

1 **An Autoimmune Transcriptional Circuit Driving Foxp3⁺ Regulatory T cell Dysfunction**

2

3 Tomokazu S. Sumida^{1,2,8}, Matthew R. Lincoln^{1,2,8}, Liang He^{2,3,4,8}, Yongjin Park^{2,3, 8}, Mineto Ota⁶,
4 Helen A. Stillwell¹, Greta A. Leissa¹, Keishi Fujio⁶, Alexander M. Kulminski⁴, Charles B. Epstein²,
5 Bradley E. Bernstein^{2,7}, Manolis Kellis^{2,3}, David A. Hafler^{1,2, †}

6

7 ¹Departments of Neurology and Immunobiology, Yale School of Medicine, New Haven, CT, USA

8 ²Broad Institute of MIT and Harvard, Cambridge, MA, USA

9 ³Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, USA

10 ⁴Biodemography of Aging Research Unit, Social Science Research Institute, Duke University,
11 Durham, NC, USA

12 ⁶Department of Allergy and Rheumatology, Graduate School of Medicine, The University of Tokyo,
13 Tokyo, Japan

14 ⁷Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA.

15 ⁸These authors contributed equally

16 † Correspondence to David Hafler (david.hafler@yale.edu)

17

18 Present Address:

19 M.R.L.: Division of Neurology, Department of Medicine, University of Toronto, and Keenan
20 Research Centre for Biomedical Science of St. Michael's Hospital, Toronto, ON, Canada

21 L.H.: Department of Public Health, Faculty of Health Sciences, University of Southern Denmark,
22 Odense M, Denmark

23 Y.P.: Department of Pathology and Laboratory Medicine, Department of Statistics, University of
24 British Columbia, Vancouver, BC, Canada

25 **Abstract**

26 Autoimmune diseases, among the most common disorders of young adults, are mediated by
27 genetic and environmental factors. While CD4⁺Foxp3⁺ regulatory T cells (Tregs) play a central role
28 in preventing autoimmunity, the molecular mechanism underlying their dysfunction is unknown.
29 Here, we performed comprehensive transcriptomic and epigenomic profiling of Tregs in the
30 autoimmune disease multiple sclerosis (MS) to identify central transcriptional programs regulating
31 human autoimmunity. We discovered that upregulation of a primate-specific short *PRDM1* isoform
32 (*PRDM1-S*) induces *SGK1* independent from evolutionally conserved long *PRDM1*, leading to
33 destabilization of Foxp3 and Treg dysfunction. This aberrant *PRDM1-S/SGK1* axis is shared
34 among other autoimmune diseases. Furthermore, by chromatin landscape profiling in MS Tregs
35 we identified a *PRDM1-S* specific *cis*-regulatory element associated with enriched binding of AP-
36 1/IRF transcription factors. Our study identifies evolutionally emerged *PRDM1-S* and epigenetic
37 priming of AP-1/IRF as key drivers of pathogenic Treg programs leading to human autoimmune
38 disease.

39 **(146 words)**

40 Introduction

41 Human autoimmune diseases constitute a leading cause of death among young adults with
42 an increasing incidence in recent years. Multiple sclerosis (MS) is a canonical, genetically mediated
43 autoimmune disease induced by environmental factors where genetic perturbation of *cis*-regulatory
44 elements in pathogenic immune cells leads to immune dysregulation and generation of
45 autoreactive T cells and antibodies¹⁻³. Among immune cells, CD4⁺ T cells play a central role in both
46 mediating and regulating autoimmunity. As CD4⁺ T cells display a large degree of functional
47 diversity, interrogation of total CD4⁺ T cell populations has not to date identified causal
48 transcriptional changes^{4,5}. Thus, we hypothesized that interrogation of CD4⁺ T cell subpopulations
49 is required to elucidate the pathophysiological characteristics of CD4⁺ T cells in autoimmune
50 disorders, allowing identification of central transcriptional factors associated with loss of immune
51 regulation

52 Human CD4⁺ Foxp3⁺ regulatory T cells (Tregs) play a central role in the maintenance of
53 immune homeostasis and prevention of autoimmunity⁶⁻⁸. We first demonstrated that Tregs from
54 patients with autoimmune disease exhibit a dysfunctional phenotype⁹⁻¹¹ and this has been
55 subsequently found among multiple autoimmune disorders. Recent evidence has shown that
56 environmental factors, such as vitamin D and fatty acids affect Treg phenotype and function. In
57 addition, high salt intake has been epidemiologically linked with autoimmune diseases¹² and higher
58 physiologic salt concentrations induce proinflammatory Th17 cells mediated by SGK1¹³ and
59 modulate the stability of Tregs^{14,15}, resembling the phenotype observed in autoimmune diseases
60 including MS¹⁶. These environmental factors are known to act, in part, through epigenetic changes
61 that can be identified by examining the histone and methylation landscape of cells¹⁷. A recent
62 investigation of the genetic architecture of 20 autoimmune diseases and MS showed that causal
63 variants are enriched in regulatory elements that are active in immune cells³. Moreover, many of
64 these genetic loci are common among multiple autoimmune diseases, implicating the shared
65 immunomodulatory mechanism in human autoimmunity^{18,19}. These findings, together with
66 increases in the incidence of autoimmune diseases over the past three decades that cannot be
67 explained by genetic factors alone, points to both genetic and epigenetic factors as key mediators
68 of risk for autoimmunity. While these lines of evidence strongly suggest that both genetic and
69 epigenetic alterations might disturb Treg homeostasis, the underlying mechanism in Treg
70 dysfunction in human autoimmune diseases has not been elucidated.

71 We interrogated the phenotypic and functional characteristics of human CD4⁺ T cells,
72 focusing primarily on memory Treg (mTreg) through comprehensive transcriptomic and epigenetic

73 profiling. We adopted both bulk and single-cell RNA-seq (scRNA-seq) for transcriptomic profiling
74 and Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) for probing
75 epigenetic regulation to understand the molecular mechanisms that drive dysfunctional programs
76 in MS Tregs. Here, we reveal that *PRDM1*, which encodes Blimp1, is upregulated in both mTreg
77 and mTconv from patients with MS with a more significant increase in mTreg. This transcriptional
78 signature of Tregs is shared across different autoimmune diseases, suggesting the common
79 transcriptional signature driving dysfunctional Tregs in human autoimmunity. Specifically, an
80 alternative short isoform of *PRDM1* (*PRDM1-S*), which codes for protein in primates but not in
81 rodents, is primarily elevated in MS mTreg compared to the evolutionary conserved long *PRDM1*
82 isoform (*PRDM1-L*). Gene overexpression experiments in primary human Tregs, together with bulk
83 and single-cell RNA-seq transcriptional profiling, demonstrate a unique link between the *PRDM1-S*
84 and *SGK1* that accounts for Treg dysfunction observed in MS. Moreover, while genome-wide
85 chromatin accessibility in mTreg remains comparable between MS and healthy controls, the
86 transcription factor (TF) footprints and motifs within accessible chromatin regions are different with
87 significantly enriched binding of AP-1 and IRF family TFs observed in MS mTreg, suggesting
88 rewiring of regulatory circuits. Finally, CRISPR activation (CRISPRa)-based exploration of *cis*-
89 regulatory elements for the *PRDM1-S* identified a *cis*-regulatory element harboring AP-1 and IRF4
90 composite motif. These findings suggest a possible novel regulatory program by which Tregs
91 become dysfunctional in humans that is shared across multiple autoimmune diseases. Our
92 multimodal datasets of human CD4⁺ T cells provide a rich resource for understanding the loss of
93 immune regulation in autoimmune diseases and suggest that the primate specific short *PRDM1*
94 isoform is a critical, targetable transcriptional regulator in human autoimmunity.

95

96 **Results**

97 ***Deep transcriptomic analysis of memory Treg and Tconv highlight PRDM1 upregulation in*** 98 ***MS (Figure 1)***

99 We sought to identify CD4⁺ T cell transcriptional differences between patients with MS and
100 healthy controls. As previous studies had not identified significant differences in bulk transcriptional
101 profiles of whole CD4⁺ T cells between MS subjects and healthy controls^{4,5}, we divided CD4⁺ T
102 cells into four major subpopulations based on two categories; Tconv vs. Treg, and naïve vs.
103 memory (Figure S1A), where each population is hypothesized to be involved in MS pathogenesis.
104 For example, mTconv contains pathogenic CD4⁺ T cells in patients with MS, such as myelin-
105 reactive T cells, which display the signatures of Th1 and/or Th17 cells²⁰⁻²³. Moreover, mTregs with

106 reduced suppressive function are implicated in MS pathophysiology^{24,25}. To provide a
107 comprehensive transcriptomic catalogue of memory CD4⁺ T cell subpopulations in patients with
108 relapsing-remitting MS (RRMS), we performed RNA-seq on *ex vivo* mTreg and mTconv isolated
109 from the peripheral blood of untreated RRMS patients free of steroid treatment and healthy control
110 subjects as a discovery cohort (Figure 1A, S1A). The control subjects were matched by age, sex,
111 and ethnicity (clinical characteristics are described in Supplementary Table 1). Among a total of 90
112 RNA-seq samples (48 mTconv and 42 mTreg samples) in the discovery cohort (HC; n = 21, MS; n
113 = 30), 84 samples (HC; n = 20, MS; n = 26) passed quality control (Method Details). We identified
114 21 and 243 differentially expressed genes (DEGs, defined as $|\log_2FC| > 0.6$, $FDR < 0.1$) between
115 MS and healthy subjects for mTreg and mTconv, respectively (Figure 1B). Several DEGs were up-
116 or down-regulated in the same direction in both mTreg and mTconv: *PRDM1*, *BCL3*, and *PIM3*
117 were upregulated genes, and *ID3*, *TOB2*, and *LBH* were downregulated genes in MS (Figure 1C,
118 S1B). *PRDM1* was identified as one of the top genes significantly upregulated in MS in both mTreg
119 and mTconv (Figure S1C). Reduced expression of *ID3* in both cell types in MS reflects the negative
120 regulation of *ID3* by *PRDM1* consistent with the known function of *ID3* in maintaining Foxp3
121 expression in Treg²⁶ (Figure 1D). We confirmed that the frequency of Foxp3⁺ Tregs within CD4⁺ T
122 cells were not changed in patients with MS (Figure S1D), consistent with previous studies that
123 quantitative loss of Foxp3⁺ Tregs is unlikely to be the primary feature of disturbed peripheral T cell
124 tolerance in patients with MS or other autoimmune diseases.

125 We validated several of the top DEGs (*PRDM1*, *BCL3*, *RHBDD2*, *TOB2*, *LBH*) in mTreg by
126 using qPCR with an independent cohort of patients with MS (n = 16) and healthy controls (n = 23)
127 (Figure 1A and E, Figure S1E, patient characteristics are described in Supplementary Table 1). We
128 further confirmed the up regulation of *PRDM1* in MS mTreg and mTconv at the protein level by
129 intracellular staining of Blimp1 using flow cytometry (Figure 1E). We then asked whether this Treg
130 transcriptional signature observed in MS is shared across different autoimmune diseases using
131 ImmuNexUT data²⁷ where transcriptomic profiles of multiple peripheral immune cells, including
132 mTreg and mTconv, were explored at population scale across multiple autoimmune diseases.
133 Notably, the transcriptional signature that we observed in MS Treg was also observed in Treg
134 among most of the autoimmune diseases we analyzed (total 12 diseases: 10 diseases from
135 ImmuNexUT, T1D from P. Gao *et al.*²⁸, and MS from this study) (Figure 1F, S1F). Inverse regulation
136 of *PRDM1* and *ID3* observed in MS was highlighted in Tregs from patients with systemic lupus
137 erythematosus (SLE), Idiopathic inflammatory myopathy (IIM), and ANCA-associated vasculitis

138 (AAV) (Figure S1G). Our bulk RNA-seq transcriptional profiling of memory Tconv and Treg
 139 highlighted *PRDM1* as a key regulatory factor in dysfunctional Tregs in autoimmune diseases.

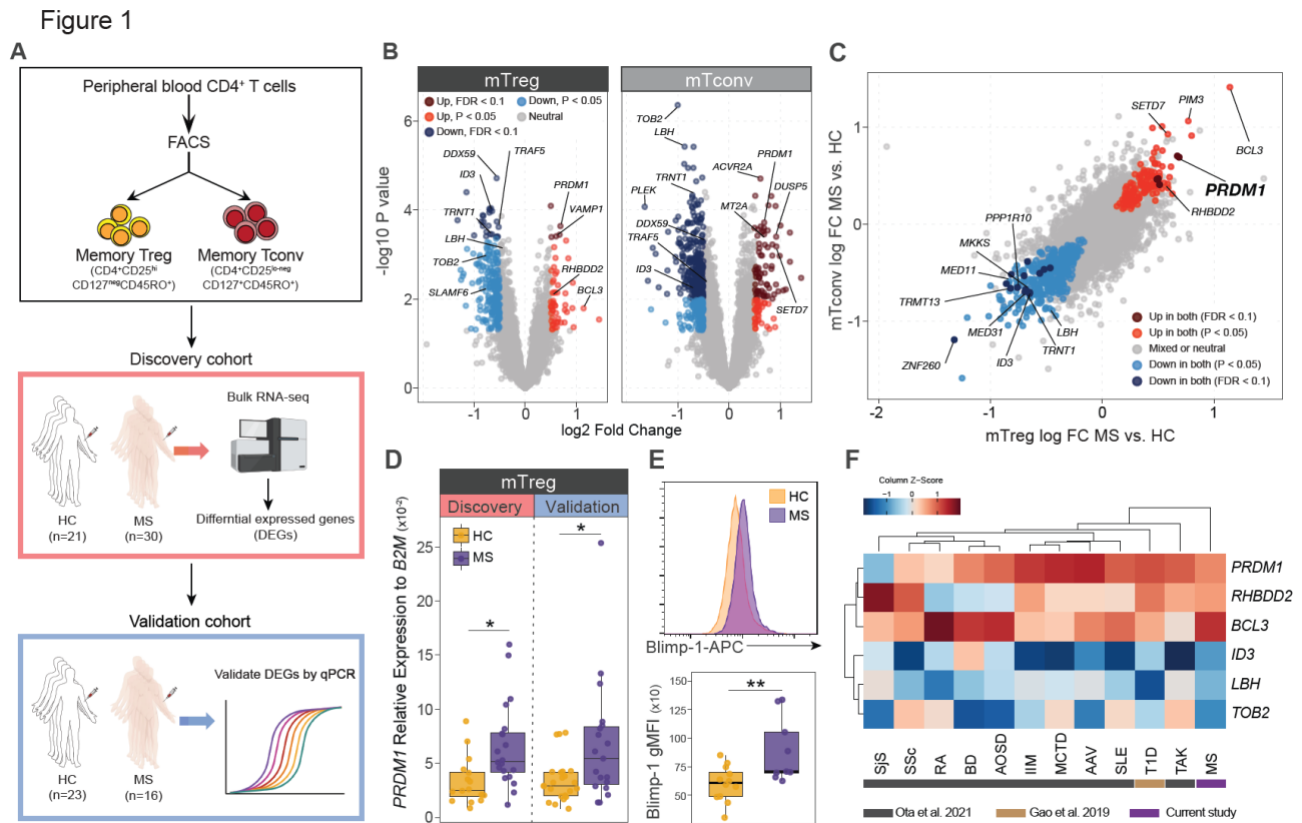


Figure 1 Deep transcriptomic analysis of memory Treg and Tconv highlight *PRDM1* upregulation in MS. (A) Schematic of study design. Memory Tconv (mTconv) or memory Tregs (mTreg) were FACS sorted from peripheral blood CD4⁺ T cells from patients with MS and healthy control subjects (HC). Bulk RNA-seq was performed on the discovery cohort and DEGs were identified (HC: n=20, MS: n=26). The selected DEGs were validated by an independent validation cohort (HC: n=23, MS: n=16). (B) Volcano plots showing DEGs for mTreg and mTconv between MS and HC. (C) Overlapped DEGs between mTreg and mTconv. (D) qPCR validation for *PRDM1* expression in both discovery and validation cohorts. (E) Protein validation for Blimp1 expression using flow cytometry (HC; n=12, MS; n=9). (F) Heatmap depicting expression patterns of selected six genes in mTreg/Fr2 eTreg across 12 autoimmune diseases (data were extracted from M. Ota *et al.* and P. Gao *et al.*). P* < 0.05, P** < 0.01; Statistical significance computed by unpaired t-test (D, E).

140
 141 **Single-cell dual omics analysis reveals elevated *PRDM1* in Th17-like Treg in MS. (Figure 2)**
 142 To gain a deeper understanding of cellular heterogeneity and novel cell types underlying
 143 disease mechanisms, we designed and performed single-cell RNA-sequencing (scRNA-seq) to
 144 profile CD4⁺ T cells. To overcome the sparsity of scRNA-seq data^{29,30}, we used Cellular Indexing
 145 of Transcriptomes and Epitopes by sequencing (CITE-seq)^{31,32}, and profiled 44 surface protein
 146 markers (Supplementary Table 2) and mRNA expression simultaneously in total CD4⁺ T cells
 147 enriched with Treg cells. Additionally, to avoid experimental batch effects between controls and

148 MS, we adopted hashing technology to pool cells into a single run of the 10x Genomics platform.
149 Across five experimental batches, we obtained a comparable number of Treg and Tconv cells
150 (Figure S2A). Demographic backgrounds (age, sex, and ethnicity) are controlled in each batch of
151 10x Genomics processing between controls and MS, and a total of five paired MS and control
152 samples were included (Supplementary Table 1).

153 We established cell identities based on surface marker proteins, eliciting prior knowledge
154 of cell-type-specific signatures. We first built a data matrix of 25,267 features (proteins and genes)
155 and 36,983 cells, combining five batches of CITE-seq profiles. We conducted a basic quality control
156 procedure to remove low-quality cells (Figure S2B) followed by batch normalization to control
157 batch-specific bias (Methods). We annotated cell type identities in semi-supervised training guided
158 by combinations of surface markers (Figure 2A), clearly demonstrating distinctive cell populations
159 in both the protein and transcriptomic space as can be seen in the gene-expression-based
160 clustering patterns in the UMAP visualization (Figure 2B). We confirmed that cell type annotation
161 results were not affected by batch labels or disease groups (Figure S2C). Our experimental design
162 also consistently enriched rare Treg populations and provided sufficient cells to study the variation
163 within Tregs (Figure S2D). The markers for Treg vs Tconv (*FOXP3*, *IKZF2*, *IL2RA/CD25*, and
164 *IL7R/CD127*), and memory vs naive (*CD45RA*, *CD45RO*) clearly distinguished four CD4⁺ T cell
165 subpopulations (Figure 2C, Figure S2E).

