology and regulation of IL-2: from molecular mechanisms
1290 to human therapy. *Nat Rev Immunol* **18**, 648 659 (2018).
1291 66. Abbas, A. K., Trotta, E., Simeonov, D. R., Marson, A. & Bluestone, J. A. Revisiting IL-2:
1292 Biology and therapeutic prospects. *Sci Immunol* **3**, eaat1482 (2018).

1293 67. Joosse, M. E. *et al.* Duplication of the IL2RA locus causes excessive IL-2 signaling and may
1294 predispose to very early onset colitis. *Mucosal Immunol* 1–11 (2021) doi:10.1038/s41385-
1295 021-00423-5.
1296 68. Fujita, T., Shibuya, H., Ohashi, T., Yamanishi, K. & Taniguchi, T. Regulation of human
1297 interleukin-2 gene: Functional DNA sequences in the 5′ flanking region for the gene
1298 expression in activated T lymphocytes. *Cell* **46**, 401–407 (1986).
1299 69. Durand, D. B., Bush, M. R., Morgan, J. G., Weiss, A. & Crabtree, G. R. A 275 basepair
1300 fragment at the 5' end of the interleukin 2 gene enhances expression from a heterologous
1301 promoter in response to signals from the T cell antigen receptor. *J Exp Medicine* **165**, 395
1302 407 (1987).
1303 70. Novak, T. J., White, P. M. & Rothenberg, E. V. Regulatory anatomy of the murine interleukin-
1304 2 gene. *Nucleic Acids Res* **18**, 4523–4533 (1990).
1305 71. Mehra, P. & Wells, A. D. Long-Range Transcriptional Control of the Il2 Gene by an Intergenic
1306 Enhancer. *Mol Cell Biol* **35**, 3880–3891 (2015).
1307 72. Consortium, T. E. P. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**,
1308 636–640 (2004).
1309 73. Lin, X. *et al.* Nested epistasis enhancer networks for robust genome regulation. *Science* **377**,
1310 1077–1085 (2022).
1311 74. Hedlund, G., Dohlsten, M., Ericsson, P. O. & Sjögren, H. O. Rapid response to Con A by
1312 CD4+CD45R− rat memory lymphocytes as compared to CD4+CD45R+ lymphocytes. *Cell*
1313 *Immunol* **119**, 317–326 (1989).
1314 75. Picker, L. *et al.* Direct demonstration of cytokine synthesis heterogeneity among human
1315 memory/effector T cells by flow cytometry. *Blood* **86**, 1408–1419 (1995).
1316 76. Sojka, D. K., Bruniquel, D., Schwartz, R. H. & Singh, N. J. IL-2 Secretion by CD4+ T Cells In
1317 Vivo Is Rapid, Transient, and Influenced by TCR-Specific Competition. *J Immunol* **172**,
1318 6136–6143 (2004).
1319 77. Powrie, F., Correa-Oliveira, R., Mauze, S. & Coffman, R. L. Regulatory interactions between
1320 CD45RBhigh and CD45RBlow CD4+ T cells are important for the balance between
1321 protective and pathogenic cell-mediated immunity. *J Exp Medicine* **179**, 589–600 (1994).
1322 78. Morawski, P. A., Mehra, P., Chen, C., Bhatti, T. & Wells, A. D. Foxp3 Protein Stability Is
1323 Regulated by Cyclin-dependent Kinase 2. *J Biol Chem* **288**, 24494–24502 (2013).
1324 79. Smigiel, K. S. *et al.* CCR7 provides localized access to IL-2 and defines homeostatically
1325 distinct regulatory T cell subsets. *J Exp Medicine* **211**, 121–136 (2014).
1326 80. Transcriptional state and function of CD8+ memory T cells is linked to past proliferation. *Nat*
1327 *Immunol* **23**, 658–659 (2022).
1328 81. Kovanen, P. E. *et al.* T-cell development and function are modulated by dual specificity
1329 phosphatase DUSP5. *J Biological Chem* **283**, 17362–9 (2008).
1330 82. Chuang, H.-C. & Tan, T.-H. MAP4K Family Kinases and DUSP Family Phosphatases in T-
1331 Cell Signaling and Systemic Lupus Erythematosus. *Cells* **8**, 1433 (2019).
1332 83. Danileviciute, E. *et al.* PARK7/DJ-1 promotes pyruvate dehydrogenase activity and maintains
1333 Treg homeostasis during ageing. *Nat Metabolism* **4**, 589–607 (2022).
1334 84. McHugh, D. *et al.* N-arachidonoyl glycine, an abundant endogenous lipid, potently drives
1335 directed cellular migration through GPR18, the putative abnormal cannabidiol receptor.
1336 *Bmc Neurosci* **11**, 44 (2010).
1337 85. Kan, J. L., Jannatipour, M., Taylor, S. M. & Moran, R. G. Mouse cDNAs encoding a
1338 trifunctional protein of de novo purine synthesis and a related single-domain glycinamide
1339 ribonucleotide synthetase. *Gene* **137**, 195–202 (1993).
1340 86. Varney, M. J. *et al.* G protein–coupled receptor kinase 6 (GRK6) regulates insulin processing
1341 and secretion via effects on proinsulin conversion to insulin. *J Biol Chem* **298**, 102421
1342 (2022).

87. Gilic, M. B. & Tyner, A. L. Targeting protein tyrosine kinase 6 in cancer. *Biochimica Et Biophysica Acta Bba - Rev Cancer* **1874**, 188432 (2020).
88. Slattery, M. L., Lundgreen, A. & Wolff, R. K. MAP kinase genes and colon and rectal cancer. *Carcinogenesis* **33**, 2398–2408 (2012).
89. Knackmuss, U. *et al.* MAP3K11 is a tumor suppressor targeted by the oncomiR miR-125b in early B cells. *Cell Death Differ* **23**, 242–252 (2016).
90. Hammond, R. K. *et al.* Biological constraints on GWAS SNPs at suggestive significance thresholds reveal additional BMI loci. *Elife* **10**, e62206 (2021).
91. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).
92. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
93. Lun, A. T. L. & Smyth, G. K. csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res* **44**, e45–e45 (2016).
94. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinform Oxf Engl* **26**, 139–40 (2009).
95. Wingett, S. *et al.* HiCUP: pipeline for mapping and processing Hi-C data. *F1000research* **4**, 1310 (2015).
96. Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinform Oxf Engl* **36**, 311–316 (2019).
97. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods* **9**, 999 1003 (2012).
98. Nora, E. P. *et al.* Molecular basis of CTCF binding polarity in genome folding. *Nat Commun* **11**, 5612 (2020).
99. Cresswell, K. G. & Dozmorov, M. G. TADCompare: An R Package for Differential and Temporal Analysis of Topologically Associated Domains. *Frontiers Genetics* **11**, 158 (2020).
100. Kaul, A., Bhattacharyya, S. & Ay, F. Identifying statistically significant chromatin contacts from Hi-C data with FitHiC2. *Nat Protoc* **15**, 991–1012 (2020).
101. Stansfield, J. C., Cresswell, K. G., Vladimirov, V. I. & Dozmorov, M. G. HiCcompare: an R-package for joint normalization and comparison of HI-C datasets. *Bmc Bioinformatics* **19**, 279 (2018).
102. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinform Oxf Engl* **31**, 166–9 (2014).
103. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *Bmc Bioinformatics* **14**, 7–7 (2013).
104. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst* **1**, 417–425 (2015).
105. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47–e47 (2015).
106. Li, Z. *et al.* Identification of transcription factor binding sites using ATAC-seq. *Genome Biol* **20**, 45 (2019).
107. Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **48**, D87–D92 (2020).
108. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228–35 (2015).
109. 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.
110. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *Plos Comput Biol* **9**, e1003118 (2013).

111. Lopez-Delisle, L. *et al.* pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinform Oxf Engl* **37**, 422–423 (2020).
112. 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).
113. 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).
114. Wagner, A. H. *et al.* DGIdb 2.0: mining clinically relevant drug–gene interactions. *Nucleic Acids Res* **44**, D1036–D1044 (2016).
115. Consortium, T. Gte. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
116. Wells, A. D., Gudmundsdottir, H. & Turka, L. A. Following the fate of individual T cells throughout activation and clonal expansion. Signals from T cell receptor and CD28 differentially regulate the induction and duration of a proliferative response. *J Clin Invest* **100**, 3173 3183 (1997).