166 For differentially expressed gene³³ analysis between MS and control subjects at the pseudo-
167 bulk level³⁴, we quantified subject-level gene expression profiles after adjusting contributions of
168 putative confounding factors, which stem from unmeasured technical covariates but may create
169 spurious associations with disease status and cell types (see Methods for details)³⁵. We found that
170 single-cell-based analysis identified 90 up-regulated genes, and 16 down-regulated genes in MS
171 mTreg (Figure S2F). We compared the DEG effect size calculated in the scRNA-seq data (y-axis)
172 with the bulk DEG results (x-axis) (Figure 2D and E), which confirmed that a substantial number of
173 DEGs found in the bulk RNA-seq analysis are replicated in the matched subpopulations (mTreg
174 and mTconv) with a statistically significant correlation in both mTreg and mTconv (mTreg; $r = 0.28$,
175 $p=2^{-178}$, mTconv; $r = 0.25$, $p=2^{-149}$). For example, upregulation of *PRDM1* and downregulation of
176 *LBH* were demonstrated in MS mTreg (Figure 2D). *PRDM1* was also upregulated in mTconv, which
177 further validated the importance of *PRDM1* in the MS T cell signature (Figure 2D and E, Figure
178 S2G). *DDIT4*, which suppresses mTOR function, was upregulated in mTreg as well as mTconv
179 and downregulation of *TRAF3IP3* in MS mTreg, supporting the dysfunctional Treg property and
180 skewed Th17 signature in MS³⁶⁻³⁹. Of note, expression of *CD45RO* and *FOXP3* was not

181 significantly altered between MS and control subjects (Figure S2G), indicating our differential
 182 analysis based on dual omics single-cell analysis was not biased by skewed memory differentiation
 183 or significant loss of *FOXP3* gene expression in MS mTreg.

184 To elucidate the plasticity of the memory CD4⁺ T cell population, we further determined the T
 185 helper cell subtypes (Th1, Th2, Th17, Th1/17) in mTreg and mTconv by using key marker
 186 expression with CITE-seq (Figure 2F, G, Figure S3A-E, see also Methods). We then analyzed the
 187 gene expression changes between MS and controls at each mTreg subtype and found that *PRDM1*
 188 was significantly upregulated in Th17-like mTreg (mTreg17) (Figure 2H). Upregulation of CD226

Figure 2

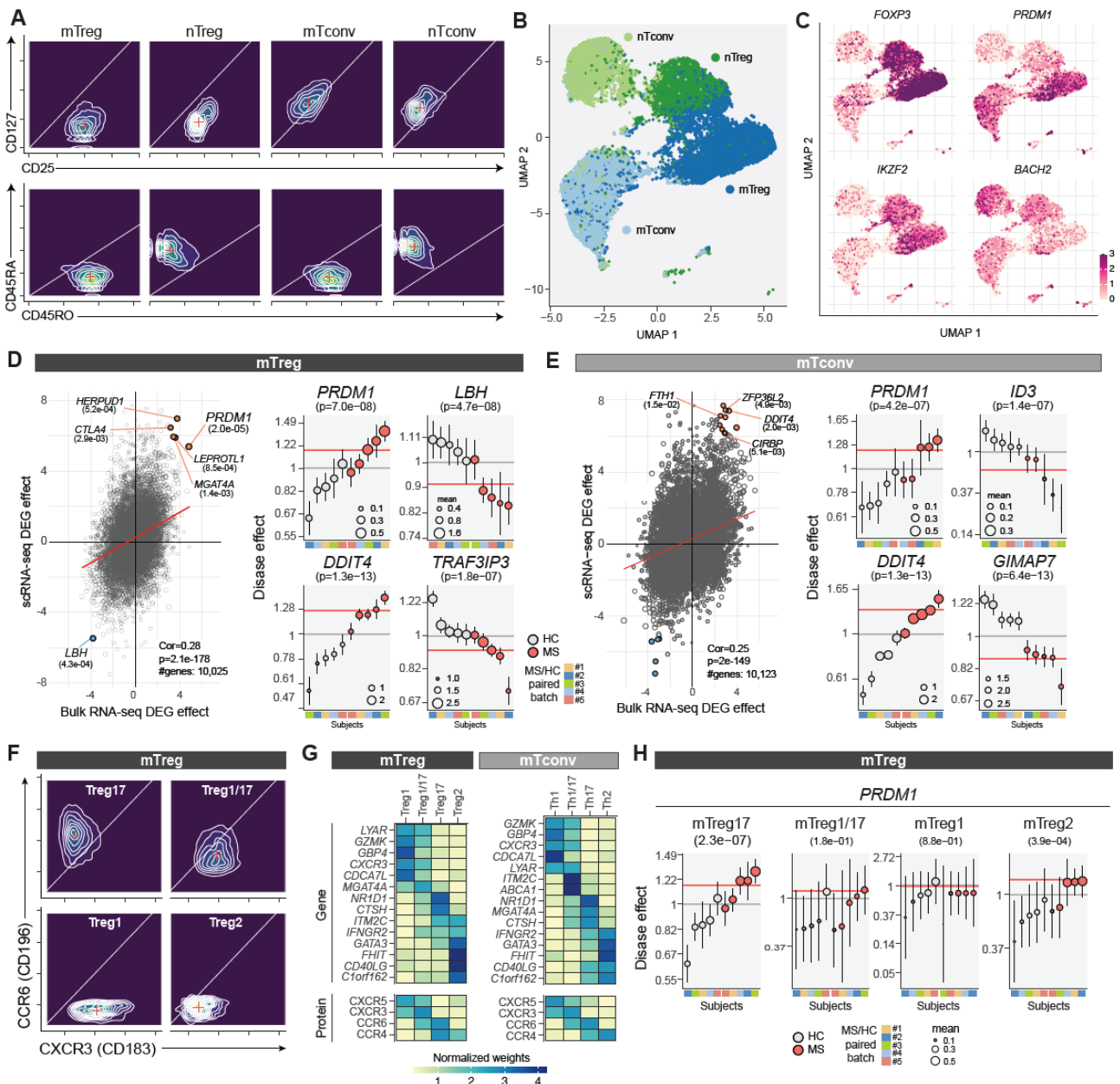


Figure 2 Single-cell dual omics analysis reveals elevated *PRDM1* in Th17-like Treg in MS. (A) Surface protein guided CD4⁺ T cells subtype annotation. Four CD4⁺ T cell subtypes were distinguished by CD25, CD127, CD45RO, and CD45RA protein expression. **(B)** UMAP based on gene expression for CD4⁺ T cells demonstrating decent overlap with protein-based subtype annotation. **(C)** *FOXP3*, *IKZF2*, *PRDM1*, and *BACH2* expressions on UMAP (all cells passed QC are plotted). **(D)** Combined differential analysis for bulk- and scRNA-seq in mTreg. Representative differentially expressed genes with pseudo-bulk analysis with scRNA-seq are shown. Gene expression changes in MS relative to control with indicated genes are computed at the single-cell level (see detailed in Methods). The size of dots is scaled proportionally to the average number of mRNA reads quantified within each batch and condition; the y-axis (disease effect) shows average gene expression after adjusted by confounding factors by matching; the error bars capture one standard deviation for the disease effects in Bayesian inference. Experimental batches for paired HC and MS samples are color coded (#1-5). **(E)** Combined differential analysis for bulk- and scRNA-seq in mTconv. **(F)** Sub-cell type analysis based on CITE-seq. Surface CXCR3 (CD183) and CCR6 (CD196) protein expressions of four subtypes are shown (log₁₀ scale). **(G)** Heatmaps showing the marker genes and proteins to define subtypes for each mTconv and mTreg. **(H)** The changes of *PRDM1* expression at subtype-level analysis in mTreg between MS and control subjects.

189 and *CD6* in mTreg17 and *CD278* (ICOS) and *JUNB* in Th17 indicates the skewed pathogenic Th17-
190 like signature in MS memory T cells (Figure S3F)⁴⁰⁻⁴³. Given the causative role of Th17 in MS
191 pathophysiology, our results at single-cell resolution further dissected the features of MS mTreg
192 and highlighted the potent role of *PRDM1* in skewing mTreg and mTconv toward the pathogenic
193 Th17-like signature in the context of MS.

194

195 ***Elevated alternative short PRDM1 isoform in MS mTreg. (Figure 3)***

196 Our unbiased transcriptomic profiling of MS mTreg using bulk and single-cell RNA-seq
197 identified *PRDM1* as a top candidate transcription factor accounting for dysfunctional Treg
198 properties in MS. *PRDM1* encodes the Blimp1 protein which functions as a zinc finger
199 transcriptional repressor initially identified as a protein that binds to the promoter of *IFNB1* and
200 suppresses its expression⁴⁴. The development and function of a variety of immune cells are also
201 under the control of Blimp1. CD4⁺ T cell-specific *PRDM1* deletion exaggerates the proinflammatory
202 reaction in multiple murine models of autoimmune diseases, including EAE^{45,46}; in contrast,
203 previous studies demonstrated Blimp1 as an essential factor driving inflammatory programs in
204 Th17 differentiation⁴⁷. This contradictory evidence suggests the context-dependent roles of *PRDM1*
205 in CD4⁺ T cells in mice. Recently, the role of *PRDM1* in Treg cells was studied by using Treg-
206 specific *PRDM1* knockout mice in the context of autoimmunity. Loss of their suppressive capacity
207 in *PRDM1*-deficient Treg indicates that *PRDM1* plays a critical role in maintaining Treg
208 homeostasis in tissue and positively regulates its suppressive function⁴⁸⁻⁵⁰. Thus, our result
209 showing upregulation of *PRDM1* in MS mTreg fundamentally contradicted observations in mice.

210 In humans, it has been appreciated that *PRDM1* has two major isoforms: the original full-
211 length isoform and another short isoform generated by alternative promoter usage⁵¹. The short
212 *PRDM1* isoform (*PRDM1-S*; encoding a short form of Blimp1, Blimp1-S) arose during dry-nosed
213 primate evolution, and thus is not coded in the mouse genome. Blimp1-S lacks the N-terminal
214 region of Blimp1, which results in missing a part of the PR domain that is important in mediating
215 interaction with co-factors of Blimp1 (Figure 3A). Thus, Blimp1-S is implicated as a dominant
216 negative form against the full-length Blimp1^{51,52}. Chromatin accessibility analysis at the *PRDM1*
217 locus showed that the *PRDM1-S* promoter region is significantly accessible in human Treg and our
218 previous HiDRA analysis⁵³ revealed stronger activity at the promoter of *PRDM1-S* than that of full-
219 length *PRDM1* isoform (*PRDM1-L*), which is consistent with DNase I hypersensitive site (DHS)
220 data of human primary total Treg⁵⁴ and T cells⁵⁵ (Figure 3B). In addition, bulk RNA-seq with nine
221 different peripheral blood human immune cells confirmed that both short and long *PRDM1* isoforms
222 were expressed in a cell-type-dependent manner. More specifically, monocytes and B cells mainly
223 express *PRDM1-L*. In contrast, NK cells dominantly express *PRDM1-S*, and CD4⁺ T cells (including
224 Treg) and CD8⁺ T cells express more *PRDM1-S* than *PRDM1-L*, especially in the memory
225 population, suggesting cell type-specific roles for each isoform (Figure 3C, S4A). This cell-type
226 specific expression patterns of *PRDM1-S* and *-L* were further validated at protein level by western
227 blot (Figure 3D, S4B). There are three bands detected between 80-110 kD consistent with previous
228 studies⁵⁶, though the band sizes are slightly higher in our data. To confirm the band size of Blimp1-
229 S and Blimp1 protein, we overexpressed the open reading frame of Blimp1-S or Blimp1 in 293T
230 cells by lentiviral transduction and found that lower two bands are derived from Blimp1-S while the
231 higher molecular weight band is from Blimp1 (Figure S4B). We then analyzed our mTreg bulk RNA-
232 seq at transcript level and found that *PRDM1-S* is significantly increased in MS mTregs compared
233 to healthy controls, though there is a moderate difference for *PRDM1-L* expression between MS
234 and control subjects (Figure 3E). This alteration was further validated by qPCR in both discovery
235 and validation cohorts (Figure S4C). Notably, this upregulation of *PRDM1-S* expression was also
236 observed in SLE Tregs from ImmuNexUT data²⁷ (Figure S4D). These data suggest that *PRDM1-S*
237 is a key regulator of Treg function in autoimmune diseases. These lines of data prompted us to
238 hypothesize that the aberrant induction of *PRDM1-S* may confer dysfunctional properties to MS
239 Tregs. Given that Blimp1-S can serve as a dominant negative isoform against conventional Blimp1,
240 we first examined whether the ratio of *PRDM1-S* and *PRDM1-L* is altered between MS vs HC.
241 However, there was no significant difference observed in this ratio with either bulk RNA-seq or
242 qPCR data, suggesting that the balance between *PRDM1-S* and *PRDM1-L* was not significantly
243 disrupted in MS Tregs (Figure S4E). Although *PRDM1-S* level was significantly upregulated in MS

244 Tregs, *PRDM1-L* levels were not changed or slightly increased in MS Treg (Figure 3E, S4C), thus
 245 *PRDM1-S* mediated effects in MS Tregs could be independent from its dominant negative function
 246 against *PRDM1-L*.

247 In order to identify genes that are differentially correlated between *PRDM1-S* and *PRDM1-L*,
 248 we further analyzed the co-expression pattern of immune-related genes with *PRDM1-S* and
 249 *PRDM1-L* in mTreg from HC and MS (Figure 3F); genes associated with effector and tissue
 250 resident Treg signature (i.e. *BATF*, *CCR8*, *ICOS*, *CD69*, *IL1RL1*, *AREG*, *IRF4*) were co-expressed
 251 with both *PRDM1-L* and *PRDM1-S*, particularly in MS Tregs, which reflects the skewing feature of
 252 MS mTreg towards effector/tissue resident properties. Differentially expressed genes in MS
 253 mTregs, such as *BCL3*, were positively correlated with both *PRDM1* isoforms in MS but negatively
 254 correlated in healthy controls. Of note, these positive correlations in MS are stronger for *PRDM1-S*
 255 compared to *PRDM1-L*. This type of correlation pattern (positive correlation for MS but negative
 256 for control Tregs and higher correlation with *PRDM1-S* than *PRDM1-L*: Box 1 and 2 in Figure 3F)
 257 was observed for the following genes; *IRF1*, which acts as a negative regulator for Foxp3
 258 expression⁵⁷; *BATF* and *FOSL2*, which belong to AP-1 family, with crucial functions for Treg
 259 differentiation and maintenance^{58,59}; *SGK1*, which is known to disrupt Treg homeostasis^{14,15} and

Figure 3

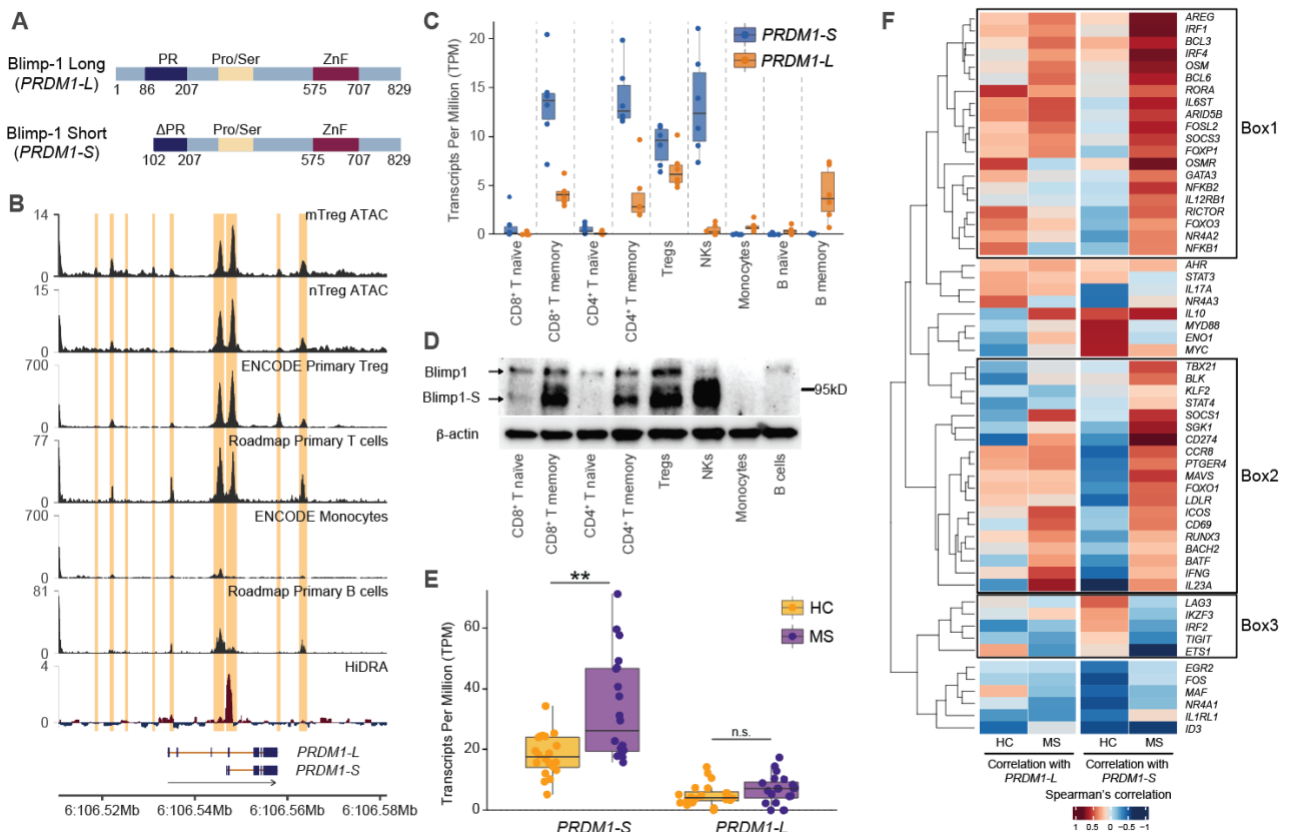


Figure 3 Elevated alternative short PRDM1 isoform in MS mTreg. (A) Schematic of *PRDM1* short and long isoforms. PR; PR domain, Pro/Ser; Proline/serine rich region, ZnF; five C2H2 zinc fingers. (B) ATAC-seq (mTreg and nTreg), DHS (ENCODE primary Treg, Roadmap primary T cells, ENCODE monocytes, Roadmap Primary B cells), and HiDRA peaks at *PRDM1* locus. (C) *PRDM1-S* and *PRDM1-L* isoform expression across 9 different immune cell types in peripheral blood assessed by bulk RNA-seq (n=6). (D) Western blot analysis of Blimp1 expression from 8 different immune cell types in peripheral blood. Conventional Blimp1 and alternative Blimp1-S are distinguished by different size. (E) *PRDM1-S* and *PRDM1-L* gene expression assessed by bulk RNA-seq in mTreg between MS and HC. $P^{**} < 0.01$; Statistical significance computed by one way ANOVA with Dunn's multiple comparisons tests. (F) Heatmaps depicting the Spearman's correlation for curated immune related genes with *PRDM1-S* or *PRDM1-L* in HC or MS.