Figure 1

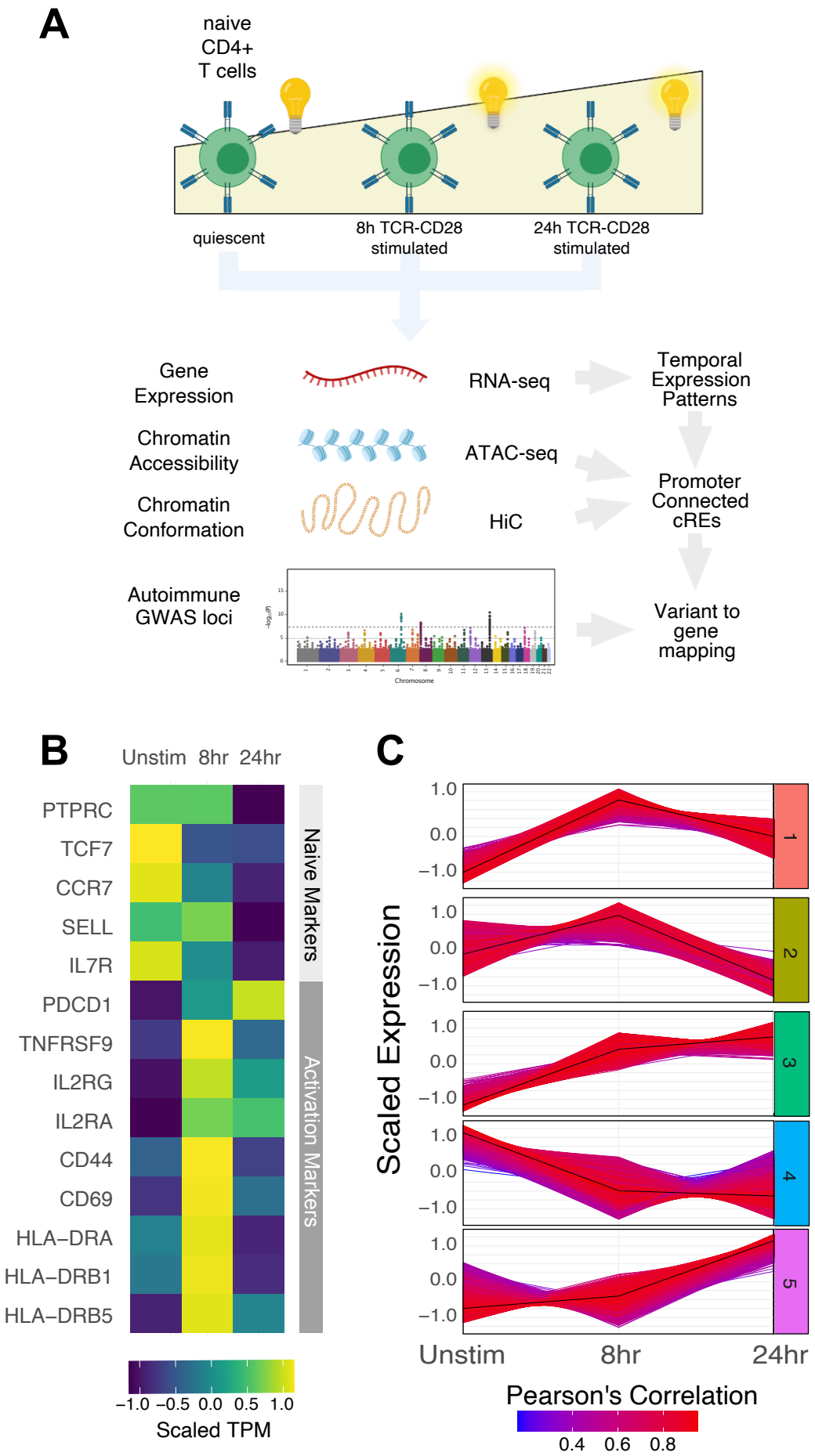

A
naive
CD4+
T cells
quiescent
8h TCR-CD28
stimulated
24h TCR-CD28
stimulated
Gene
Expression
RNA-seq
Temporal
Expression
Patterns
Chromatin
Accessibility
ATAC-seq
Chromatin
Conformation
HiC
Promoter
Connected
cREs
Autoimmune
GWAS loci
Chromosome
Variant to
gene
mapping
B
Unstim 8hr 24hr
PTPRC
TCF7
CCR7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

Figure 2

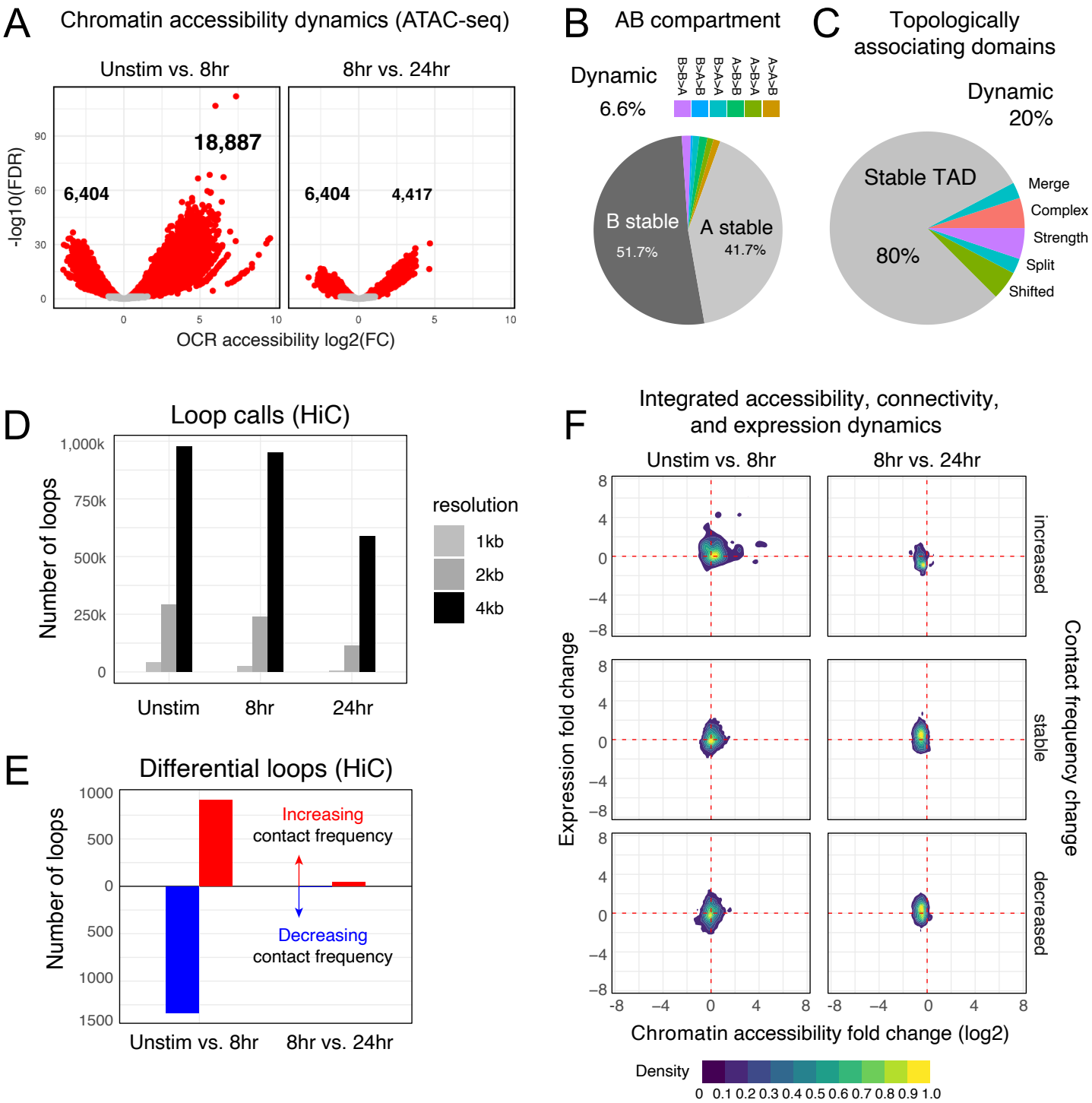

A
Chromatin accessibility dynamics (ATAC-seq)
Unstim vs. 8hr
8hr vs. 24hr
-log10(FDR)
18,887
6,404
6,404
4,417
OCR accessibility log2(FC)
B
AB compartment
Dynamic
6.6%
B>B>A
B<A>B
B>A>A
A>B>B
A<B>A
A>A>B
B stable
51.7%
A stable
41.7%
C
Topologically associating domains
Dynamic
20%
Stable TAD
80%
Merge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

Figure 3

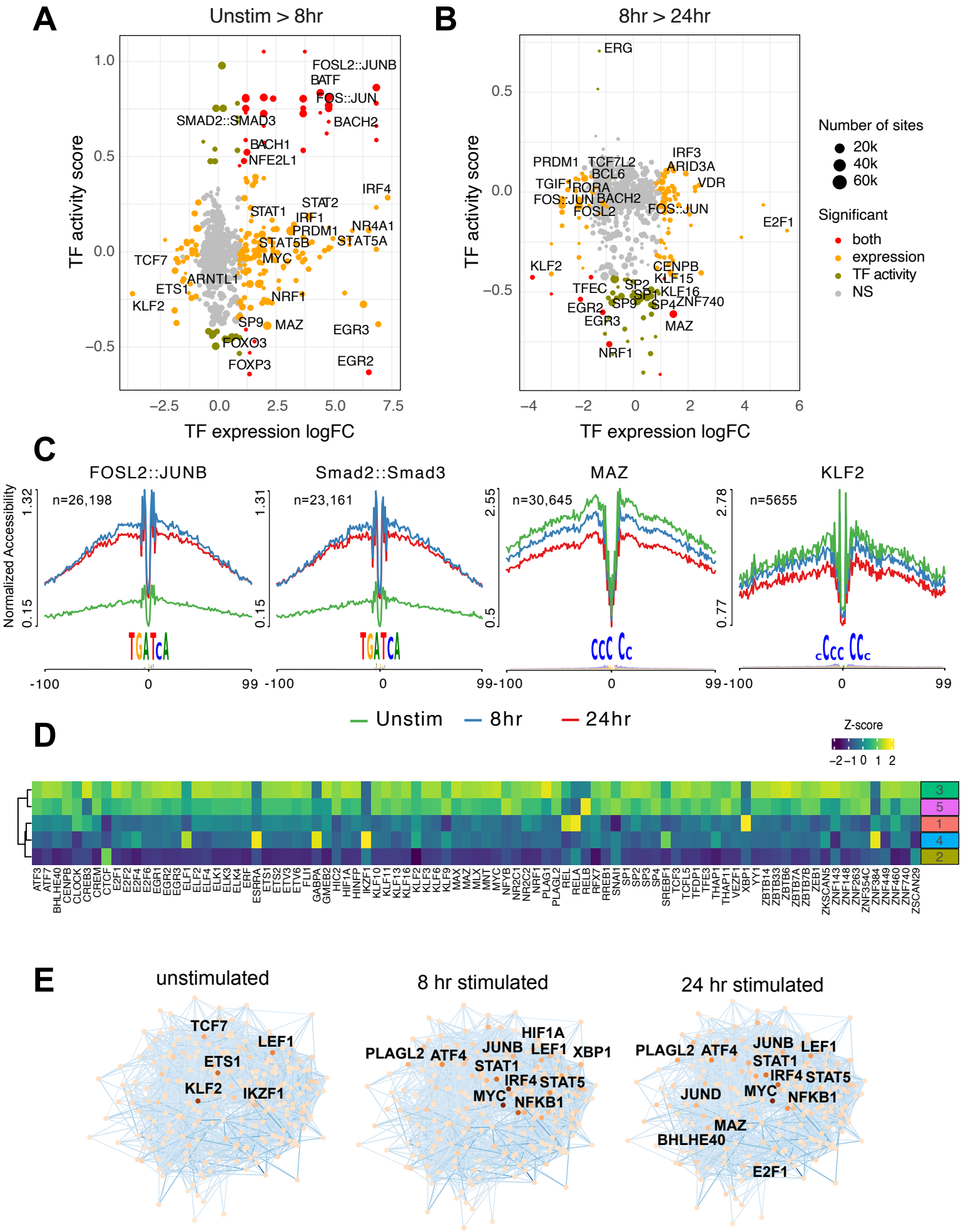

A
Unstim > 8hr
TF activity score
TF expression logFC
FOSL2::JUNB
BATF
FOS::JUN
BACH2
SMAD2::SMAD3
BACH1
NFE2L1
IRF4
STAT2
STAT1
IRF1
PRDM1
NR4A1
STAT5B
STAT5A
TCF7
MYC
ARNTL1
ETS1
NRF1
KLF2
SP9
MAZ
EGR3
FOXO3
FOXP3
EGR2
B
8hr > 24hr
ERG
PRDM1
TCF7L2
BCL6
IRF3
ARID3A
VDR
TGIF1
RORA
FOS::JUN
BACH2
FOSL2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

Figure 4

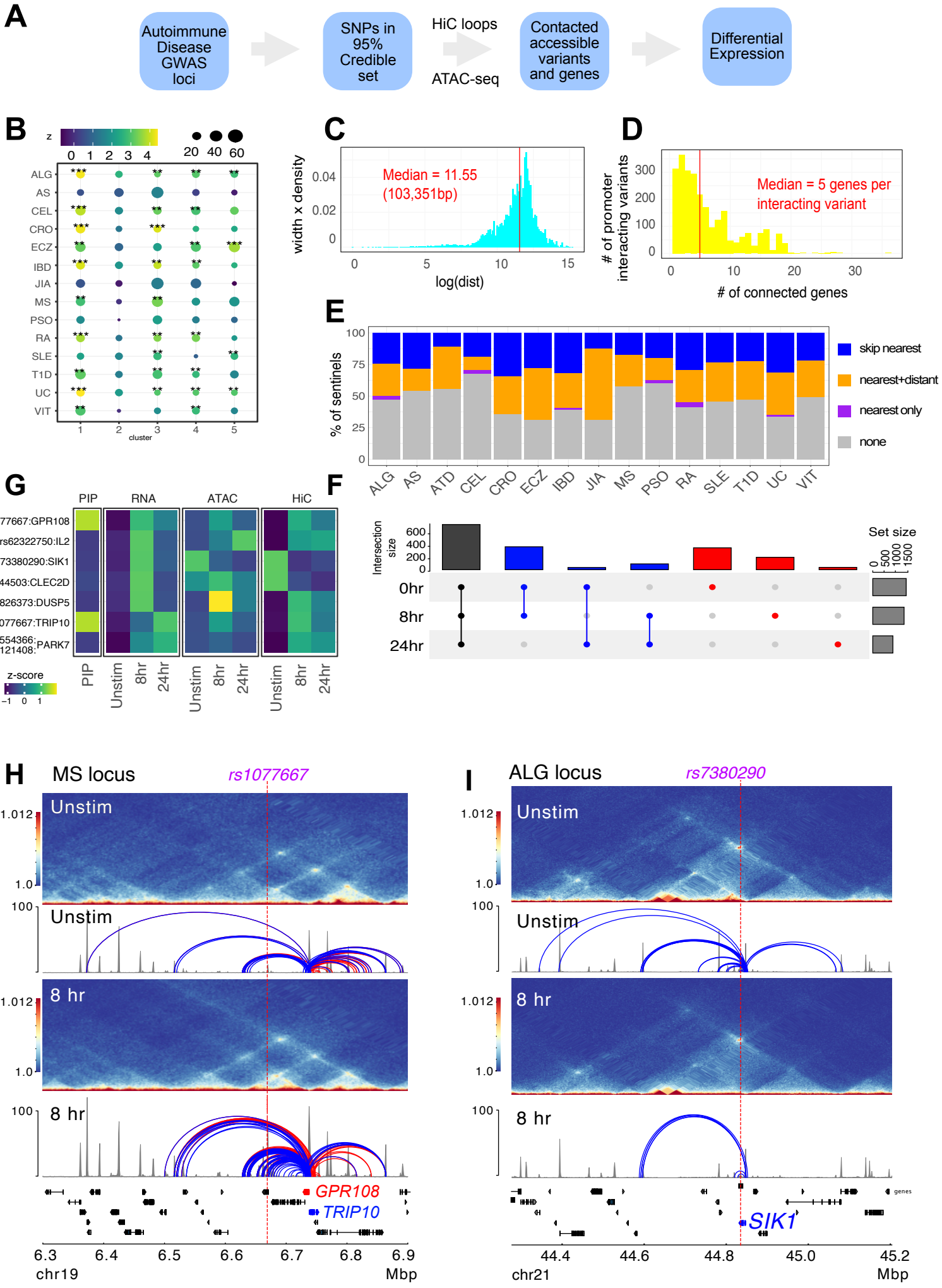
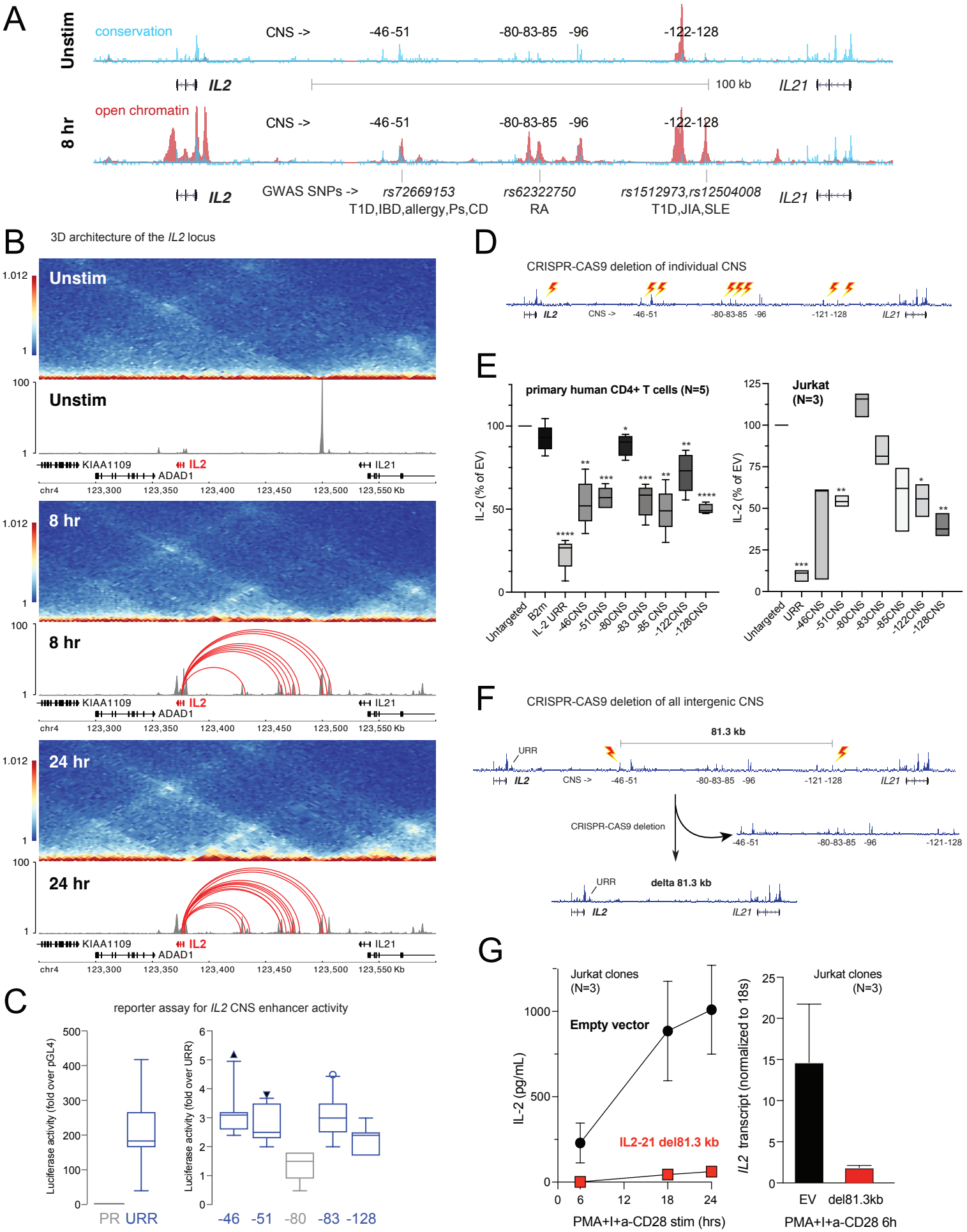