260 play pathogenic roles in EAE⁶⁰. We also noted that suppressive molecules enhancing Treg function
261 (i.e. *IKZF4*, *TIGIT*, *LAG3*, *ID3*) are negatively correlated with *PRDM1-S* in MS but not in healthy
262 controls (Box 3 in Figure 3F), supporting the association between *PRDM1-S* and dysfunctional
263 Treg properties in MS. Taken together, we identified *PRDM1-S* as a significantly upregulated
264 alternative isoform that partially accounts for the dysfunctional gene signature of MS mTreg.

265 **Short PRDM1 induces SGK1 and Treg dysfunction. (Figure 4)**

266 Next, we hypothesized that *PRDM1-S* has a unique feature independent of interfering from
267 *PRDM1-L* function. To test this hypothesis, we transduced *PRDM1-S* or *PRDM1-L* into primary
268 human Tregs by using a lentivirus-based overexpression (OE) system. Isolated human primary
269 Tregs were infected with lentivirus encoding *PRDM1-S* or *PRDM1-L* with GFP reporter and GFP
270 positive transduced cells were sorted by FACS after four days of culture. Overexpression for each
271 transcript was confirmed by qPCR (Figure S5A). Since only three amino acids are unique to
272 *PRDM1-S* (Blimp1-S) coding sequence compared to *PRDM1-L* (Blimp1), overexpression of
273 Blimp1-S open reading frame cannot be detected by qPCR, because qPCR primers are targeted
274 to unique sequence of 5' UTR region of *PRDM1-S*. Instead, we confirmed upregulation of total
275 *PRDM1* but not *PRDM1-L* with Blimp1-S OE by qPCR (Figure S5A). Protein level induction of
276 Blimp1-S with this system was confirmed by western blot (Figure S4B). Bulk RNA-seq was
277 performed on sorted GFP positive cells and highlighted 100 genes exhibiting nominal evidence of
278 differential expression ($|\log_2FC| > 0.6$, P value < 0.05) between *PRDM1-S* overexpression and
279 GFP control (Figure 4A). We found that *SGK1* was one of the major upregulated genes induced by
280 *PRDM1-S* but not by *PRDM1-L* overexpression (Figure 4A, S5B). We validated this observation by
281 qPCR in human primary Treg cells and Jurkat T cells (Figure 4B). Moreover, *SGK1* was
282 upregulated in our scRNA-seq analyses, particularly in the Th17-like Treg subset together with
283 *PRDM1*, consistent with the role of SGK1 skewing pathogenic Th17 like signature in Tregs⁶⁰ (Figure
284 4C, S5C). This *PRDM1-SGK1* axis was a common feature among other autoimmune diseases in
285 a published dataset²⁷, where both *PRDM1* and *SGK1* were significantly increased in Tregs from

286 patients with SLE and ANCA-associated vasculitis (Figure S1E, S5D). Finally, we examined the
 287 impact of *PRDM1-S* on Treg function by *in vitro* Treg suppression assay and found that Tregs with
 288 *PRDM1-S* overexpression (OE) exhibited lower suppressive function than control GFP OE Treg,
 289 strongly indicating that aberrant *PRDM1-S* expression causes Treg dysfunction (Figure 4D).
 290 *PRDM1-S* OE decreased the level of both full-length Foxp3 and the exon 2-containing suppressive
 291 isoform of Foxp3 protein, further confirming the impaired suppressive function on Tregs with
 292 *PRDM1-S* OE (Figure 4E, S5E). These data support the unique role of *PRDM1-S* as a positive
 293 regulator for *SGK1* in human Treg cells, especially in the subset of potentially pathogenic Th17-
 294 like Treg, where *SGK1* plays a proinflammatory role under high sodium conditions and is
 295 responsible for pathogenic features in a murine MS model via suppressing Foxp3 expression and
 296 stabilization^{14,15}. Thus, these data suggest that the *PRDM1-S/SGK1* axis underlies Treg dysfunction
 297 in MS.

Figure 4

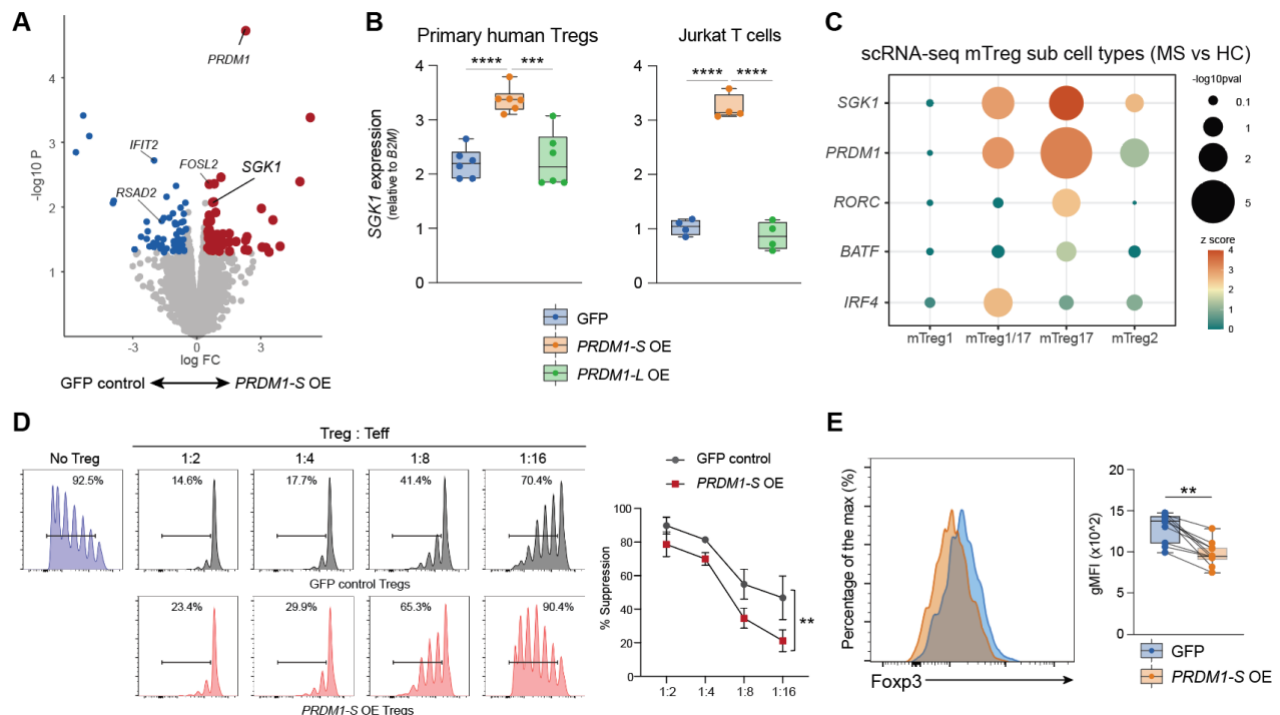


Figure 4 Short *PRDM1* induces *SGK1* and Treg dysfunction. (A) Volcano plot showing statistical significance and fold change for genes differentially expressed by *PRDM1-S* overexpression in primary mTreg. (B) *SGK1* expression was assessed by qPCR for overexpression of *PRDM1-S* and *PRDM1-L* with primary human mTregs (left) and Jurkat T cells (right). $P^{***} < 0.001$, $P^{****} < 0.0001$; Statistical significance computed by one way ANOVA with Dunn's multiple comparisons tests. (C) Pseudo-bulk analysis of *SGK1*, *PRDM1*, *RORC*, *IRF4* and *BATF* expression in scRNA-seq at each of four major mTreg subtypes. (D) *In vitro* Treg suppression assay with human primary Tregs. T effector cell proliferation was assessed after 5 days of co-culture with Tregs transfected with GFP control vector vs *PRDM1-S* OE vector. $P^{**} < 0.01$; Statistical significance computed by two-way repeated measures ANOVA. (E) Flow cytometry analysis for Foxp3 in primary Treg cells by overexpression of *PRDM1-S* compared to GFP control (n=10). $P^{**} < 0.01$; Statistical significance computed by paired t test.

298 **Comprehensive analysis for chromatin accessibility reveals AP-1 and IRF enrichment in MS**
299 **mTreg. (Figure 5)**

300 To further elucidate the regulatory mechanisms underlying the dysfunctional properties of MS
301 mTregs, we performed Assay for Transposase Accessible Chromatin sequencing (ATAC-seq) to
302 identify chromatin-accessible signatures and epigenetic regulatory elements. We first determined
303 the change of mTreg chromatin accessibility between MS and healthy controls. To our surprise,
304 there was no significant difference in genome-wide chromatin accessibility between MS and
305 healthy controls in mTreg (FDR<0.05), suggesting that global chromatin accessibility is not per se
306 the major factor regulating gene expression in MS mTreg. We hypothesized that differential binding
307 of regulatory TFs may account for the changes in mTreg gene expression between MS and control
308 subjects. To identify TFs that potentially drive the observed gene expression signature in MS mTreg,
309 we analyzed the enrichment of TF motifs and TF footprints within accessible regions between MS
310 and healthy controls in mTreg (Figure 5A). We observed an enrichment of AP-1 and IRFs TF motifs
311 that are important for CD4⁺ T cell activation and differentiation in MS mTreg (Figure 5B).

312 To gain deeper insight into captured accessible regulatory elements, we performed a
313 differential footprint analysis on ATAC-seq peaks by using HINT (H_{mm}-based I_{de}N_{tification} of
314 T_ranscription factor footprints)⁶¹ and TOBIAS⁶². Consistent with our motif analysis, footprint
315 analysis demonstrated the enrichment of AP-1 family TFs and IRFs in mTregs from MS compared
316 to control subjects⁶³⁻⁶⁶ (Figure 5C). AP-1 transcriptional activity is negatively regulated by direct
317 interaction with Foxp3⁶⁷ and AP-1 is postulated to serve as a pioneer factor at Treg-specific
318 regulatory elements where Foxp3 subsequently replaces it to establish Treg-specific enhancer
319 architecture and DNA methylation^{68,69}. Our observation of AP-1 enrichment in dysfunctional Tregs
320 in MS could thus stem from the lower or impaired activity of Foxp3 in MS mTreg and reflect a more
321 effector Tconv-like mTreg function in patients with MS^{70,71}.

322 IRF family TFs share a common DNA binding sequence (IRF binding element). Of interest,
323 IRF-1 and IRF-2, but not IRF-4 nor IRF-8, are known to compete with evolutionally conserved
324 *PRDM1-L*^{72,73}. IRF-1 plays a critical role in Treg differentiation and maintenance^{57,74} and it was of
325 interest that *IRF1* was more co-expressed with *PRDM1-S* as compared to *PRDM1-L* in MS mTreg
326 (Figure 3F). In addition, the enrichment of IRF-1 TF motif and footprint in MS mTreg (Figure 5B, C)
327 could reflect the disrupted *PRDM1-L*-mediated gene regulation in MS mTreg. These results
328 suggest that *PRDM1-L* mediated gene regulation could be disrupted in the context of MS (Figure
329 S6A). To test this hypothesis, we first defined the set of genes that are specifically regulated by
330 *PRDM1-L* by performing *PRDM1-L*-specific gene knockdown in human primary Tregs. A total of

331 1566 differentially expressed genes (defined as $|\log_2FC| > 1$, $FDR < 0.05$; 753 upregulated and 813
332 downregulated genes) were identified and defined as “*PRDM1-L* signature genes” (Figure S6A and
333 B). We performed single-cell level correlation analysis with our scRNA-seq data using a scalable
334 negative binomial mixed model, NEBULA⁷⁵. We then investigated putative pathogenic downstream
335 effects of *PRDM1* expressions in single-cell analysis, examining whether the *PRDM1-L* signature
336 genes were perturbed by *PRDM1* genes between the mTreg cell groups derived from the healthy
337 and MS samples. We first ranked genes according to the correlation with single-cell *PRDM1*
338 expression levels in two independent cell groups derived from the healthy and MS samples,
339 respectively (thus, two lists of genes--one for the healthy and the MS). At a p-value threshold (the
340 x-axes of Figure S6C), we investigated whether the number of genes in the top lists were from the
341 *PRDM1-L* signature genes. So as not to be biased by a fixed p-value threshold, we conducted
342 bootstrapped enrichment tests for each p-value threshold. We found that the *PRDM1-L* signature
343 genes were significantly depleted in the top list of the genes derived from the MS mTreg group,
344 suggesting that the interactions between *PRDM1* and the target genes were more frequently
345 disrupted in the disease cells. We further explored the isoform level transcriptional co-regulation
346 by using our bulk RNA-seq data. Approximately half of positively correlated *PRDM1-L* signature
347 genes in Tregs from healthy controls lost their positive correlation in MS Tregs (Figure S6D,
348 Supplementary Table 3). Given that *PRDM1-L* plays a crucial role in maintaining Treg function⁴⁸⁻
349⁵⁰, we next assessed the correlation between *PRDM1-L* and Treg signature genes^{76,77}. As observed
350 with *PRDM1-L* signature genes, approximately half of the positively correlated Treg signature
351 genes in healthy controls lost their positive correlation in MS Tregs (Figure S6E). Taken together,
352 our ATAC-seq results revealed that while genome-wide chromatin accessibility was not significantly
353 altered, differential binding of TFs (AP-1 and IRFs) to regulatory elements could serve as key
354 upstream regulatory factors and drive dysfunctional Treg gene programs by possibly disturbing
355 *PRDM1-L* mediated gene regulation in patients with MS.

356

Figure 5

357

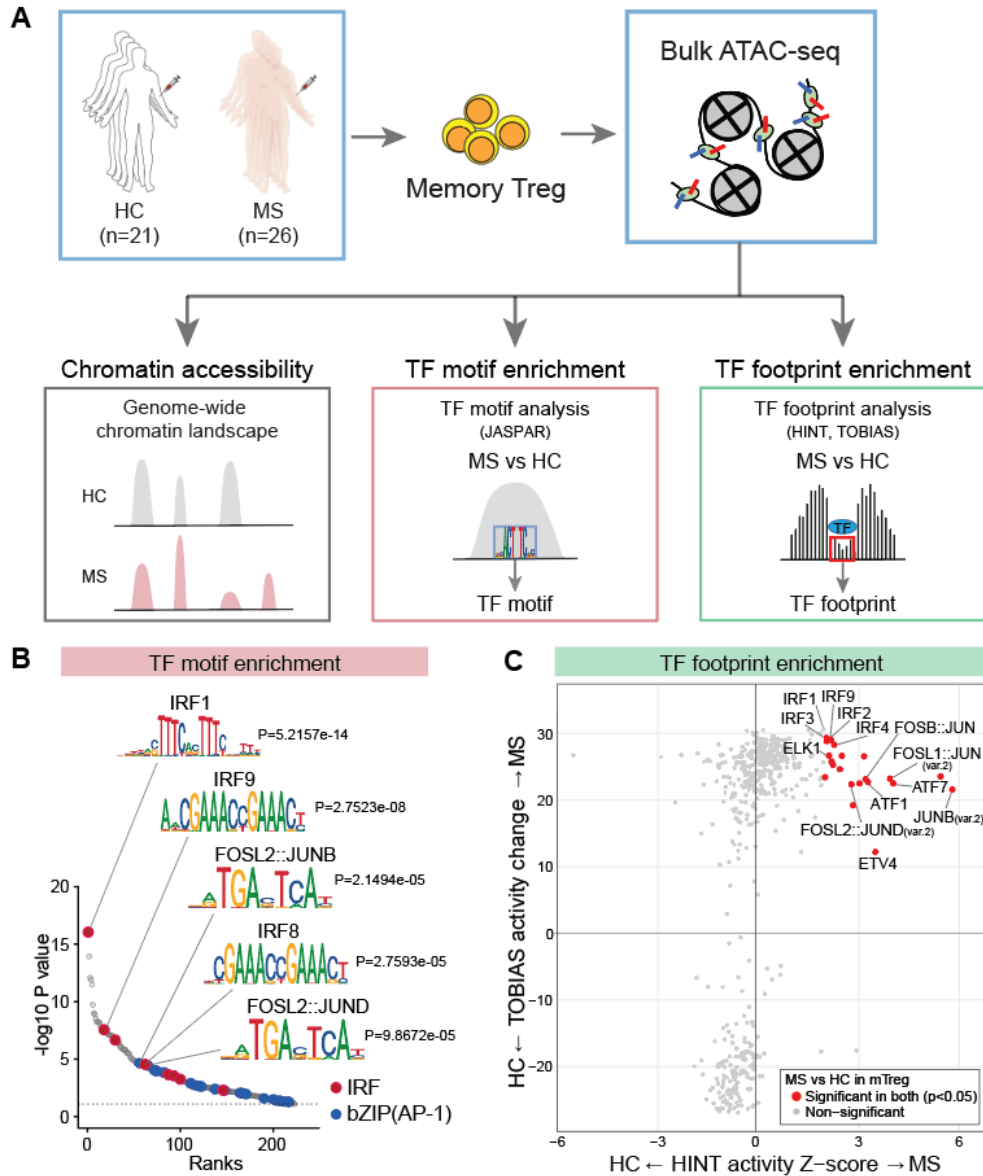


Figure 5 AP-1 and IRF TF bindings are enriched in MS mTreg. (A) Schematic of ATAC-seq experiments for mTreg from MS (n=26) and healthy control (n=21). (B) TF motif enrichment analysis in mTreg between MS and HC by HINT. IRF and AP-1 motifs are significantly enriched in MS mTreg. (C) TF footprint enrichment analysis in mTreg between MS and HC by HINT and TOBIAS. IRF and AP-1 footprints are significantly enriched in MS mTreg.

358 **Identification of active enhancer for Short PRDM1 in human T cells. (Figure 6)**

359 We sought to determine the regulatory mechanisms that induce upregulation of *PRDM1-S* as
 360 opposed to *PRDM1-L* in human mTreg. We reasoned that the enriched binding of AP-1 and IRFs
 361 exert its regulatory function through binding to *cis*-regulatory elements for *PRDM1*, especially
 362 *PRDM1-S*. To identify the *cis*-regulatory elements for *PRDM1-S*, we prioritized the accessible
 363 chromatin regions surrounding the *PRDM1* locus (\pm 0.5MB window around the transcriptional start

364 site of *PRDM1*) and examined the association between *PRDM1* expression and chromatin
365 accessibility in our human primary T cell ATAC-seq and RNA-seq data. Twenty significant
366 accessible regions ($p < 0.001$) were identified as potential regulatory elements associated with
367 *PRDM1* expression (Figure 6A). The majority of these peaks overlapped with H3K27ac ChIP-seq
368 signals for primary human Tregs⁵⁵, nominating them as potential enhancers. To functionally
369 validate these accessible chromatin regions, we adopted the CRISPR activation (CRISPRa)
370 system in Jurkat T cells that stably expresses catalytically inactive Cas9 fused to the transcriptional
371 activator VP64 (dCas9-VP64)⁷⁸ (Figure 6A).

372 First, we confirmed that *PRDM1-S* and *PRDM1-L* are independently regulated through
373 different promoter activity by our CRISPRa method with sgRNAs targeting each promoter region,
374 though there may be interactions between the transcription start site (TSS) of *PRDM1-L* and
375 *PRDM1-S* (Figure S7A). Next, we designed three sgRNAs for each accessible region and
376 generated sgRNA expressing lentiviral particles for twenty candidate regulatory elements (coded
377 as #1 to #20) (Supplementary Table 4, Figure 6A). dCas9-VP64 expressing Jurkat T cells were
378 infected by lentivirus encoding each sgRNA, then the double positive cells for GFP (dCas9-VP64)
379 and RFP (sgRNA) were sorted by FACS (Figure 6A). We observed that sgRNAs targeting the #2
380 peak region (-339,554 bp upstream of the *PRDM1-L* TSS) mediated a unique induction of *PRDM1-S*
381 but not *PRDM1-L* compared to control sgRNAs (Figure 6B). This #2 peak region is reported as
382 one of the “double elite” regulatory elements for *PRDM1* in the GeneHancer dataset⁷⁹, which
383 reflects a higher likelihood of prediction accuracy for both enhancer and target gene (Figure 6B;
384 top). The enhancer function of the #2 peak *cis*-regulatory element was further validated by
385 independent experiments with not only CRISPRa but also CRISPRi (Figure 6C, S7B).

386 To further clarify the function of this region as a *cis*-regulatory element at *ex vivo* human
387 primary Tregs, we sought to decode the histone modification. The main obstacle in the investigation
388 of chromatin state by using conventional ChIP-seq technique is the limited number of *ex vivo*
389 primary Tregs available for determining multiple histone marks from the same sample. Recent
390 development of multiplexed, indexed T7 ChIP-seq (Mint-ChIP)⁸⁰ allows us to identify both active
391 and repressive epigenetic marks by using histone modification-specific antibodies with limited cell
392 numbers. We performed Mint-ChIP on human primary Tregs in collaboration with the ENCODE
393 project. Histone modifications (H3K27ac, H3K4me1, and H3K4me3) distinguishing active
394 enhancers were assessed on *ex vivo* human primary Treg from MS and control subjects. The #2
395 peak region was marked by H3K4me1 and H3K27ac but without H3K4me3, suggesting the #2
396 peak region functions as an active enhancer (Figure 6D; top). Importantly, this #2 peak region

397 overlaps with AP-1 family and IRF4 ChIP-seq peaks and contains AP-1/IRF composite motif⁸¹,
398 suggesting this enhancer element regulates *PRDM1-S* via AP-1 and IRF4 binding (Figure 6D;
399 bottom).

400 Given that BATF and IRF4 are known to bind in a cooperative fashion on the AP-1/IRF
401 composite motif facilitating Th17 differentiation^{63,82}, we examined the role of IRF4 and BATF on
402 *PRDM1-S* expression in Tregs by performing IRF4 and BATF knockdown experiments. Surprisingly,
403 *PRDM1-S* was upregulated by knocking down *IRF4* or *BATF* while in contrast, *PRDM1-L* was
404 downregulated by *IRF4* KD (Figure S7C). These data indicate that *IRF4* differentially regulates
405 *PRDM1-S* and *PRDM1-L* in human primary Tregs. Of note, *IRF4* or BATF KD did not affect *FOXP3*
406 expression, suggesting that the loss of IRF4 and BATF in human Tregs can induce *PRDM1-S* and
407 *SGK1* expression without significant reduction of *FOXP3* expression, further indicating that the
408 dysfunctional *PRDM1-S/SGK1* axis observed in MS Tregs is counter regulated by the core effector
409 Treg regulator IRF4 and BATF. These data also suggest that the *cis*-regulatory element (#2 peak
410 region identified as upstream of *PRDM1-S*, Figure 6) serves as a negative regulatory element for
411 *PRDM1-S* expression via IRF4/BATF binding. Indeed, CRISPRi targeting on this #2 peak region
412 suppress *PRDM1-S* expression (Figure 6C). Thus, these results highlight IRF4/BATF as potential
413 upstream transcription factors negatively regulating *PRDM1-S* and indicating that a newly identified
414 *cis*-regulatory element that contains the AP-1/IRF composite motif may account for aberrant
415 *PRDM1-S* induction in MS mTreg.

416

Figure 6

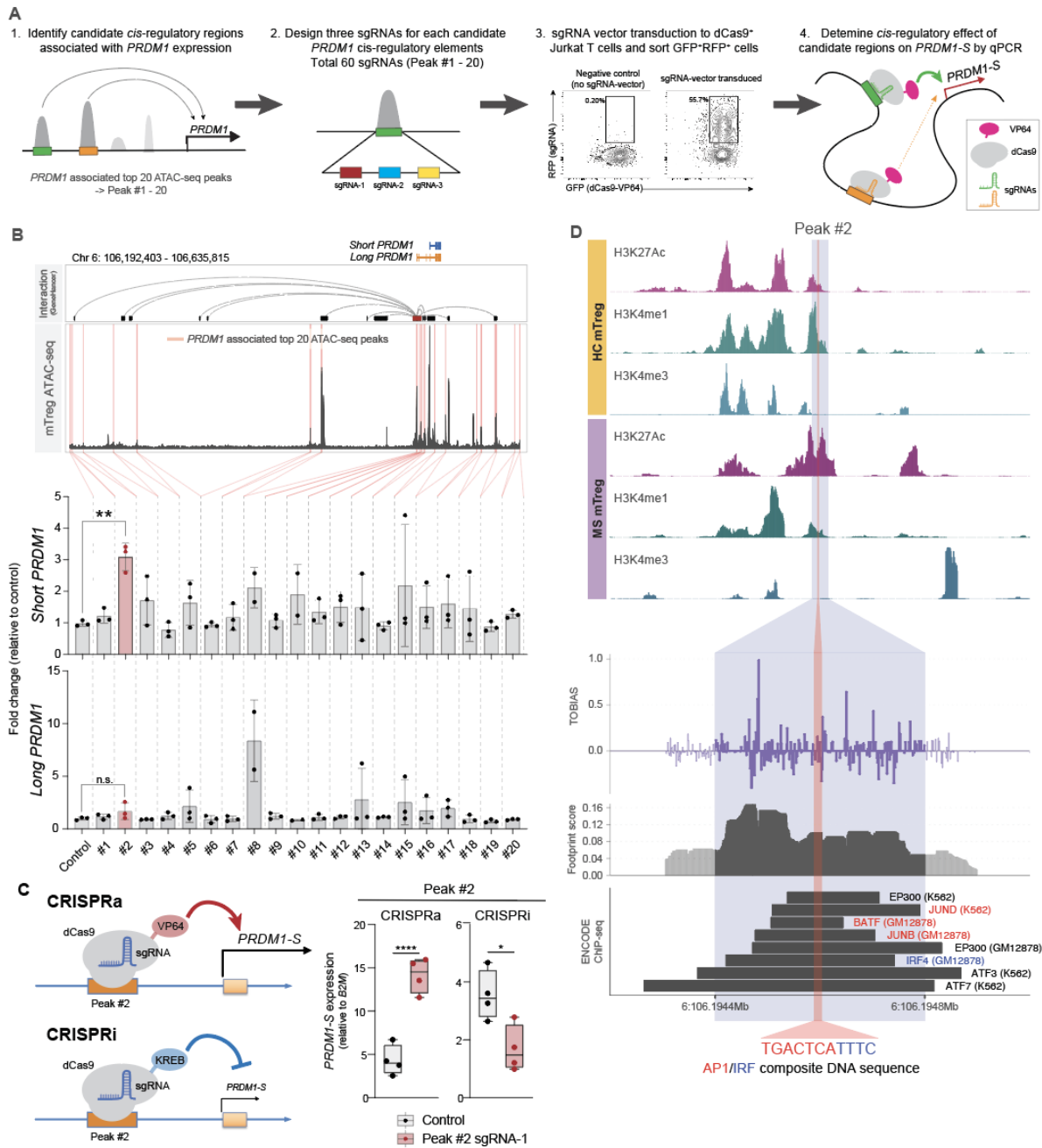


Figure 6 Active enhancer element for short *PRDM1* with AP-1 and IRF bindings. (A) Schematic experimental overview. 1. Identification of candidate *cis*-regulatory elements regulating *PRDM1* expression from ATAC-seq peaks. 2. CRISPRa based examination of *PRDM1*-S specific *cis*-regulatory elements. (B) CRISPRa validation for top 20 *PRDM1* associated regulatory elements. Top: Top 20 accessible chromatin elements that are associated with *PRDM1* expression are highlighted in red. Potential interactions of regulatory elements with *PRDM1* gene are analyzed by GeneHancer database and shown on the top. Middle and bottom; CRISPRa-induced expression of short *PRDM1* (middle) and Long *PRDM1* (bottom) were assessed by qPCR. Detailed information for all 20 regions is shown in Supplementary Table 4. (C) Validation CRISPRa and CRISPRi experiment for #2 peak sgRNA-1 (n=4). (D) Top: H3K27ac, H3K4me1, and H3K4me3 MINT-ChIP signal on the #2 peak region in mTreg from HC and MS. Four replicates of HC and two replicates of MS are merged into one representative track respectively. Middle: Footprint analysis on #2 peak region with TOBIAS footprint score. Bottom: AP-1 and IRF ChIP-seq signals identified on #2 peak region in ENCODE data are shown. AP-1/IRF composite motif identified in #2 peak region is highlighted. P* < 0.05, P** < 0.01, P*** < 0.0001; Statistical significance computed by one way ANOVA with Dunn's multiple comparisons tests.

418 Discussion

419 Disruption of peripheral CD4⁺ T cell homeostasis is a central component driving pathogenesis
420 of autoimmune disease where autoreactive T cells lose tolerance to self-antigen by both intrinsic
421 and extrinsic mechanisms. Treg-mediated surveillance is central for controlling activation of
422 autoreactive CD4⁺ T cells and dysfunctional Tregs are a hallmark of MS and other autoimmune
423 diseases^{7,8}. In addition, recent analysis of genome-wide association studies emphasizes the
424 substantial contribution of CD4⁺ T cells, including Tregs, as potentially causal mediators of
425 autoimmune disease³. Although several phenotypic changes have been identified in dysfunctional
426 Tregs, the underlying molecular mechanisms leading to breakdown of Treg suppressive function
427 in patients with autoimmune diseases are unknown. Here, by using MS as a model for studying the
428 molecular mechanisms of human Treg dysfunction, we examined transcriptional and epigenetic
429 alteration in human Tregs, identifying a previously unknown role of an alternative short *PRDM1*
430 isoform in dysfunctional Treg. *SGK1* was identified as a target of short *PRDM1*, which has been
431 reported to confer the pathogenic function of Treg in both human and mouse^{14,15,60}. Moreover, both
432 *PRDM1* and *SGK1* were upregulated in Tregs from different autoimmune diseases such as SLE
433 and ANCA-associated vasculitis²⁷, suggesting that the *PRDM1/SGK1* axis could serve as a
434 common feature of Treg dysfunction in the context of human autoimmunity. Finally, exploration of
435 epigenetic changes in MS Treg revealed an active enhancer element that induces short *PRDM1*
436 transcription, which was validated by CRISPRa/i experiments. AP-1 family, especially BATF, and
437 IRF4 directly bound to this regulatory element, implicating the role of these TFs in contributing to
438 *PRDM1-S* mediated Treg dysfunction in autoimmune disease. Thus, this study provides a novel
439 mechanistic insight of dysfunctional Tregs in the context of MS and potential therapeutic targets to
440 reverse Treg dysfunction in autoimmune diseases.

441 Previous studies exploring transcriptional alterations that occur in CD4⁺ T cells in patients
442 with MS as compared to control subjects did not identify significant differences^{4,5}. We hypothesized
443 that it was critical to examine CD4⁺ T cell subpopulations. Thus, we further segregated CD4⁺ T
444 cells into four subpopulations and performed transcriptional characterization using bulk RNA-seq
445 with deeper sequencing. This allowed us to detect significant differences in gene expression
446 between patients with MS and control subjects at both the gene and transcript level. Our bulk RNA-
447 seq based findings in the discovery cohort were further confirmed by two different means; (1)
448 utilization of a validation cohort with the same method as the discovery cohort, and (2) utilization
449 of CITE-seq to determine the subpopulations and assess the difference within the third cohort,
450 leading to highly reproducible findings. One caveat is that we were unable to quantitate *PRDM1*

451 isoform expression at single cell resolution in our 10x genomics dataset. Full length mRNA-
452 capturing scRNA-seq has the potential to differentiate between *PRDM1* isoforms; thus, this
453 technique could be applied to elucidate further insight at single-cell resolution⁸³. This approach will
454 also enable characterization of further differences in T cell transcriptomics between MS and healthy
455 subjects.

456 Our transcriptomic analysis of *PRDM1* isoforms identified expression patterns between
457 *PRDM1-S* and *PRDM1-L* that are highly cell type specific. Memory T cells and NK cells expressed
458 higher levels of *PRDM1-S* as compared to *PRDM1-L*, while monocytes/DCs, and B cells
459 preferentially express *PRDM1-L*. Of note, *PRDM1* expression across all B cell lineage (naïve,
460 unswitched memory, switched memory, double negative B cells, and plasmablasts) is strictly
461 limited to *PRDM1-L* over *PRDM1-S* (ImmuNexUT dataset). We also observed that the balance
462 between *PRDM1-S* and *-L* is dynamic and changes after T cell activation and cytokine milieu *in*
463 *vitro*. Thus, our findings demonstrate a tight linkage between *PRDM1* isoforms and immune
464 pathways related to AP-1/IRFs in autoimmune diseases and highlight a fundamental role in
465 primates for *PRDM1* isoform switching in regulating immune responses.

466 Epigenetic alterations in MS Tregs were assessed by using bulk ATAC-seq. Conventional
467 characterization with differential analysis of chromatin accessibility did not reveal significant
468 alterations of chromatin accessibility between patients with MS and healthy donors at a genome-
469 wide level. We then used our ATAC-seq data to elucidate TF footprint enrichment and found that
470 AP-1 and IRF family TFs are significantly enriched in MS mTreg compared to that of healthy
471 controls. It was of interest that neither AP-1 nor IRFs were detected as differentially expressed
472 genes by transcriptomic analysis, indicating that TF activity cannot be inferred by RNA expression.
473 Conventional TF ChIP-seq is thought to be the best method to provide direct evidence of TF activity;
474 however, it is technically challenging to achieve with primary human *ex vivo* mTregs due to their
475 low frequency in circulating blood. Thus, here we employed ATAC-seq footprint analysis to
476 overcome these issues for epigenetic profiling in human Tregs and provided novel evidence of
477 enriched AP-1 and IRFs bindings in MS mTregs that contributing to the Treg transcriptional
478 signature. These data agree with previous studies highlighting the important role of AP-1 in
479 establishing epigenetic state and linking genetic susceptibility to T cell activation⁸⁴. Of note, recent
480 studies identified Treg specific eQTL effects with MS associated CD28 susceptible locus⁸⁵, which
481 was also confirmed with ImmuNexUT and DICE data (data not shown). Given that CD28 signaling
482 is central in the activation of AP-1 function in T cells, our findings of AP-1 enrichment in MS mTregs
483 potentially suggest that the genomic susceptibility of MS could be mediated by high activity of the

484 AP-1 family in mTregs, resulting in increased *PRDM1-S*. Further studies focused on the functional
485 properties of the MS-associated SNP at the CD28 locus are warranted.

486 A fundamental question relates to the elucidation of molecular interactions between
487 environment triggers and gene transcription driven by allelic variation associated with disease risk
488 that lead to autoimmune disease. We have previously shown that high Na⁺ induces SGK-1 and
489 subsequent inactivation of Foxo1 leading to dysfunctional Tregs. This SGK1-Foxo pathway plays
490 a role in driving pathogenic Th17 cells especially under high salt environment²². Other recent
491 studies in humans have demonstrated higher Na⁺ tissue levels in a subset of patients with MS⁸⁶.
492 Of note, gene set enrichment analysis (GSEA) demonstrated a dysfunctional Foxo1 and Foxo3 KO
493 Treg signature that was enriched in MS and SLE mTreg (Figure S8), consistent with our previous
494 studies showing impaired function of Foxo1 in dysfunctional Tregs in MS^{15,87,88}. Thus, we
495 hypothesize that molecular interactions leading to dysfunctional Tregs in autoimmunity can be
496 driven in part through SGK-1 mediated by high sodium concentration as an environmental factor.
497 However, it should be pointed out that this could be just one of many environmental factors that
498 drive autoimmune disease.