A
Autoimmune Disease GWAS loci
SNPs in 95% Credible set
HiC loops
ATAC-seq
Contacted accessible variants and genes
Differential Expression
B
z
0 1 2 3 4
20 40 60
ALG
AS
CEL
CRO
ECZ
IBD
JIA
MS
PSO
RA
SLE
T1D
UC
VIT
1 2 3 4 5
cluster
C
width x density
Median = 11.55 (103,351bp)
log(dist)
D
# of promoter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

Figure 5


A
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

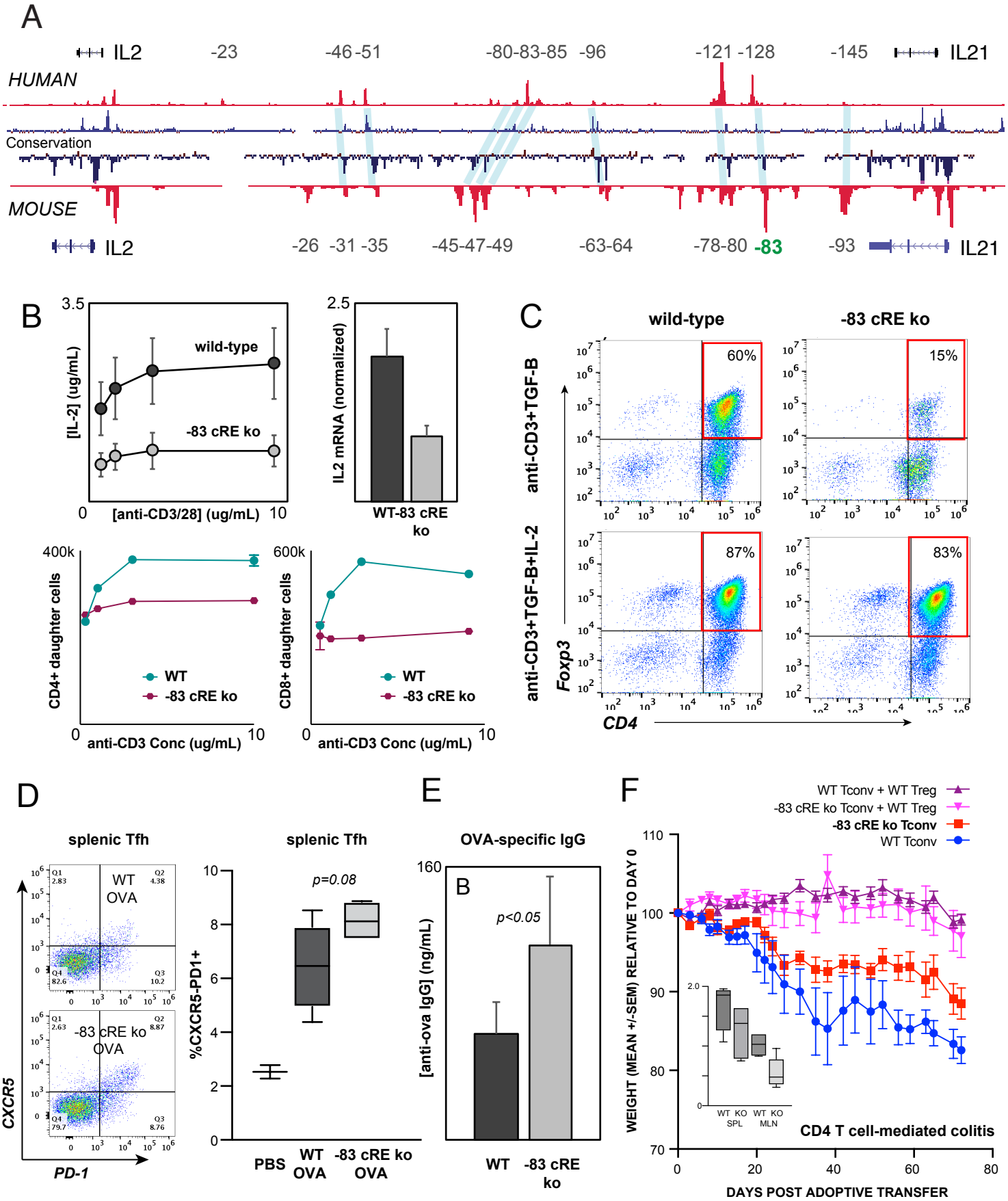

A
IL2
-23
-46 -51
-80-83-85 -96
-121 -128
-145
IL21
HUMAN
Conservation
MOUSE
IL2
-26 -31 -35
-45-47-49
-63-64
-78-80 -83
-93
IL21
B
[IL-2] (ug/mL)
wild-type
-83 cRE ko
[anti-CD3/28] (ug/mL)
IL2 mRNA (normalized)
WT-83 cRE ko
CD4+ daughter cells
CD8+ daughter cells
WT
-83 cRE ko
anti-CD3 Conc (ug/mL)
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 7

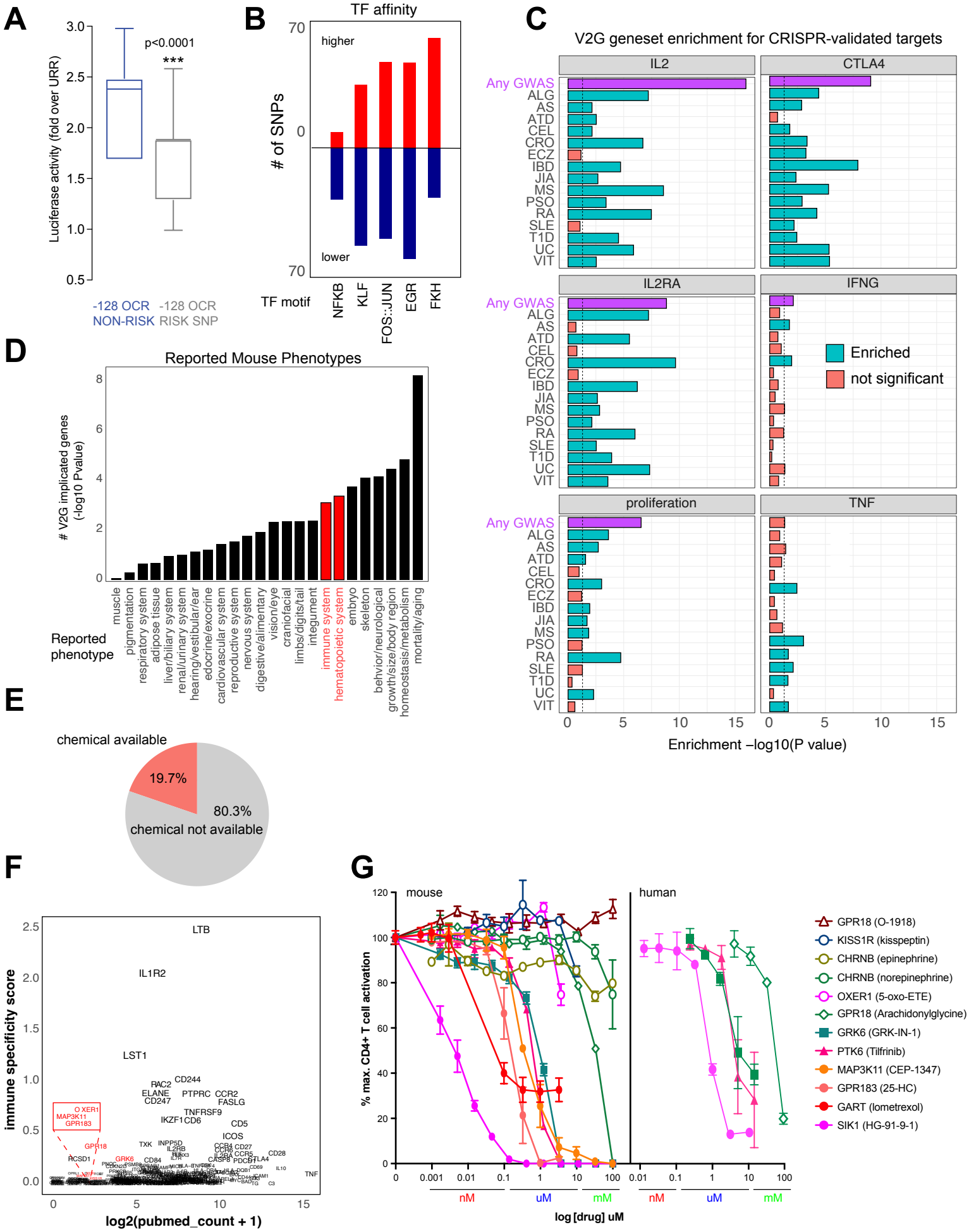

A
Luciferase activity (fold over URR)
p<0.0001
***
-128 OCR NON-RISK
-128 OCR RISK SNP
B
TF affinity
higher
lower
# of SNPs
TF motif
NFKB
KLF
FOS::JUN
EGR
FKH
C
V2G geneset enrichment for CRISPR-validated targets
IL2
CTLA4
IL2RA
IFNG
proliferation
TNF
Any GWAS
ALG
AS
ATD
CEL
CRO
ECZ
IBD
JIA
MS
PSO
RA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

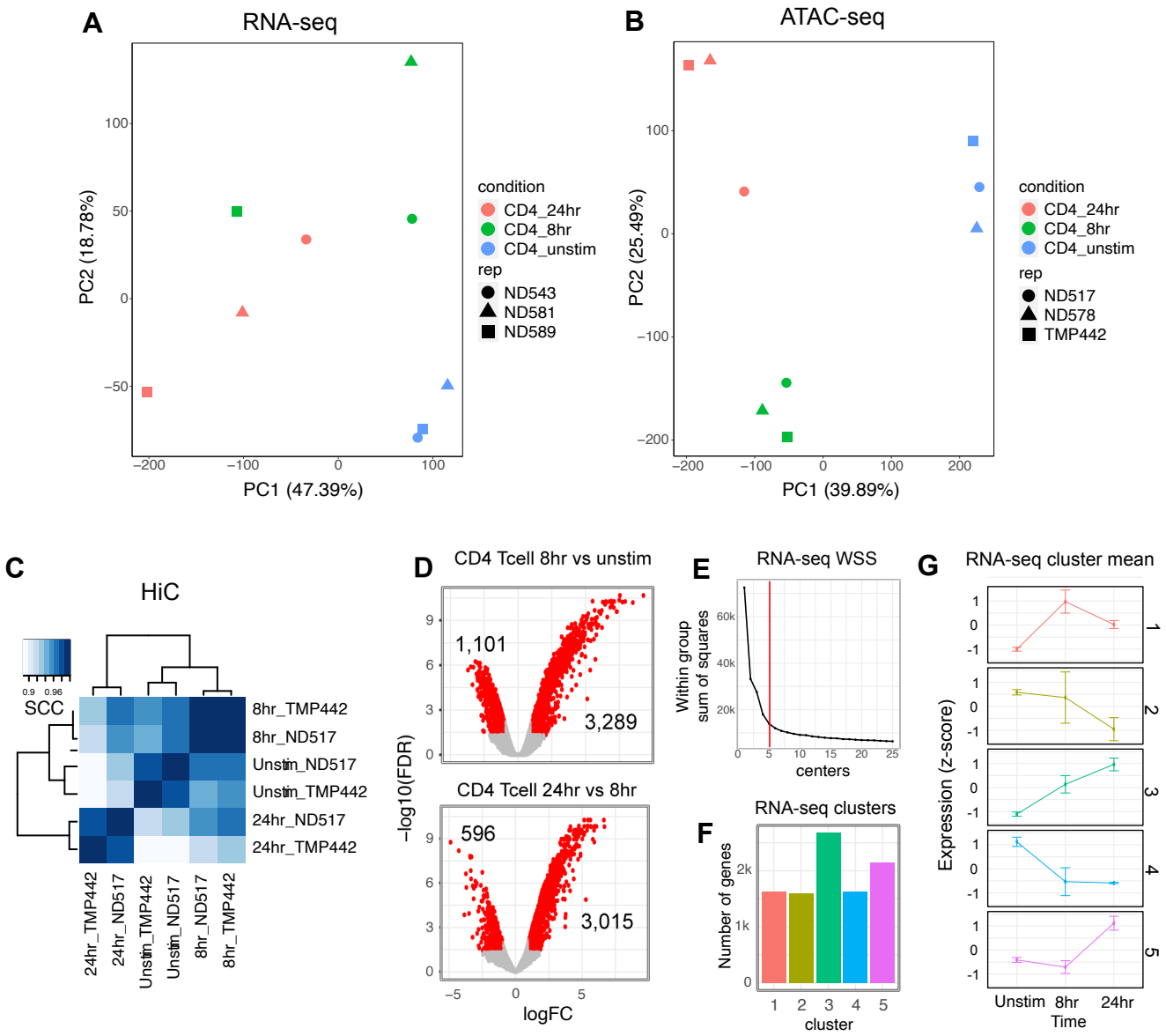


Figure S2: Characterization of CD4+ T cell epigenomic data. (A) Accessibility of OCR and (B) expression of genes located in A vs. B compartments (log2 fkpm). Differential accessibility (C) and expression (D) of cRE-gene pairs in regions exhibiting differential TAD structure. (E) Enrichment of promoter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

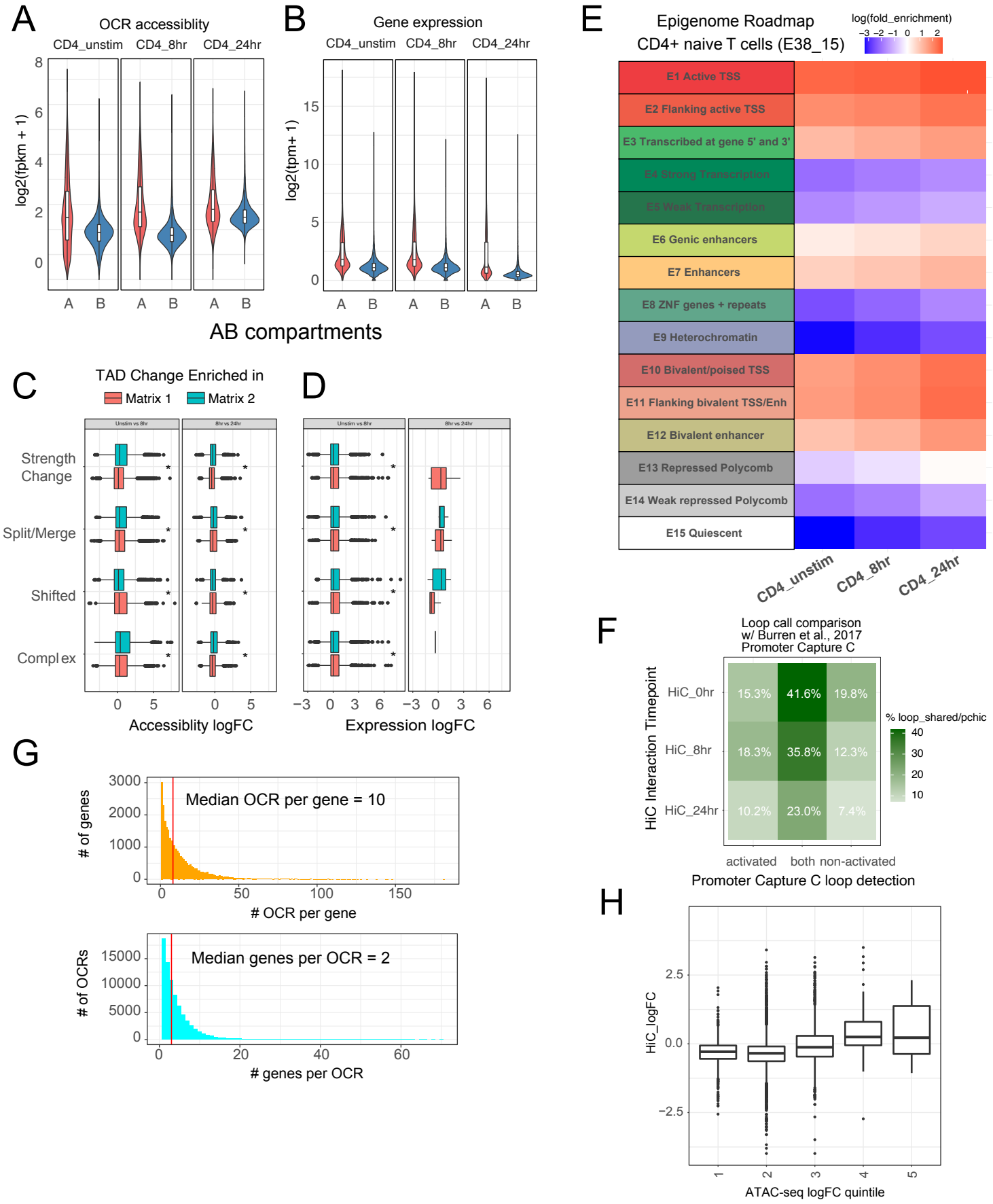


Figure S3: Characterization of chromatin stripes during CD4+ T cell activation. (A) Stripes called per timepoint. (B) Proportion of SNPs annotated in A vs B compartments. Accessibility of OCR (C) and expression of genes (D) located in anchor (small more interactive region of stripe), stripe excluding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