499 In summary, our data uncover fundamental molecular mechanisms by which Treg dysfunction
500 is triggered in patients with MS and potentially other autoimmune diseases. Identification of the
501 primate-specific alternative short *PRDM1* isoform induction in MS Tregs highlights the importance
502 of studying human tissues in addition to mouse models in obtaining insight into disease
503 pathogenesis. Enhancement of the *PRDM1/SGK1* axis in mTreg was observed in the other
504 autoimmune diseases, suggesting shared mechanisms among dysfunctional Tregs in the context
505 of autoimmunity. Furthermore, our findings link epigenetic priming of AP-1 and IRF TF binding in
506 MS Tregs to short *PRDM1* induction. Finally, we believe our rich data of both transcriptome and
507 epigenome profiles on human memory CD4⁺ T cells and Tregs will be a useful tool to explore further
508 insights into pathogenic mechanisms of dysfunctional T cells in autoimmune diseases.

509

510 **Methods**

511 ***Study subjects and ethics statement***

512 Peripheral blood was drawn from people with MS and healthy controls who were recruited as part
513 of an Institutional Review Board (IRB)-approved study at Yale University, and written consent was
514 obtained. All experiments conformed to the principles set out in the WMA Declaration of Helsinki
515 and the Department of Health and Human Services Belmont Report.

516 ***Human T cell isolation and culture***

517 Peripheral blood mononuclear cells (PBMCs) were prepared from whole blood by Ficoll gradient
518 centrifugation (Lymphoprep, STEMCELL Technologies) and used directly for total CD4+ T cell
519 enrichment by negative magnetic selection using Easysep magnetic separation kits (STEMCELL
520 Technologies). Cell suspension was stained with anti-CD4 (RPA-T4), anti-CD25 (clone 2A3), anti-
521 CD45RO (UCHL1), anti-CD45RA (HI100) and anti-CD127 (hIL-7R-M21, all from BD Biosciences)
522 for 30 minutes at 4°C. Naïve Tconv (CD4+/CD25neg/CD127+/CD45ROneg/CD45RA+), Naive
523 Treg (CD4+/CD25hi/CD127neg/CD45ROneg/CD45RA+), Memory Tconv
524 (CD4+/CD127+/CD45RO+/CD45RAneg), and Memory Treg
525 (CD4+/CD25hi/CD127neg/CD45RO+/CD45RAneg) were sorted on a FACSAria (BD Biosciences).
526 Sorted cells were plated in 96-well round-bottom plates (Corning) and cultured in RPMI 1640
527 medium supplemented with 5% Human serum, 2 nM L-glutamine, 5 mM HEPES, and 100 U/ml
528 penicillin, 100 µg/ml streptomycin, 0.5 mM sodium pyruvate, 0.05 mM nonessential amino acids,
529 and 5% human AB serum (Gemini Bio-Products). Cells were seeded (30,000-50,000/well) into
530 wells pre-coated with anti-human CD3 (2 µg/ml, clone UCHT1, BD Biosciences) along with soluble
531 anti-human CD28 (1 µg/ml, clone 28.2, BD Biosciences) in the presence or absence of human IL-
532 2 (50 U/ml).

533 ***Lentiviral transduction***

534 *Lentiviral production*

535 Lentiviral plasmids encoding shRNA for gene knockdown for *PRDM1-L* or open reading frame of
536 overexpression for *PRDM1-S* and *PRDM1-L* were obtained from Sigma-Aldrich (MISSION shRNA)
537 and Horizon Discovery Biosciences (Precision LentiORF), respectively. dCas9-VP64-2A-GFP
538 (Addgene 61422) and pHR-SFFV-dCas9-BFP-KRAB (addgene 46911) were used for generating
539 Jurkat T cell lines for CRISPRa and CRISPRi, respectively. EF1a-RFP-H1-gRNA vector
540 (CASLV502PA-R from System bioscience) was modified to introduce Bsal cut site and single
541 sgRNAs were cloned into it by using Golden Gate Assembly kit (BsmBI-v2, New England Biolabs
542 #E1602). All single sgRNAs used in this study are listed in Supplementary Table 5. Each plasmid

543 was transformed into One Shot Stbl3 chemically competent cells (Invitrogen) and purified by
544 ZymoPURE plasmid Maxiprep kit (Zymo research). Lentiviral pseudoparticles were obtained after
545 plasmid transfection of 293T cells using TurboFectin 8.0 Transfection Reagent (Origene). The
546 medium was replaced after 6-12 h with fresh media with 1X Viral boost (Alstem). The lentivirus
547 containing media was harvested 72 h after transfection and concentrated 80 times using Lenti
548 Concentrator (Origene). LV particles were then resuspended in RPMI 1640 media without serum
549 and stored at -80°C before use. Virus titer was determined by using Jurkat T cells and Lenti-X
550 GoStix Plus (Takara Clontech).

551 Lentiviral transduction

552 FACS-sorted Tregs were plated at 50,000 cells/well in round bottom 96 well plates pre-coated with
553 anti-human CD3 (2 µg/ml, clone UCHT1, BD Biosciences) and soluble anti-human CD28 (1 µg/ml,
554 clone 28.2, BD Biosciences), in the presence of human IL-2 (50 U/ml). After 24 h, cells were
555 transferred into Retronectin coated 96 well plates and 25-50 µl of lenti particles were added to each
556 well, then spininfected with high-speed centrifugation (1000 g) for 1.5 hour at 32 °C. Immediately
557 after centrifugation, cells are placed back to the culture. On day 5, cells are harvested and GFP
558 positive cells are sorted by FACS Aria or analyzed by Fortessa. Gene knockdown effect by *PRDM1-*
559 *L* shRNA is shown in Figure S6B.

560 Jurkat T cells were plated at 50,000 cells/well in round bottom 96 well plates and 25 µl of lenti
561 particles were added to each well, and spininfected as well as above. On day 3-5, cells were scaled
562 up to 12 well plates and followed by the second scale-up at day 7-9 into 6 well plates. Cells were
563 stimulated with PMA and Ionomycin (50 ng/ml and 250 ng/ml respectively) for four hours and
564 GFP⁺/RFP⁺ double positive cells were sorted directly into RNA lysis buffer by FACS Aria.

565 **Suppression assay**

566 CD4⁺CD25⁺CD127^{neg} Treg cells were sorted on a FACS Aria (BD Biosciences). Treg cells were
567 transduced with lentiviral particles containing *PRDM1-S* ORF or GFP control. GFP positive cells
568 were sorted by FACS at day 5, and freshly sorted T effector cells were labeled with cell trace
569 violet dye and then co-cultured with GFP⁺ T_{reg} cells (1 x 10⁴) at different ratio with human Treg
570 inspector beads at 2:1 bead-to-cell ratio. The proliferation of T effector cells was determined at
571 day 4 on a BD Fortessa instrument (BD Bioscience).

572

573 ***Flow cytometry analysis***

574 Cells were stained with LIVE/DEAD Fixable Near-IR Dead Cell Stain kit (Invitrogen) and surface
575 antibodies for 30 min at 4°C. For intracellular cytokine staining, cells were fixed with BD Cytofix™
576 Fixation Buffer (BD Biosciences) for 10 min at RT, then washed with PBS. Intracellular staining
577 was performed in Foxp3 permeabilization buffer (Thermo Fisher) for 30 min at 4°C. The following
578 antibodies were used: anti-Foxp3 (clone 150D, Biolegend, and clone PCH101, Thermo Fisher),
579 anti-Blimp1 (clone 3H2-EB, Thermo Fisher). All antibody information is listed in Supplementary
580 Table 6. Cells were acquired on a BD Fortessa flow cytometer and data was analyzed with FlowJo
581 software v10 (Threestar).

582 ***Immunoblotting***

583 Cells were lysed with RIPA buffer containing EDTA-free cOmplete protease inhibitor cocktail
584 (Roche) and Halt phosphatase inhibitor cocktail (Thermo Fisher Scientific). Extracted protein was
585 quantified with a BCA kit (Thermo Scientific). 0.8-1 µg of protein extract was loaded in each lane,
586 followed by separation by 7.5% SDS-PAGE and transfer to a nitrocellulose membrane. After 1 hour
587 of blocking with 2.5 % bovine serum albumin (BSA) containing 1X Tris-Buffered Saline, 0.1%
588 Tween20 Detergent (TBST), the blotted membranes were then incubated overnight with primary
589 antibodies (anti-Blimp1 (C14A4, 1:1,000, Cell Signaling Technologies or C-7, 1:200, Santa Cruz
590 Biotechnology), anti-β-actin (A5441, 1:10,000, Sigma-Aldrich)) in TBST with 2.5% BSA. Primary
591 antibodies were detected by the secondary antibody horseradish peroxidase (HRP)-conjugated
592 anti-rabbit or mouse (Cell Signaling Technology), and SuperSignal West Femto Maximum
593 Sensitivity ECL Substrate (Pierce) was used as a chemiluminescent substrate for detecting HRP.
594 The images were obtained with a ChemiDoc Imaging system (Bio-Rad).

595 ***Bulk RNA-seq and ATAC-seq library preparation and sequencing***

596 *Bulk RNA-seq*: FACS sorted cells (5,000 cells) were subjected to cDNA synthesis using SMART-
597 Seq v4 Ultra Low Input RNA Kit for sequencing (Takara/Clontech). Barcoded libraries were
598 generated by the Nextera XT DNA Library Preparation kit (Illumina) and sequenced with a 2x100
599 bp paired-end protocol on the HiSeq 4000 Sequencing System (Illumina).

600 *Bulk ATAC-seq*: We adopted FAST-ATAC for our FACS sorted CD4⁺ T cells (5,000 cells) (Corces
601 et al., 2016). Cells were pelleted by centrifugation at 500 g for 7 min at 4°C, then resuspended with
602 50 µl of transposase mixture (25 µL of 2x TD buffer (Illumina), 2.5 µL of TDE1 (Illumina), 0.5 µL of
603 1% digitonin (Thermo Fisher), 22 µL of nuclease-free water). Transposition reactions were
604 incubated at 37°C for 30 minutes in a thermal shaker with agitation at 300 RPM. Transposed DNA

605 was purified using a MinElute Reaction Cleanup kit (QIAGEN) and purified DNA was eluted in 20
606 μ L elution buffer. Transposed fragments were amplified and purified as described previously⁸⁹ with
607 modified primers⁹⁰. Libraries were quantified using qPCR (KAPA Library Quantification Kit) prior
608 to sequencing. All Fast-ATAC libraries were sequenced using a 2x100 bp paired-end protocol on
609 the HiSeq 4000 Sequencing System (Illumina).

610 ***Mint-ChIP library preparation and sequencing***

611 We used 800,000-1,000,000 cryopreserved sorted human primary Tregs for Mint-ChIP profiling.
612 Cells were thawed and immediately pelleted by centrifugation at 400 g for 7 min at 4°C, then
613 resuspended with 100 μ L of ice cold PBS. Approximately 100,000 cells per antibody are lysed in
614 detergent and chromatin is digested with micrococcal nuclease. Double stranded adapters (that
615 contain both a promoter for transcription by T7 RNA polymerase and a demultiplexing adapter) are
616 ligated to the chromatin. Chromatin is mixed overnight with antibodies recognizing histone
617 modifications, and immune complexes are captured using protein A / protein G magnetic bead
618 mixtures. In addition, some chromatin is set aside overnight to enable preparation of an antibody
619 free, input control library. Immobilized immune complexes are washed, and immunoprecipitated
620 DNA is eluted using proteinase K. Recovered DNA is purified with SPRI beads and subject to T7
621 RNA synthesis, thus creating an RNA copy of the immunoprecipitated DNA. RNA is copied back
622 to cDNA using a random primer containing a 5' extension, enabling subsequent PCR amplification
623 with Illumina indexed sequencing primers. PCR products are purified and mixed together to enable
624 multiplex Illumina sequencing. DNA is sequenced using a paired end protocol; in Mint-ChIP3, the
625 first 8 bases of Illumina Read2 serve as an inline barcode enabling demultiplexing of the chromatin
626 using the ligated barcoded adapter. The detailed protocol and primer/adaptor sequences are
627 described in [dx.doi.org/10.17504/protocols.io.wbefaje](https://doi.org/10.17504/protocols.io.wbefaje).

628 ***Mint-ChIP data processing***

629 We processed the Mint-ChIP FASTQ files of each sample using the ENCODE3 ChIP-seq pipeline
630 provided by Anshul Kundaje (<https://github.com/ENCODE-DCC/chip-seq-pipeline2>) with the
631 following parameters “chip.pipeline_type=histone, chip.aligner=bowtie2, chip.true_rep_only=true,
632 chip.paired_end=true, chip.ctl_paired_end=true, chip.always_use_pooled_ctl=false” specified in
633 its json file. We used the default value for all the other parameters. Briefly, the pipeline first mapped
634 the reads to the hg19 human reference genome using bowtie2. The aligned reads were filtered and
635 duplicated reads were removed. The peak calling was then performed using MACS2 with a control
636 sample for each individual. Sample quality was assessed with a cross-correlation plot.

637 **Bulk RNA-seq analysis**

638 After sequencing, adapter sequences and poor-quality bases (quality score < 3) were trimmed
639 with Trimmomatic. Remaining bases were trimmed if their average quality score in a 4 bp sliding
640 window fell below 5. FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was
641 used to obtain quality control metrics before and after trimming. Remaining reads were aligned to
642 the GRCh38 human genome assembly with STAR 2.5.2⁹¹. We used Picard
643 (<https://github.com/broadinstitute/picard>) to remove optical duplicates and to compile alignment
644 summary statistics and RNA-seq summary statistics. After alignment, reads were quantitated to
645 gene level with RSEM⁹² using the GENCODE annotation⁹³.

646 We conducted our initial quality control assessment on the entire dataset, including memory and
647 naïve Tconv and Treg cells obtained from MS patients and healthy controls. A subset of these
648 were stimulated with IL-2 as described above, with the remainder collected in the *ex vivo* state.
649 We used principal component analysis to identify potential sample swaps. We considered genes
650 that were quantitated >1 count per million (cpm) in ≥ 15 samples, normalizing expression values
651 by the trimmed mean of M-values as implemented in edgeR⁹⁴. We used limma⁹⁵ with the voom
652 transformation⁹⁶ to identify differentially expressed genes (DEGs) within mTconv and mTreg cell
653 populations separately. We used RUV-seq⁹⁷ to account for batch effect and other sources of
654 systematic variation; we included RUV parameters along with a sex covariate in our final model.
655 DEG list for *ex vivo* mTreg and mTconv between MS vs HC are shown in Supplementary Table
656 7.

657 **Co-expression analysis in bulk RNA-seq and scRNA-seq**

658 The co-expression between two genes was computed as the Spearman correlation coefficient of
659 normalized gene expression. The normalized gene expression was calculated by dividing the raw
660 count by the library size of the sample and its scaling factor obtained from the TMM normalization.
661 The co-expression analysis of *PRDM1* in the scRNA-seq data was conducted for genes that had
662 the average count per cell >0.005 in the mTreg and mTconv. The co-expression was measured
663 by the log(fold-change) between *PRDM1* and the gene of interest in a negative binomial mixed
664 model with the subjects as random effects implemented in NEBULA⁷⁵. We used normalized
665 expression of *PRDM1* (the raw count divided by the library size of the cell) as the explanatory
666 variable and included in the model the proportion of reads from ribosomal protein genes and
667 mitochondrial genes as covariates. To assess the differential co-expression, the model was fitted
668 for 4896 and 4676 cells from the five MS patients and five healthy controls separately.

669 **Bulk ATAC-seq analysis**

670 The official pipeline of the Encyclopedia of DNA Elements ⁵⁴ consortium
671 (https://github.com/kundajelab/atac_dnase_pipelines) {*kundajelab/atac_dnase_pipelines: 0.3.0*}
672 was adopted to preprocess the ATAC-seq raw data. The preprocessing started with the paired-end
673 ATAC-seq fastq files of each subject. More specifically, reads were trimmed for adapters using
674 cutadapt ⁹⁸ and mapped to the human reference genome (hg19) using bowtie2 ⁹⁹. The output raw
675 bam files were filtered, deduped, and converted to single-ended Tn5-shifted tagalign files, which
676 were then used as the input for peak calling. The deduped bam files were used in the downstream
677 motif and footprint analyses described in a later section.

678 We first called sample-specific narrow peaks (FDR<0.01) using MACS2 ¹⁰⁰ from the tagalign file of
679 each of the samples, separately, with the command “macs2 callpeak -f BED -g hs -q 0.01 --nomodel
680 --shift -75 --extsize 150 -B --SPMR --keep-dup all --call-summits”. We calculated the fraction of
681 reads in called peak regions (FRiP score) for each sample. We then called group-specific narrow
682 peaks (FDR<0.01) by providing MACS2 with the tagalign files of all high-quality samples
683 (FRiP>0.1) within each of the eight groups (healthy/MS mTconv, mTreg, naive Treg (nTreg), and
684 naive Tconv (nTconv)) using the command “macs2 callpeak -t <tagalign files of the high-quality
685 samples within the group> -f BED -g hs -q 0.01 --nomodel --shift -75 --extsize 150 -B --SPMR --
686 keep-dup all --call-summits”. To obtain a unified set of peak regions across all groups, we merged
687 the group-specific peak regions that were overlapping or had maximum distance <100bp using
688 BEDTools ¹⁰¹ with the command “bedtools merge -i -d 100”.

689 Given the unified set of peak regions, the number of reads overlapping a peak was called for each
690 subject using BEDTools. This raw count matrix was used for the downstream analyses. In the
691 analysis of differential accessibility, we first filtered out peaks with low counts and normalized the
692 counts of each subject using TMM ⁹⁴. Then the *voom* function in the limma package ⁹⁵ was used to
693 perform the differential analysis between the cases and controls within each cell type with FRiP,
694 sex and ethnicity as covariates.