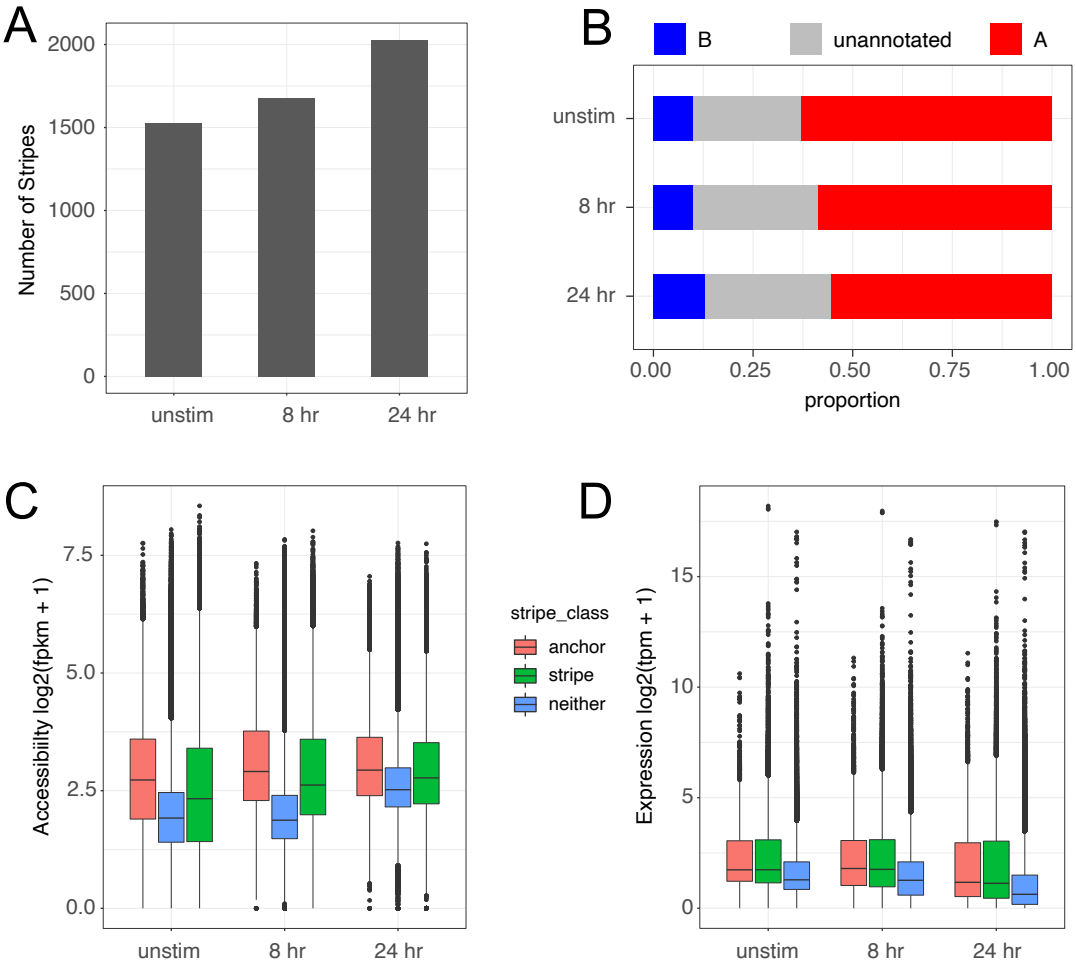


Figure S4: Large-scale epigenetic changes at dynamically expressed genes during CD4+ T cell activation. HiC contact frequency matrix (heatmap), chromatin accessibility (grey), and loop calls (red) for (A) GEM, (B) IRF4, (C) KLF2, and (D) DPEP2. The IRF4 matrix is truncated at by end of the chromosome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

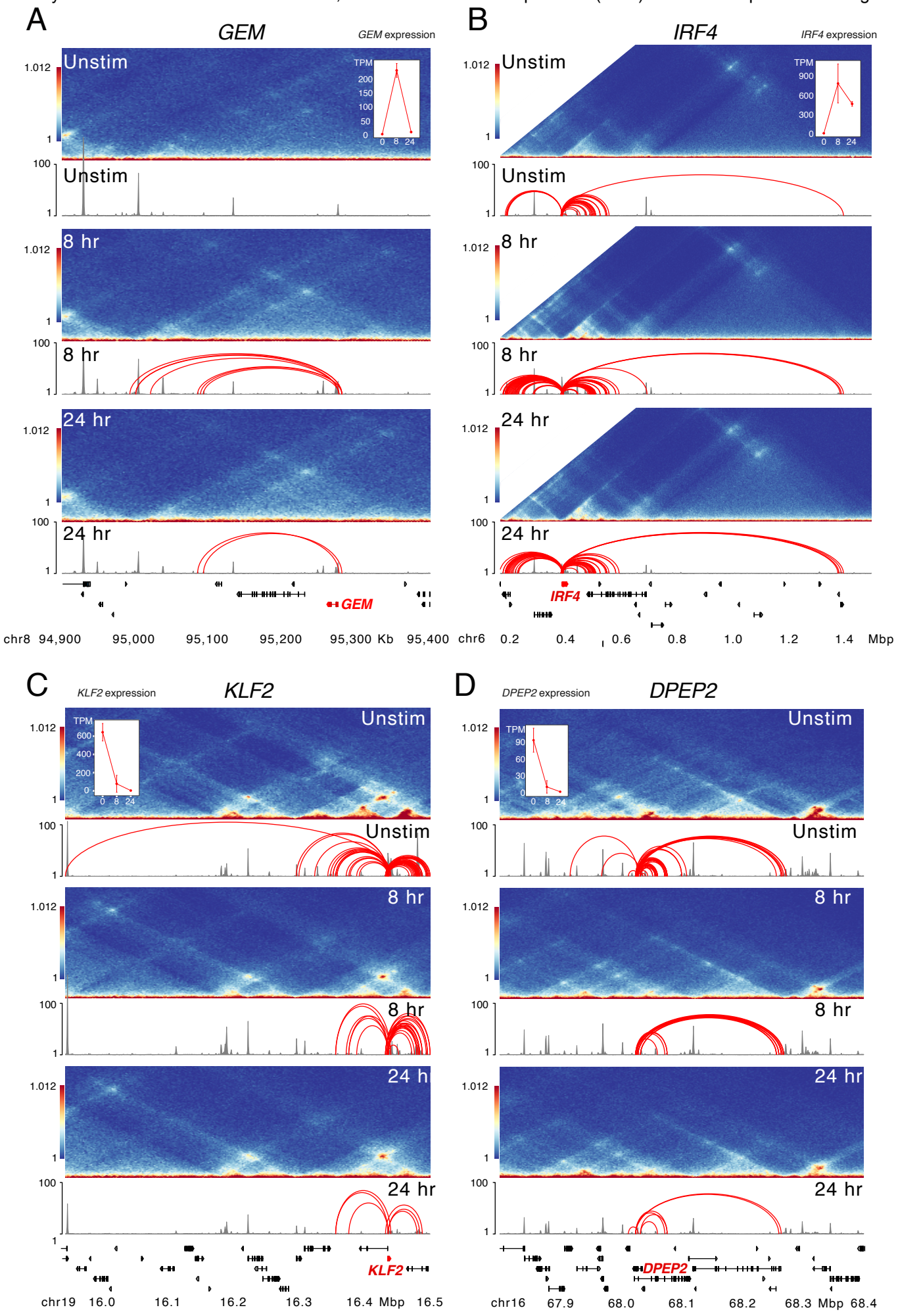


Figure S5: Enrichment of transcription family members and gene regulatory network construction. (A) -log10(P) enrichment of transcription factors annotated by family for each expression cluster. (B) Top five pathways enriched for a subset of TF-regulated gene expression.