695 **Correlation analysis between RNA-seq and ATAC-seq data**

696 The analysis of the correlation between the accessibility of the adjacent open chromatin regions
697 and the *PRDM1* expression was performed using a linear regression model in which the normalized
698 *PRDM1* expression (raw count divided by the total library size and the scaling factor) was the
699 dependent variable and the adjusted ATAC-seq peak height was the explanatory variable. The
700 adjusted ATAC-seq peak height was the residual obtained by fitting a weighted linear model using
701 limma-voom for the peak with FRiP as the covariate. We included 106 samples from all the cell

702 populations and both groups and these samples had both RNA-seq and ATAC-seq measurements.
703 We interrogated 72 ATAC-seq peaks for which *PRDM1* was annotated as the nearest gene by
704 Homer. These peaks span from ~350k bp upstream to ~100k bp downstream of the transcription
705 start site of the long isoform of *PRDM1*.

706 ***ATAC-seq footprint analysis***

707 To conduct differential footprint analysis between the disease groups and cell types, we generated
708 group-level bam files by merging the deduped bam files within each of the eight groups. A total of
709 47 high-quality ATAC-seq samples with FRiP>0.1 were included (MS group: 26 mTregs, control
710 group: 21 mTregs). The differential footprint analysis was performed using HINT v0.12.3¹⁰² and
711 TOBIAS v0.10.1⁶². In the analysis using HINT, we first called footprints using the command “rgt-
712 hint footprinting --atac-seq --paired-end” with the group-specific bam files and peaks as the input
713 files. We then identified predicted binding sites using the command “rgt-motifanalysis
714 matching” for the 579 JASPAR (2018) core motifs for vertebrates. Finally, we identified
715 footprints showing differential binding activity between cell types or disease groups using
716 the command “rgt-hint differential” based on the results from the previous two steps. The
717 same bam files and input files were used in the analysis adopting TOBIAS, in which we first ran
718 “TOBIAS ATACCorrect” to obtain bias-corrected signals and then ran “TOBIAS FootprintScores” to
719 obtain footprint scores. Finally, we performed the differential footprint analysis with the command
720 “TOBIAS BINDetect”.

721 ***Single-cell RNA-seq using 10x Genomics platform***

722 CD4⁺ T cells and CD25^{hi} CD4⁺ T cells were negatively isolated from PBMCs separately by using
723 Easysep human CD4⁺ T cell isolation kits and EasySep Human CD4⁺CD127^{low}CD25⁺ Regulatory
724 T Cell Isolation Kit (STEMCELL Technologies), respectively. To avoid batch effects between
725 healthy and MS samples, and to increase the numbers of Tregs to analyze, we used hashing
726 technology (Biolegend) to pool samples in a single run of the 10x Genomics platform. Each MS
727 sample was processed with a paired healthy control subject matched for age, sex, and ethnicity. A
728 total of five healthy and MS sample pairs were analyzed. 100,000 cells for each cell type were
729 subjected to Total-seq C and Hashtag antibody staining. 2 ug per 1 million cells for Total-seq C
730 and 1 ug per 1 million cells for Hashing antibodies were used for the staining. Cells were washed
731 three times with PBS containing 2% FBS and four hashed samples (total CD4⁺ T cells and
732 CD25^{hi}CD4⁺ T cells from each of HC and MS) were pooled into one sample. The cellular
733 concentration was adjusted to 1,000/μL and loaded into the 10x Genomics instrument aiming to

734 recover 10,000 cells for library preparation and sequencing. Generated libraries were then
735 sequenced on the NovaSeq (Illumina) with a target of 50,000 reads/cell (2×150 paired-end reads).

736 ***Dual omics single-cell analysis***

737 *Cell type assignment by protein surface markers*

738 We define a latent indicator variable \mathbf{z}_{jk} to mark the assignment of a cell j to a cell type k and
739 estimate the posterior probability of $\mathbf{z}_{jk}=1$ by the stochastic expectation maximization (EM)
740 algorithm. We assume that a normalized vector for each cell \mathbf{x}_j follows von Mises-Fisher (vMF)
741 distribution¹⁰³ with cell type k -specific mean vector $\boldsymbol{\mu}_k$ and shared concentration parameter κ :
742 $P(\mathbf{x}_j | \mathbf{z}_{jk}=1) \sim \exp[\kappa \boldsymbol{\theta}_k^\top \mathbf{x}_j]$. We modified the existing EM algorithm¹⁰⁴ and enforced the sparsity of
743 the mean vector $\boldsymbol{\mu}$ based on the prior knowledge of the cell-type-specific activity of marker
744 proteins/genes/features. Simply, we allow $\boldsymbol{\mu}_{gk}$ to take non-zero values if and only if a feature g is a
745 known marker for the cell type k .

746 To improve the quality of inference, we also take advantage of “negative” marker protein labels and
747 build an “adversarial” model for each cell type and contrast with the likelihood of the corresponding
748 “positive” model.

- 749 • nTconv: CD3+, CD4+, CD8-, CD25-/CD127+, CD45RA+/CD45RO-
- 750 • mTconv: CD3+, CD4+, CD8-, CD25-/CD127+, CD45RA-/CD45RO+
- 751 • nTreg: CD3+, CD4+, CD8-, CD25+/CD127-, CD45RA+/CD45RO-
- 752 • mTreg: CD3+, CD4+, CD8-, CD25+/CD127-, CD45RA-/CD45RO+

753 To further dissect the cell types within mTreg cells, we sorted single-cell CITE-seq vectors based
754 on the following definitions:

- 755 • Treg1: CD183+ / CD194- / CD196-
- 756 • Treg1/17: CD183+ / CD194- / CD196+
- 757 • Treg17: CD183- / CD194+ / CD196+
- 758 • Treg2: CD183- / CD194+ / CD196-

760 *Batch correction and visualization of scRNA-seq data*

761 We used the top 100 principal components of the log-transformed scRNA-seq data matrix to
762 characterize intercellular similarity and clustering patterns across ~45k cells and ~15k genes. We
763 managed to adjust discrepancies across five different batches using the batch-balancing k-nearest
764 neighborhood method¹⁰⁵ followed by adjustment of principal component, subtracting out the mean
765 difference between batches¹⁰⁶.

766 *Single-cell differential expression analysis*

767 We compared cell-type-specific gene expression profiles between the MS and HC subjects by
768 estimating unbiased subject-level pseudo-bulk profiles for each gene using CoCoA-diff³⁵. CoCoA-

769 diff can improve the statistical power in case-control scRNA-seq study while adjusting for unwanted
770 confounding effects existing across individuals. We first estimated latent factors, which may
771 confound gene and cell-type-specific expressions with the disease labels. We established a
772 controlled baseline for each cell derived from the MS subjects by imputing counterfactual gene
773 expression values based on the 100 cells found in the HC cells in the top 10 PC space (BBKNN-
774 based weighted average). Likewise, we imputed counterfactual values for the HC cells using the
775 MS cells. More precisely, for each gene \mathbf{g} and cell j , we have observed \mathbf{Y}_{gj}^{16} and counterfactual
776 (imputed) $\mathbf{Y}_{gj}^{(HC)}$ if a cell j were derived from the MS; we have observed $\mathbf{Y}_{gj}^{(HC)}$ and counterfactual
777 $\mathbf{Y}_{gj}^{(MS)}$ if a cell j were from the HC. We then aggregate both factual (observed) and counterfactual
778 cell profiles within each subject and cell type to estimate causal effects by comparing the average
779 disease effect (ADE), average disease effect in the disease subject ¹⁰⁷, and average disease effect
780 in the control subject (ADC). Denoting the subject-level aggregate profiles $\boldsymbol{\lambda}_{gi}^{16}$ and $\boldsymbol{\lambda}_{gi}^{(HC)}$ (for gene
781 \mathbf{g} and subject i), we define ADE of a gene \mathbf{g} as: $\text{ADE}(\mathbf{g}) = \sum_{i=1..10} \log[\boldsymbol{\lambda}_{gi}^{16} / \boldsymbol{\lambda}_{gi}^{(HC)}] / 10$, $\text{ADC}(\mathbf{g}) = \sum_{i$
782 $\text{in } 5 \text{ HC subjects}} \log[\boldsymbol{\lambda}_{gi}^{16} / \boldsymbol{\lambda}_{gi}^{(HC)}] / 5$, $\text{ADD}(\mathbf{g}) = \sum_{i \text{ in } 5 \text{ MS subjects}} \log[\boldsymbol{\lambda}_{gi}^{16} / \boldsymbol{\lambda}_{gi}^{(HC)}] / 5$. We implemented
783 Bayesian inference methods in a C++ program that calculates gene-level statistics, including
784 posterior mean and standard error, efficiently handling ten thousand genes and hundred thousand
785 cells ³⁵.

786 **Data and code availability**

787 The accession number for the whole transcriptome sequencing data, dual omics single-cell
788 sequencing data, and processed data reported in this paper will be provided in GEO/SRA.

789

790 **Acknowledgements**

791 We thank patients and families, members of the Yale MS clinic; Lesley Devine and Chao Wang for
792 technical assistance with flow cytometry; Mei Zhong and the staff of the Yale Stem Cell Center for
793 technical assistance with bulk RNA-seq and ATAC-seq sample preparation and sequencing.

794 **Funding**

795 This work was supported by Race to Erase MS Young investigator award to T.S.S.; a Career
796 Transition Fellowship from the Consortium of MS Centers and the National MS Society to M.R.L.;
797 NIH grants (P01 AI073748, U24 AI11867, R01 AI22220, UM 1HG009390, P01 AI039671, P50
798 CA121974, R01 CA227473) to D.A.H. NIA grants (R01 AG047310, R01 AG061853, R01
799 AG065477, and R01 AG070488) to A.M.K.

800

801 **Author contributions**

802 Conceptualization: T.S.S., M.R.L., L.H., Y.P., M.K., and D.A.H. Recruitment of study participants:
803 T.S.S., M.R.L., and D.A.H. Specimen collection: T.S.S., M.R.L., and H.A.S. Human primary cell
804 experiments: T.S.S., H.A.S., and G.A.L. Mint-ChIP experiment and analysis: T.S.S., M.R.L., L.H.,
805 C.B.E. and B.E.B. Transcriptomic and eQTL analysis of ImmuneNexUT data: M.O., K.F. Data
806 analysis: T.S.S., M.R.L., L.H., and Y.P. Visualization: T.S.S., M.R.L., L.H., and Y.P. Writing review
807 & editing: T.S.S., M.R.L., L.H., Y.P., A.M.K., M.K., and D.A.H. wrote the manuscript, with input from
808 all co-authors. Supervision: M.K., and D.A.H.

809 **Declaration of interests**

810 D.A.H. has received research funding from Bristol-Myers Squibb, Novartis, Sanofi, and Genentech.
811 He has been a consultant for Bayer Pharmaceuticals, Bristol Myers Squibb, Compass Therapeutics,
812 EMD Serono, Genentech, Juno therapeutics, Novartis Pharmaceuticals, Proclara Biosciences,
813 Sage Therapeutics, and Sanofi Genzyme. B.E.B. declares outside interests in Fulcrum
814 Therapeutics, Arsenal Biosciences, HiFiBio, Cell Signaling Technologies and Chroma Medicine.
815

816 **Supplementary Figures**

Figure S1

817

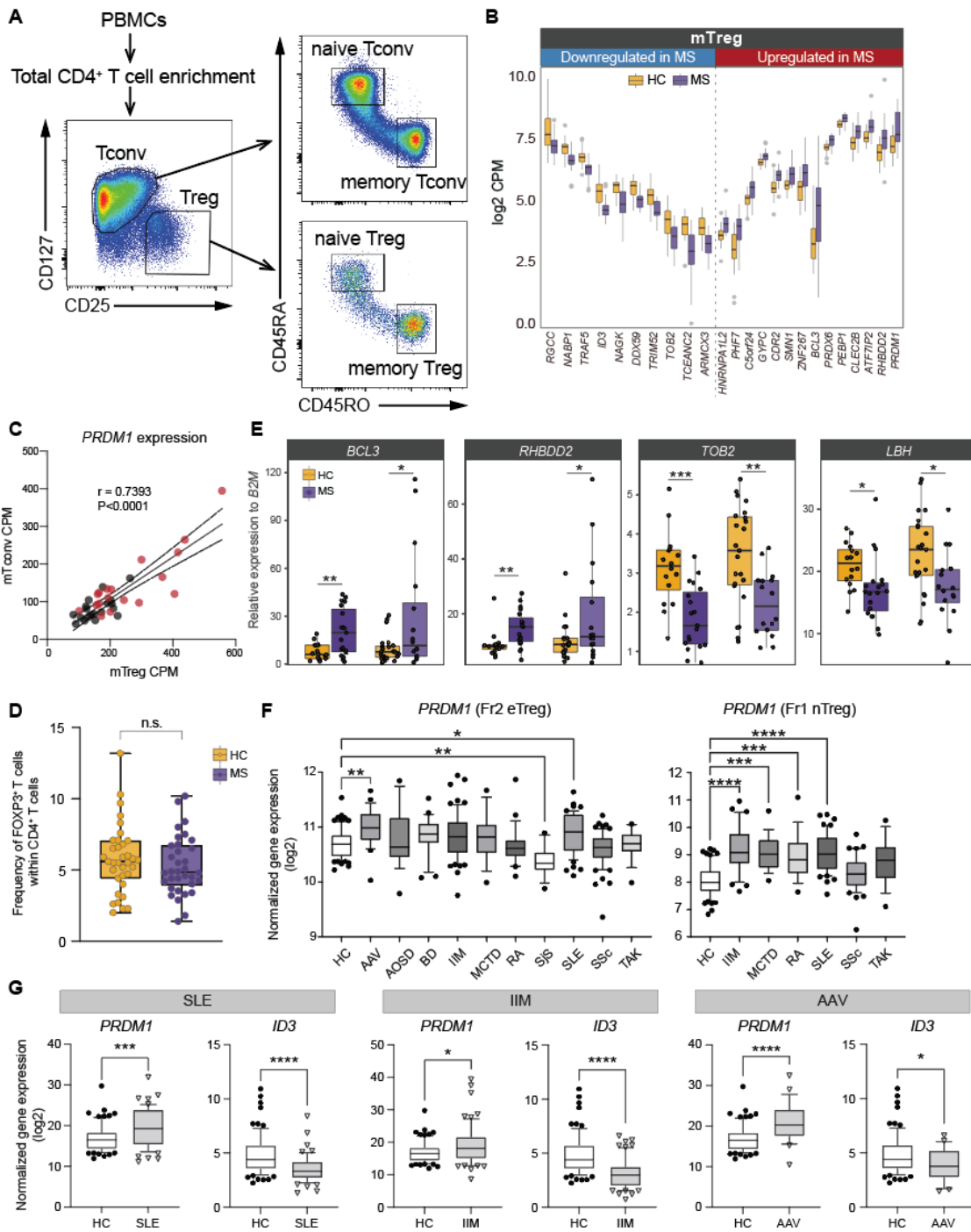


Figure S1 Transcriptomic analysis in mTreg and mTconv with MS patients, related to Figure 1. (A) Experimental workflow for isolating four major CD4⁺ T cell subpopulations by FACS. **(B)** Top individual DEGs expression at subject level. **(C)** Correlation of *PRDM1* expression in mTreg and mTconv. **(D)** Flow cytometry-based assessment of the frequency of FOXP3⁺ T cells in HC vs MS. (n=34 (HC), 35 (MS)) **(E)** qPCR validation for top four DEGs in mTreg. Data for the discovery cohort (left side) and validation cohort (right side) are shown for each HC (yellow) and MS (purple). **(F)** *PRDM1* expression in Fr2 eTreg and Fr1 nTreg across 10 and 6 autoimmune diseases are shown respectively (data were extracted from M. Ota *et al.*). **(G)** *PRDM1* and *ID3* expression in Fr2 eTreg from HC, SLE, IIM, and AAV are shown respectively (data were extracted from M. Ota *et al.*) P* < 0.05, P** < 0.01, P*** < 0.001, P**** < 0.0001; Statistical significance computed by one way ANOVA with Dunn's multiple comparisons tests.

Figure S2

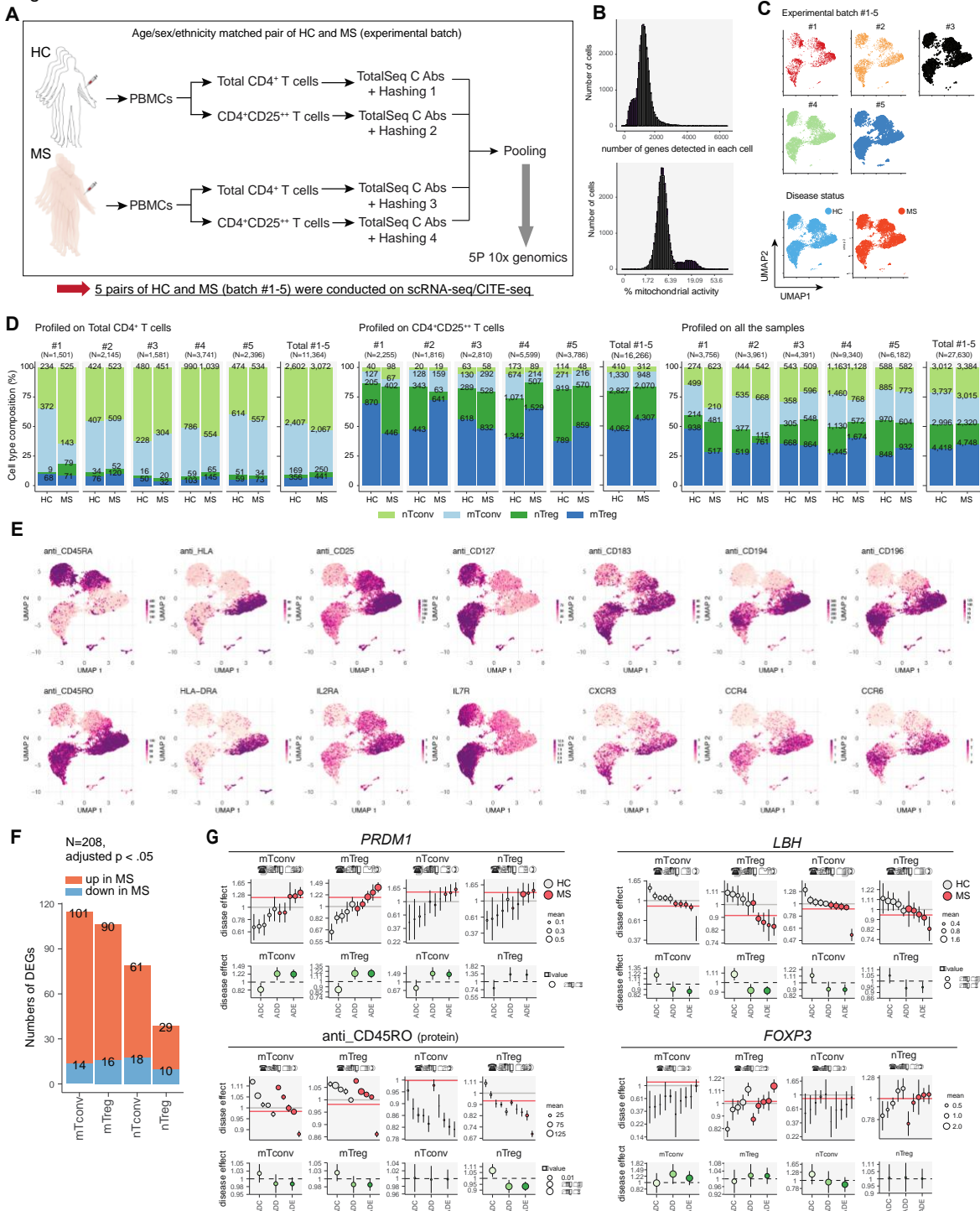


Figure S2 Single-cell dual omics analysis with CD4+ T cells in MS, related to Figure 2. (A) Experimental workflow for dual omics single-cell analysis of HC and MS CD4+ T cells. Age, sex, and ethnicity matched HC and MS subject are processed at the same time as one experimental batch. Total five experimental batches were included in this study. **(B)** Histograms showing numbers of genes detected per cell (left) and frequency of mitochondrial genes per cell (right). Cells before and after quality control selection are highlighted in purple and light gray respectively. **(C)** Gene expression UMAP of all cells color coded for experimental five batches (#1-5) (top) and disease condition (bottom). **(D)** Annotated cell numbers of CD4+ T cell subpopulation within total CD4+ T cells (left), CD4+CD25+ T cells (middle), and combined total cells (right). Cell numbers for each batch and summary of five batches from HC and MS are shown. **(E)** Representative gene and protein expressions UMAP. **(F)** Numbers of upregulated and downregulated DEGs in each CD4+ T cell subpopulation are shown. **(G)** Representative differential gene analysis in each CD4+ T cell subpopulation is depicted (See also Methods).