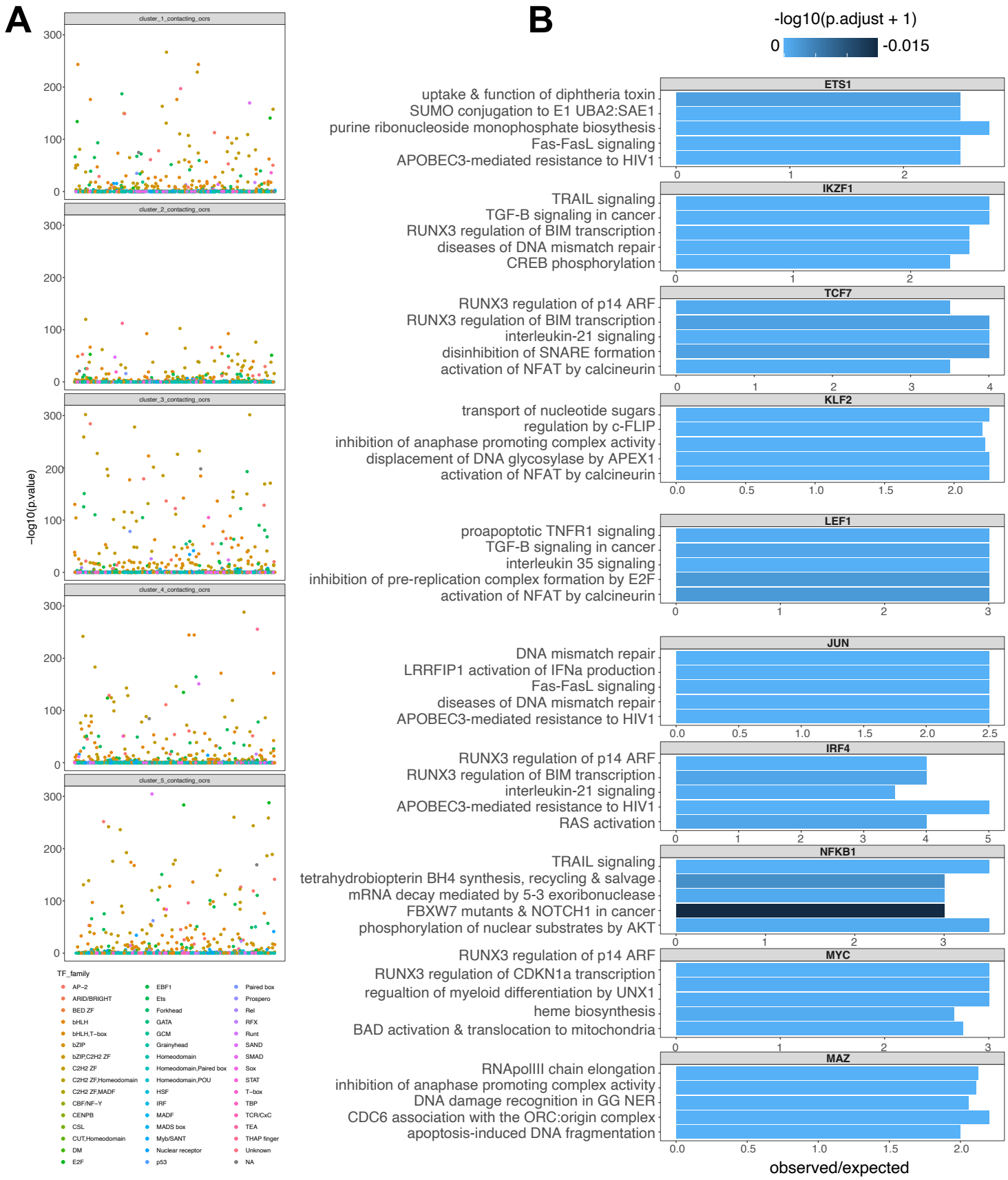


Figure S6: Comparative predictive power of various V2G approaches. (**A**) Percent overlap in predicted genes among chromatin-based and eQTL-based V2G approaches. (**B**) Expression (scaled TPM) of genes implicated by both eQTL (Soskic et al.) and HiC. (**C**) Physical variant-gene associations (HiC) are enriched for statistical variant-gene associations (eQTL) in activated CD4+ T cells. The histogram depicts the null distribution of shared variant-gene pairs expected at random (~5) while the red line indicates the observed number of variant-gene pairs (41) shared with the 127 eQTL identified by Soskic et al. in a similar CD4+ T cell activation system. (**C**) Precision-Recall analysis of V2G gene predictions against the set of monogenic human inborn errors in immunity.

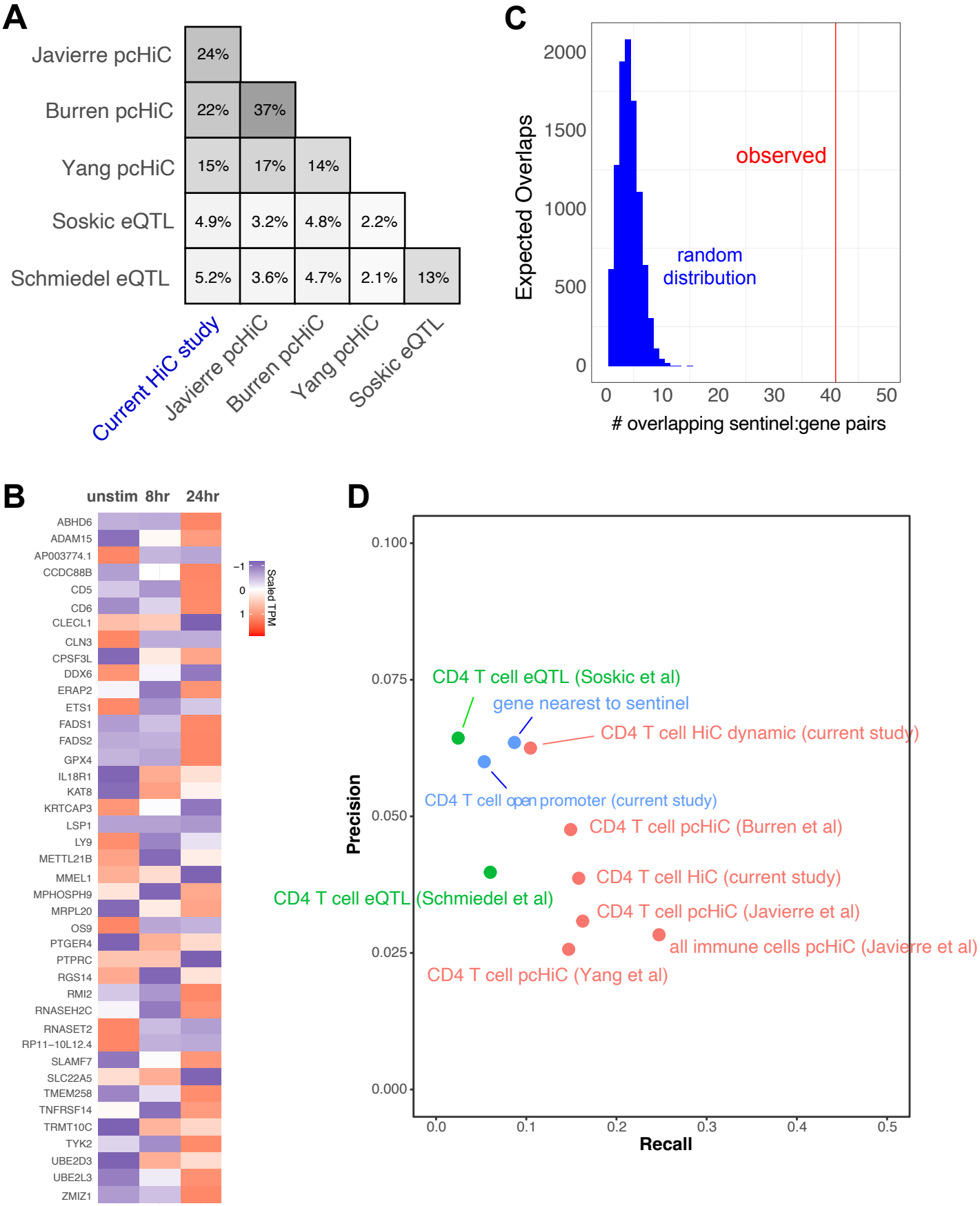


Figure S7: TF occupancy and stability of distal IL2 cRE in CD4+ T cell subsets. Evolutionary conservation (blue), open chromatin (red), and autoimmune disease-associated SNPs at the IL2 locus in quiescent (A) and 8-hour activated (B) naïve CD4+ T cells. (C) Curated TF occupancy at the IL2 URR and distal cRE from the ReMap atlas of regulatory regions.

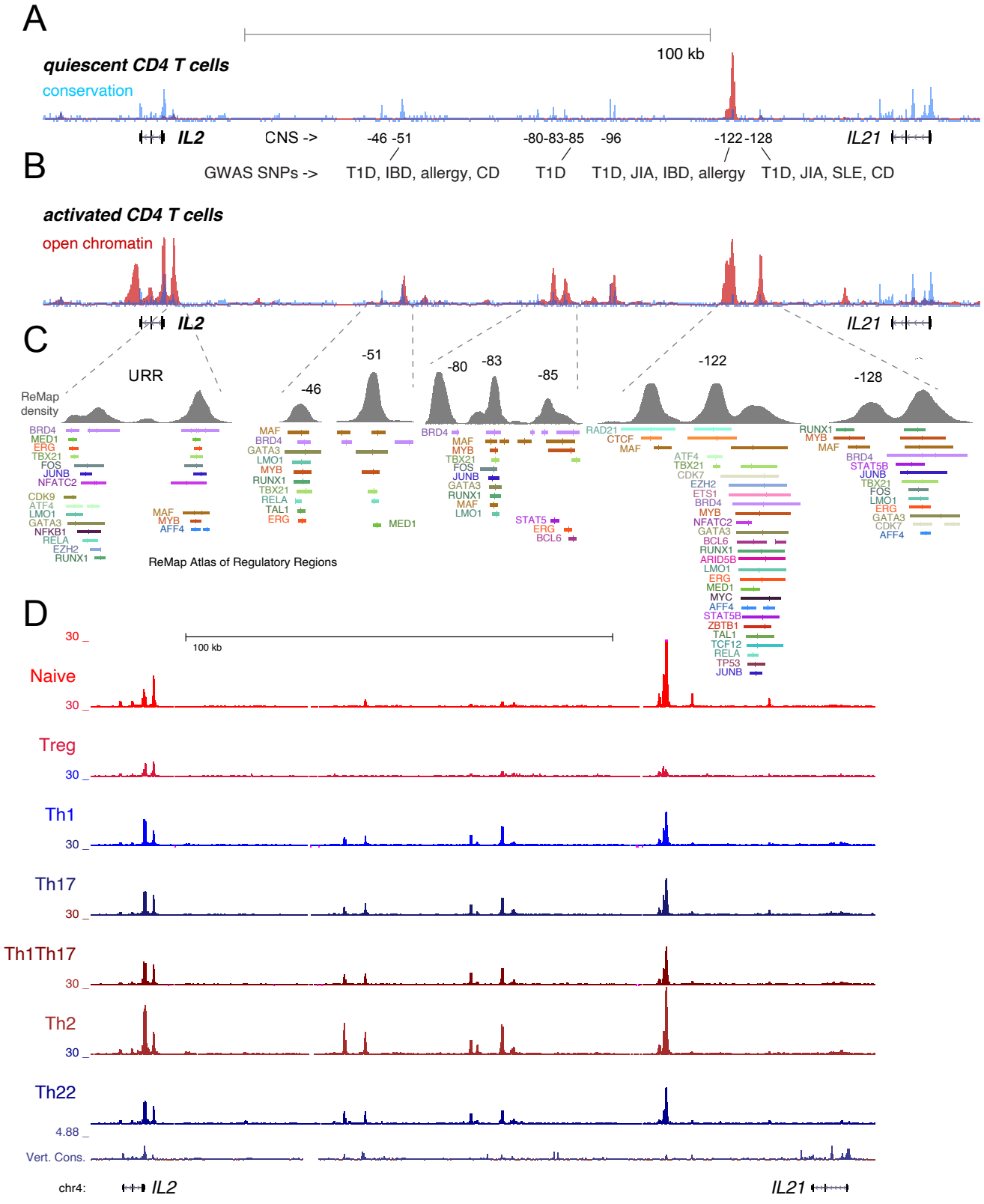


Figure S8: Predicted impact of autoimmune disease-associated genetic variation at V2G-implicated loci. (A) Top 50 TF motifs impacted by autoimmune SNPs (see Table S15). HiC contact frequency matrix (heatmap), chromatin accessibility (grey), and loop calls (blue-IL19/ITGA4, red-FAIM3) at the T1D-rs3024505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