Figure S3

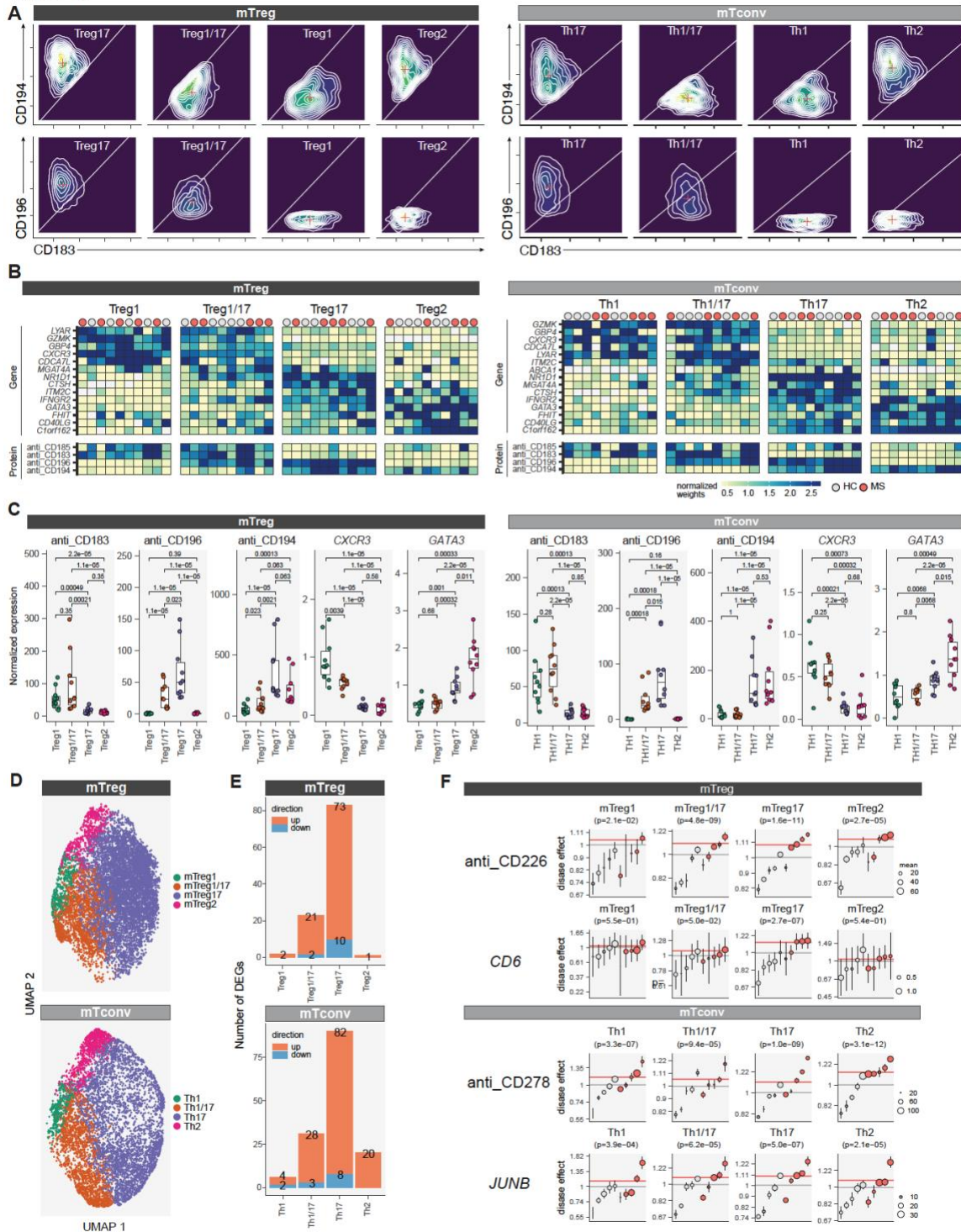
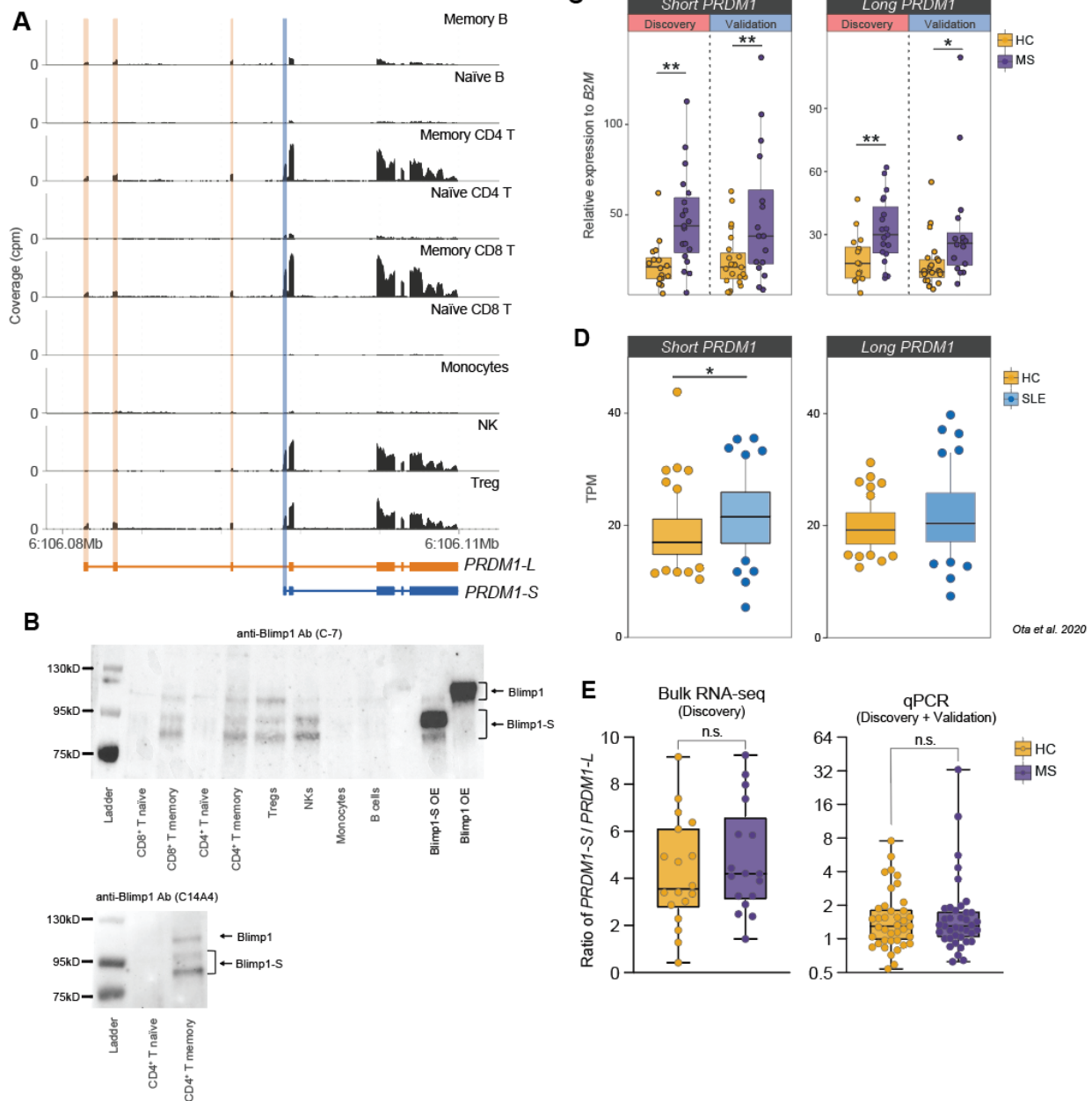


Figure S3 Single cell dual omics analysis with CD4⁺ T cells sub cell types in MS, related to Figure 2. (A) Surface protein guided mTreg and mTconv subtype annotation. CD196, CD183, and CD194 expressions were shown in four subtypes for each mTreg and mTconv. **(B)** Heatmaps showing the marker genes and proteins to define subtypes for each mTconv and mTreg at individual subject level. **(C)** Key marker expressions (CD196, CD183, CD194, CXCR3 and GATA3) for each subtype in mTreg and mTconv. **(D)** UMAP based on CITE-seq based protein expressions for the subtypes in mTreg and mTconv. **(E)** Numbers of upregulated and downregulated DEGs in each subtype for mTreg (top) and mTconv (bottom). **(F)** Representative differential gene and surface protein analysis in each subtype in mTreg and mTconv are depicted (See also Methods).

Figure S4



820

Figure S4 Elevated alternative short *PRDM1* isoform in MS mTreg, related to Figure 3. (A) Representative bulk RNA-seq coverage tracks at *PRDM1* locus from nine different immune cell types in peripheral blood. Unique exonic regions for *PRDM1-L* and *PRDM1-S* are highlighted in orange and blue respectively. (B) Western blot analysis of Blimp1 expression from 8 different immune cell types in peripheral blood and Blimp1-S or Blimp1 overexpressed (OE) 293T cells by anti-Blimp1 Ab (C-7) (top) and anti-Blimp1 Ab (C14A4) (bottom). (C) qPCR validation of short and long *PRDM1* isoform expression between HC and MS from discovery cohort (left box plot) and validation cohort (right box plot). $P < 0.05$, $P < 0.01$; Statistical significance computed by unpaired t test. (D) Short and long *PRDM1* isoform expression between HC and SLE from ImmuneNexUT data. (E) Ratio of *PRDM1-S* vs *PRDM1-L* in bulk RNA-seq and qPCR data. $P < 0.05$, $P < 0.01$; Statistical significance computed by Mann-Whitney test.

821
822
823
824
825
826
827

Figure S5

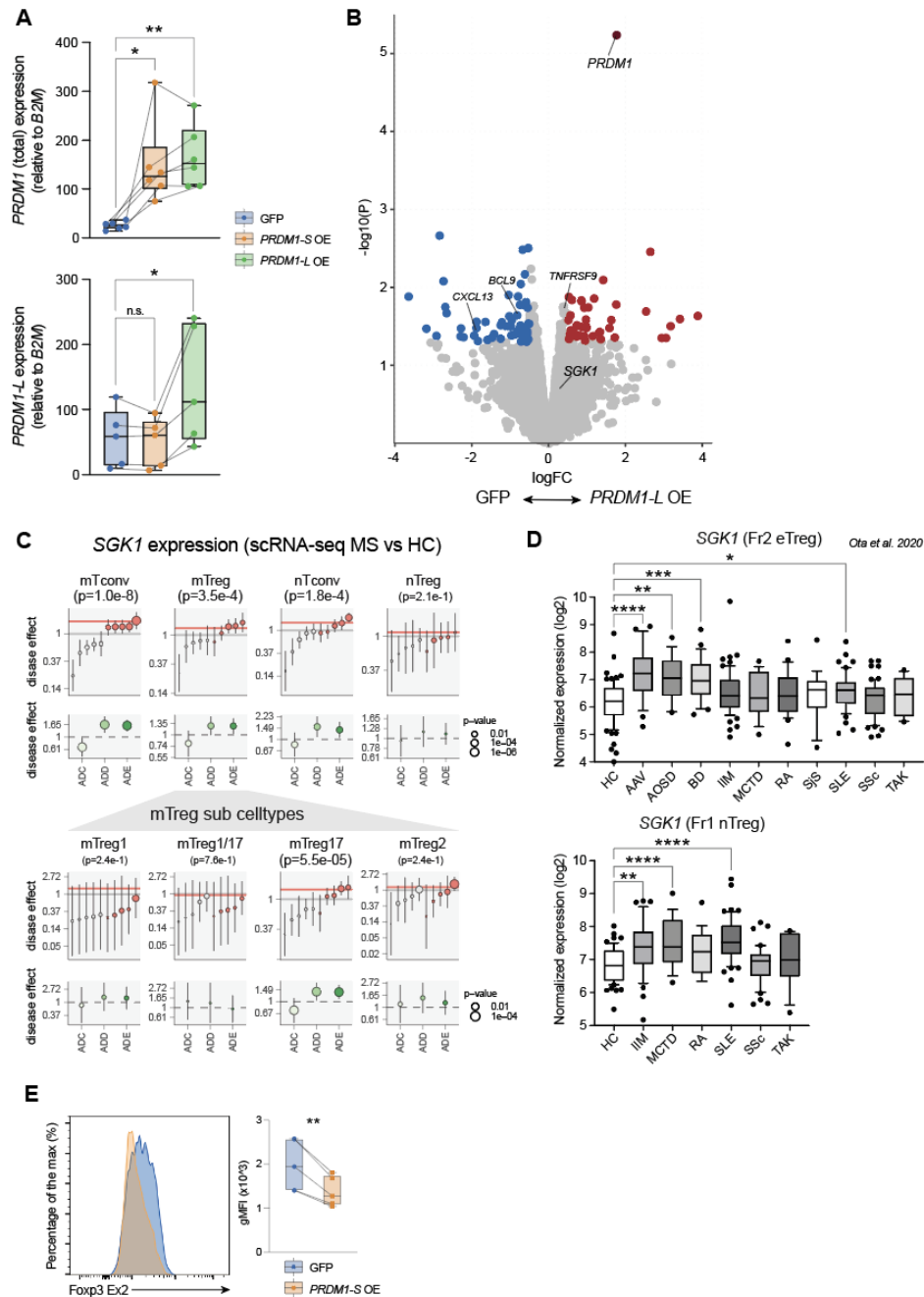


Figure S5 Short PRDM1 induces SGK1 and Treg instability, related to Figure 4. (A) qPCR validation of total PRDM1 and long PRDM1 expression by short and long PRDM1 isoform overexpression in mTreg. $P^* < 0.05$, $P^{**} < 0.01$; Statistical significance computed by paired t test. **(B)** Volcano plot showing statistical significance and fold change for genes differentially expressed by long PRDM1 overexpression in primary mTregs. **(C)** SGK1 expression assessed by scrRNA-seq in four main CD4⁺ T cell subpopulations (top) and mTreg sub cell-types (bottom). ADE: average disease effect between disease cells and matched healthy cells across all MS and HC. ADC: average disease effect only measured within the healthy control group. ADD: average disease effect only measured within the disease group. **(D)** SGK1 expression in Fr2 eTreg and Fr1 nTreg across 10 and 6 autoimmune diseases respectively (data were extracted from *M. Ota et al.*). $P^* < 0.05$, $P^{**} < 0.01$, $P^{***} < 0.001$, $P^{****} < 0.0001$; Statistical significance computed by one way ANOVA with Dunn's multiple comparisons tests. **(E)** Flow cytometry analysis for the specific splicing isoforms of FOXP3 containing exon 2 (FOXP3-Ex2) in primary Treg cells by overexpression of PRDM1-S compared to GFP control ($n=5$). $P^{**} < 0.01$; Statistical significance computed by paired t test.