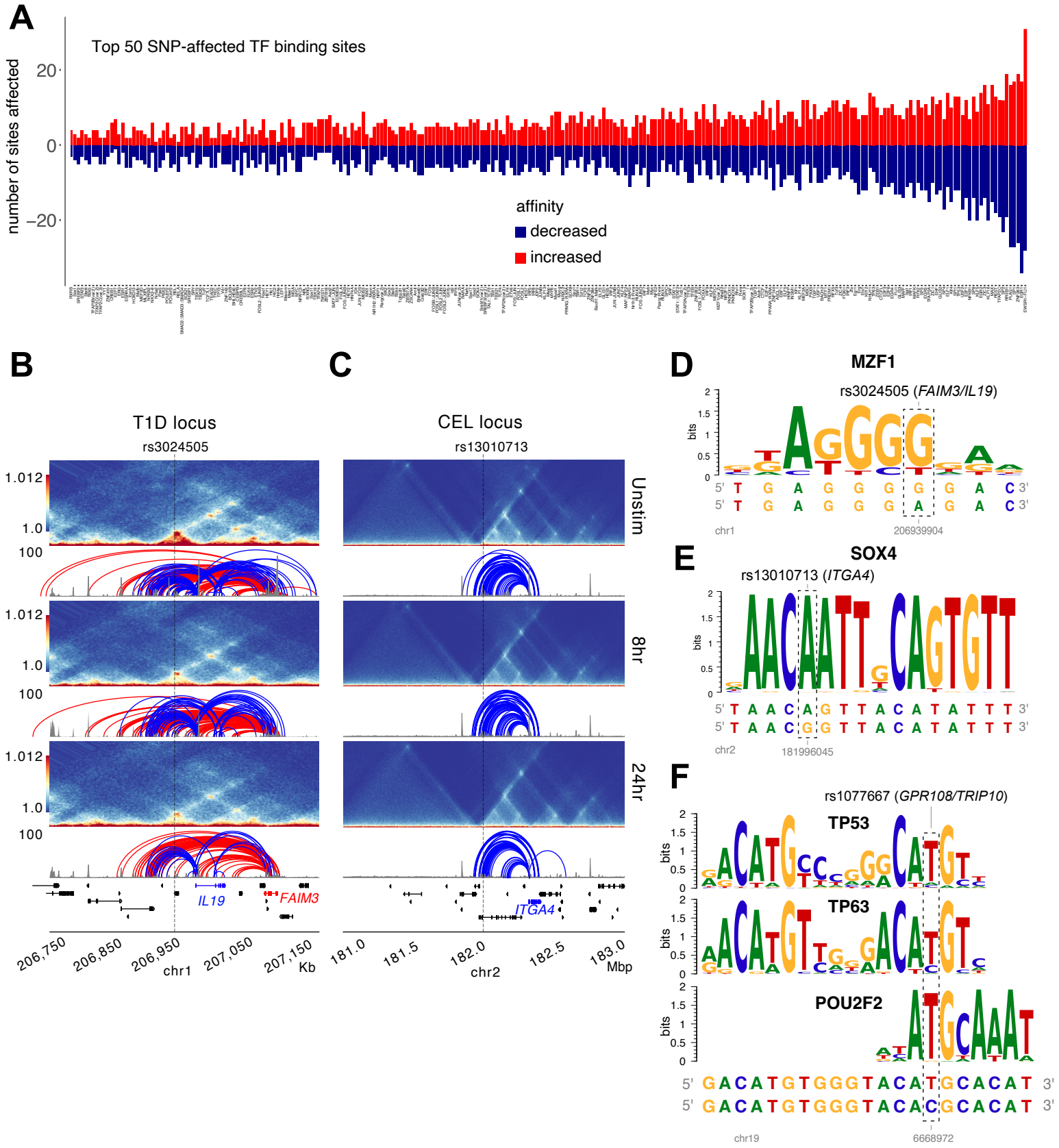
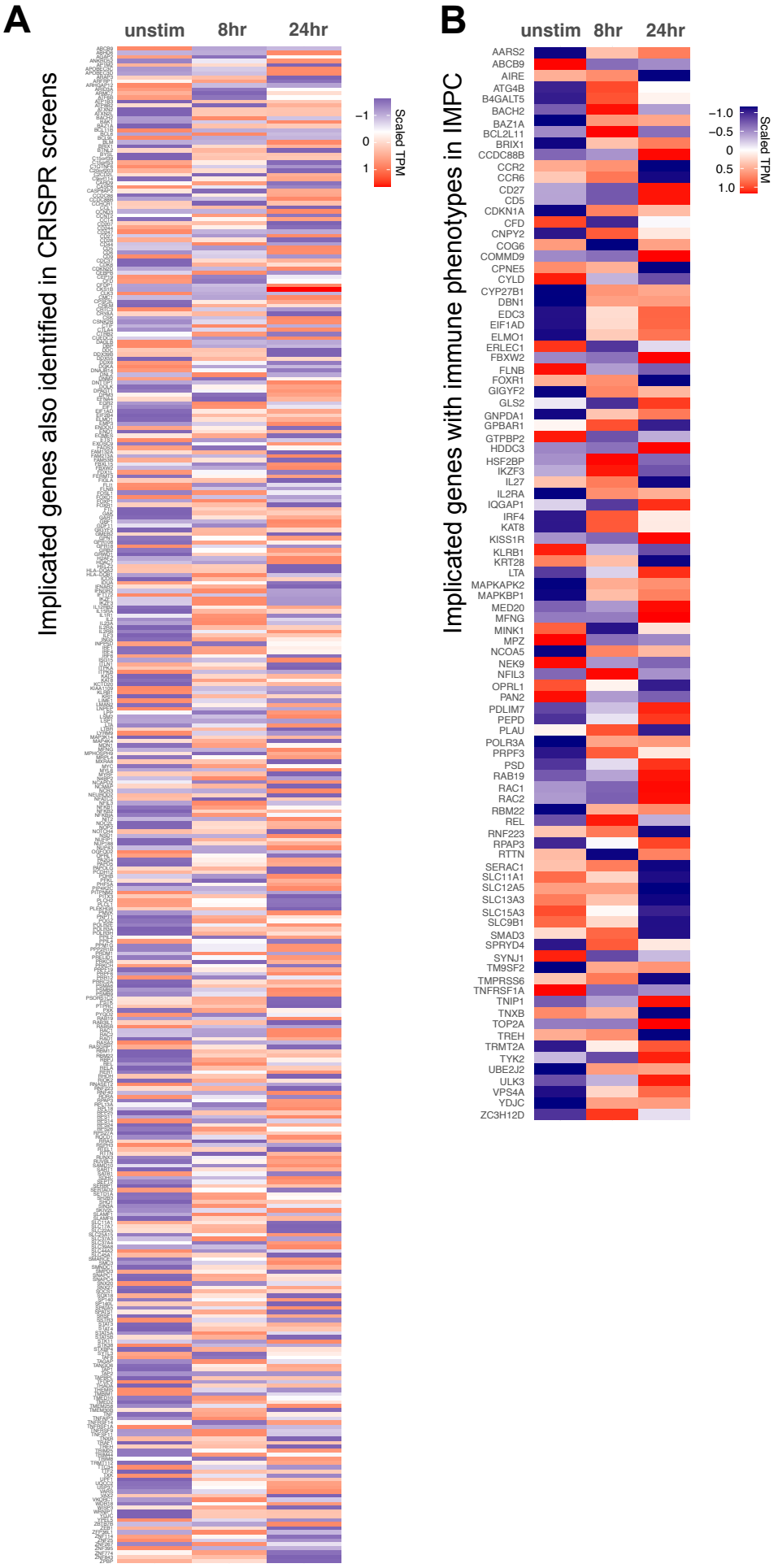



Figure S9: Orthogonal validation of 3D chromatin-based V2G genes. Expression (scaled TPM) of genes implicated by chromatin contacts in (A) prior CRISPRi/a screens and (C) genes with known immune phenotypes in the international mouse phenotyping consortium. Red = increased, blue = decreased.