Figure S6

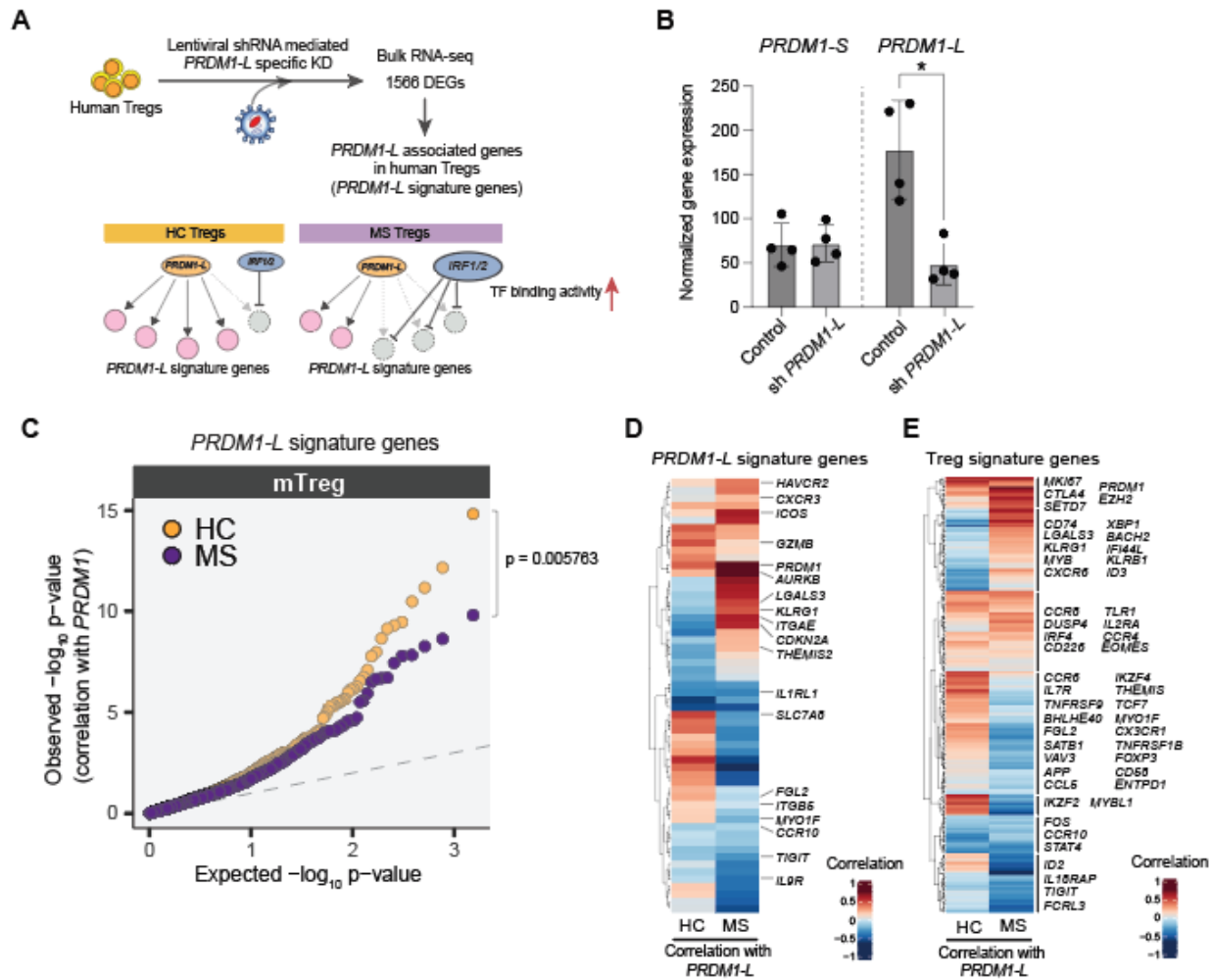


Figure S6 Disrupted *PRDM1-L* mediated gene regulation in MS mTregs, related to Figure 5. (A) Schematic of how *PRDM1-L* mediated gene regulation is disrupted by enriched TF binding of IRF1/2. **(B)** Normalized gene expression for *PRDM1-S* and *PRDM1-L* on human primary Tregs with and without *PRDM1-L* specific gene knockdown by lentiviral shRNA transduction. Lentiviral transduced GFP⁺ cells were sorted by FACS at day 5. mRNAs were isolated and bulk RNA-seq was performed. **(C)** Quantile-quantile plot showing the co-expression correlation between *PRDM1-L* signature genes with *PRDM1* at a single-cell level in mTreg for MS and HC. Correlation between *PRDM1-L* signature genes and *PRDM1* expression is stronger in HC compared to MS. Wilcoxon's test p-value as summary values between MS vs HC: p-value = 0.005763. **(D, E)** Heatmaps depicting the Spearman's correlation between *PRDM1-L* expression and *PRDM1-L* signature genes **(D)** and Treg signature genes **(E)**.

Figure S7

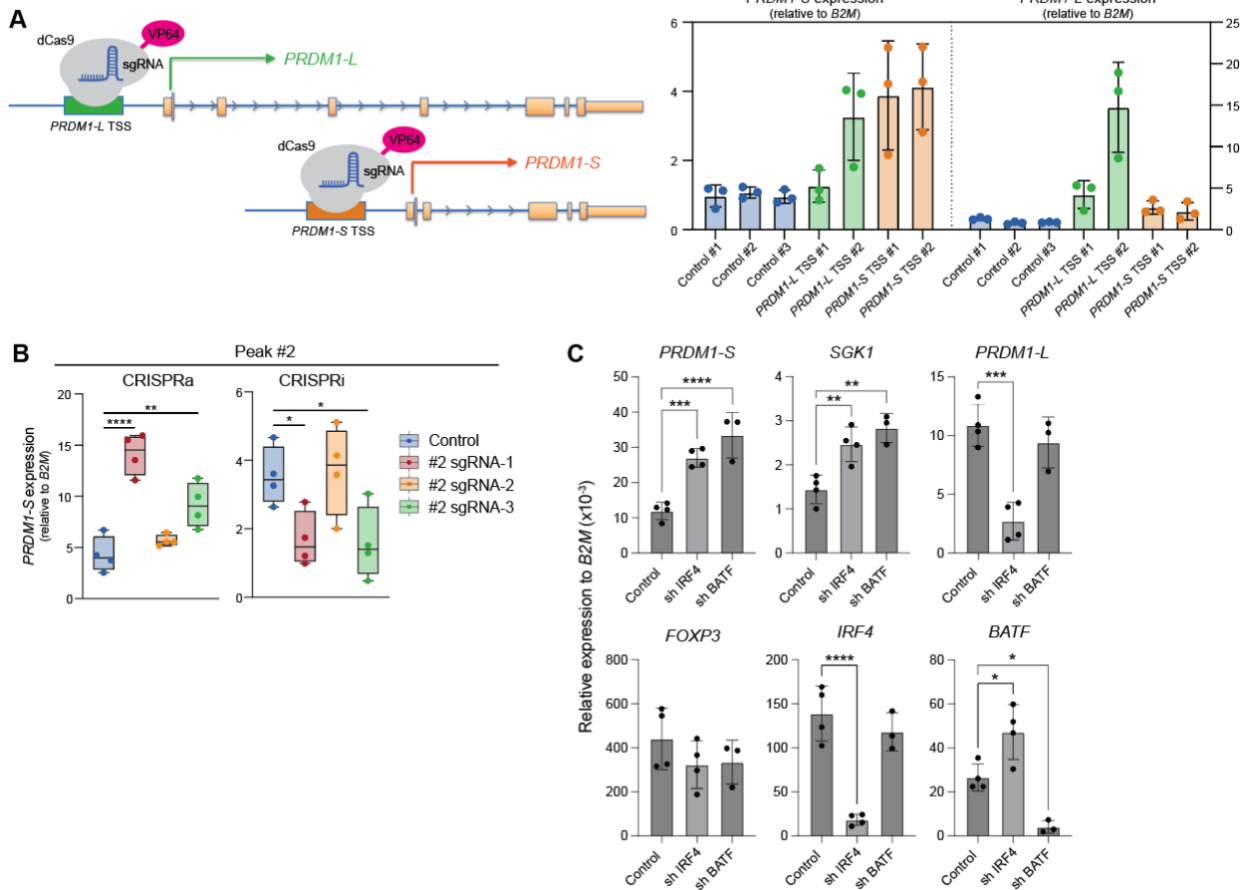


Figure S7 Identification of active enhancer for short *PRDM1* in human T cells, related to Figure 6. (A) Schematic of CRISPRa experiment for short and long *PRDM1* induction with targeting each promoter element (left) and qPCR quantification of short and long *PRDM1* expression (right). **(B)** Validation CRISPRa and CRISPRi experiment for #2 peak independent from Figure 6B (n=4). Peak #2 sgRNA-1 and -3 were validated as functional *cis*-regulatory elements for *PRDM1-S*. **(C)** Lentiviral shRNA-based gene knockdown for IRF4 and BATF in human primary Tregs. Human primary Tregs are isolated by FACS and stimulated with anti-CD3/CD28 antibodies. Lenti particles were transduced at day 1 and GFP⁺ cells were sorted by FACS at day 4-5. Expressions of each gene assessed by qPCR are shown. P* <0.05 , P** <0.01 , P*** <0.001 , P**** <0.0001 ; Statistical significance computed by one way ANOVA with Dunn's multiple comparisons tests.

Figure S8

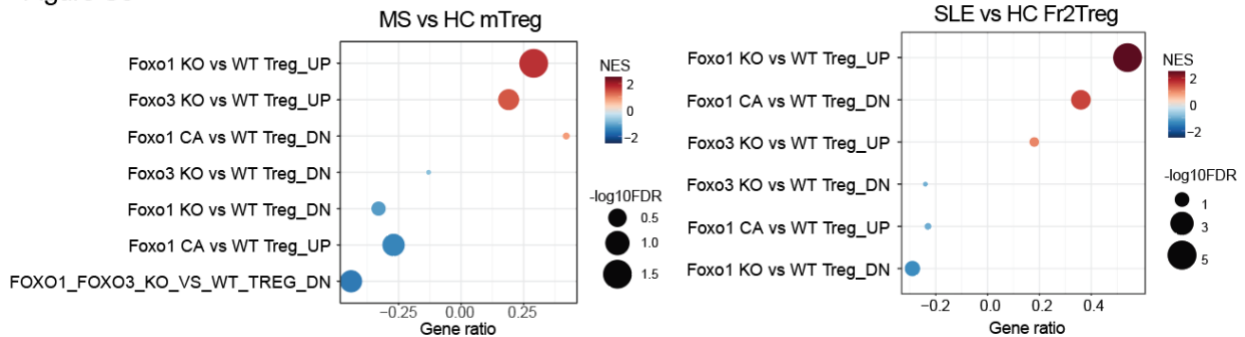


Figure S8 Dysfunctional Foxo1 and Foxo3 KO Treg signatures in MS and SLE mTregs. Dot plots showing gene set enrichment analysis (GSEA) for Foxo signaling on Tregs in MS (left) and SLE (right) Tregs. Gene sets generated by Foxo1/3 KO Tregs and Foxo1 constitutive active (CA) Tregs are used. NES; Normalized enrichment score.

830 References

- 831 1 Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in
832 regulatory DNA. *Science* **337**, 1190-1195 (2012).
833 <https://doi.org/10.1126/science.1222794>
- 834 2 Yoshida, H. *et al.* The cis-Regulatory Atlas of the Mouse Immune System. *Cell* **176**, 897-
835 912 e820 (2019). <https://doi.org/10.1016/j.cell.2018.12.036>
- 836 3 Farh, K. K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease
837 variants. *Nature* **518**, 337-343 (2015). <https://doi.org/10.1038/nature13835>
- 838 4 Hellberg, S. *et al.* Dynamic Response Genes in CD4+ T Cells Reveal a Network of
839 Interactive Proteins that Classifies Disease Activity in Multiple Sclerosis. *Cell Rep* **16**,
840 2928-2939 (2016). <https://doi.org/10.1016/j.celrep.2016.08.036>
- 841 5 Brorson, I. S. *et al.* No differential gene expression for CD4(+) T cells of MS patients and
842 healthy controls. *Mult Scler J Exp Transl Clin* **5**, 2055217319856903 (2019).
843 <https://doi.org/10.1177/2055217319856903>
- 844 6 Cvetanovich, G. L. & Hafler, D. A. Human regulatory T cells in autoimmune diseases.
845 *Curr Opin Immunol* **22**, 753-760 (2010). <https://doi.org/10.1016/j.coi.2010.08.012>
- 846 7 Miyara, M. *et al.* Human FoxP3+ regulatory T cells in systemic autoimmune diseases.
847 *Autoimmun Rev* **10**, 744-755 (2011). <https://doi.org/10.1016/j.autrev.2011.05.004>
- 848 8 Dominguez-Villar, M. & Hafler, D. A. Regulatory T cells in autoimmune disease. *Nat*
849 *Immunol* **19**, 665-673 (2018). <https://doi.org/10.1038/s41590-018-0120-4>
- 850 9 Viglietta, V., Baecher-Allan, C., Weiner, H. L. & Hafler, D. A. Loss of functional
851 suppression by CD4+CD25+ regulatory T cells in patients with multiple sclerosis. *J Exp*
852 *Med* **199**, 971-979 (2004). <https://doi.org/10.1084/jem.20031579>
- 853 10 Dominguez-Villar, M., Baecher-Allan, C. M. & Hafler, D. A. Identification of T helper type
854 1-like, Foxp3+ regulatory T cells in human autoimmune disease. *Nat Med* **17**, 673-675
855 (2011). <https://doi.org/10.1038/nm.2389>
- 856 11 McClymont, S. A. *et al.* Plasticity of human regulatory T cells in healthy subjects and
857 patients with type 1 diabetes. *J Immunol* **186**, 3918-3926 (2011).
858 <https://doi.org/10.4049/jimmunol.1003099>
- 859 12 Farez, M. F., Fiol, M. P., Gaitan, M. I., Quintana, F. J. & Correale, J. Sodium intake is
860 associated with increased disease activity in multiple sclerosis. *J Neurol Neurosurg*
861 *Psychiatry* **86**, 26-31 (2015). <https://doi.org/10.1136/jnnp-2014-307928>
- 862 13 Kleinewietfeld, M. *et al.* Sodium chloride drives autoimmune disease by the induction of
863 pathogenic TH17 cells. *Nature* **496**, 518-522 (2013). <https://doi.org/10.1038/nature11868>
- 864 14 Hernandez, A. L. *et al.* Sodium chloride inhibits the suppressive function of FOXP3+
865 regulatory T cells. *J Clin Invest* **125**, 4212-4222 (2015). <https://doi.org/10.1172/JCI81151>
- 866 15 Sumida, T. *et al.* Activated beta-catenin in Foxp3(+) regulatory T cells links inflammatory
867 environments to autoimmunity. *Nat Immunol* **19**, 1391-1402 (2018).
868 <https://doi.org/10.1038/s41590-018-0236-6>
- 869 16 Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional
870 variation in humans. *Nature* **501**, 506-511 (2013). <https://doi.org/10.1038/nature12531>
- 871 17 Wu, H. *et al.* High salt promotes autoimmunity by TET2-induced DNA demethylation and
872 driving the differentiation of Tfh cells. *Sci Rep* **6**, 28065 (2016).
873 <https://doi.org/10.1038/srep28065>
- 874 18 Cotsapas, C. *et al.* Pervasive sharing of genetic effects in autoimmune disease. *PLoS*
875 *Genet* **7**, e1002254 (2011). <https://doi.org/10.1371/journal.pgen.1002254>
- 876 19 Lincoln, M. R. *et al.* Joint analysis reveals shared autoimmune disease associations and
877 identifies common mechanisms. *medRxiv*, 2021.2005.2013.21257044 (2021).
878 <https://doi.org/10.1101/2021.05.13.21257044>

- 879 20 Panitch, H. S., Hirsch, R. L., Haley, A. S. & Johnson, K. P. Exacerbations of multiple
880 sclerosis in patients treated with gamma interferon. *Lancet* **1**, 893-895 (1987).
881 [https://doi.org/10.1016/s0140-6736\(87\)92863-7](https://doi.org/10.1016/s0140-6736(87)92863-7)
- 882 21 Lee, Y. *et al.* Induction and molecular signature of pathogenic TH17 cells. *Nat Immunol*
883 **13**, 991-999 (2012). <https://doi.org/10.1038/ni.2416>
- 884 22 Wu, C. *et al.* Induction of pathogenic TH17 cells by inducible salt-sensing kinase SGK1.
885 *Nature* **496**, 513-517 (2013). <https://doi.org/10.1038/nature11984>
- 886 23 Cao, Y. *et al.* Functional inflammatory profiles distinguish myelin-reactive T cells from
887 patients with multiple sclerosis. *Sci Transl Med* **7**, 287ra274 (2015).
888 <https://doi.org/10.1126/scitranslmed.aaa8038>
- 889 24 Schwarz, A. *et al.* Fine-tuning of regulatory T cell function: the role of calcium signals
890 and naive regulatory T cells for regulatory T cell deficiency in multiple sclerosis. *J*
891 *Immunol* **190**, 4965-4970 (2013). <https://doi.org/10.4049/jimmunol.1203224>
- 892 25 Ciccocioppo, F. *et al.* The Characterization of Regulatory T-Cell Profiles in Alzheimer's
893 Disease and Multiple Sclerosis. *Sci Rep* **9**, 8788 (2019). [https://doi.org/10.1038/s41598-](https://doi.org/10.1038/s41598-019-45433-3)
894 [019-45433-3](https://doi.org/10.1038/s41598-019-45433-3)
- 895 26 Rauch, K. S. *et al.* Id3 Maintains Foxp3 Expression in Regulatory T Cells by Controlling
896 a Transcriptional Network of E47, Spi-B, and SOCS3. *Cell Rep* **17**, 2827-2836 (2016).
897 <https://doi.org/10.1016/j.celrep.2016.11.045>
- 898 27 Ota, M. *et al.* Dynamic landscape of immune cell-specific gene regulation in immune-
899 mediated diseases. *Cell* **184**, 3006-3021 e3017 (2021).
900 <https://doi.org/10.1016/j.cell.2021.03.056>
- 901 28 Gao, P. *et al.* Risk variants disrupting enhancers of TH1 and TREG cells in type 1
902 diabetes. *Proc Natl Acad Sci U S A* **116**, 7581-7590 (2019).
903 <https://doi.org/10.1073/pnas.1815336116>
- 904 29 Grun, D., Kester, L. & van Oudenaarden, A. Validation of noise models for single-cell
905 transcriptomics. *Nat Methods* **11**, 637-640 (2014). <https://doi.org/10.1038/nmeth.2930>
- 906 30 Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The
907 technology and biology of single-cell RNA sequencing. *Mol Cell* **58**, 610-620 (2015).
908 <https://doi.org/10.1016/j.molcel.2015.04.005>
- 909 31 Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single
910 cells. *Nat Methods* **14**, 865-868 (2017). <https://doi.org/10.1038/nmeth.4380>
- 911 32 Stoeckius, M. *et al.* Cell Hashing with barcoded antibodies enables multiplexing and
912 doublet detection for single cell genomics. *Genome Biol* **19**, 224 (2018).
913 <https://doi.org/10.1186/s13059-018-1603-1>
- 914 33 International Multiple Sclerosis Genetics, C. *et al.* Analysis of immune-related loci
915 identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet* **45**, 1353-1360
916 (2013). <https://doi.org/10.1038/ng.2770>
- 917 34 Soneson, C. & Robinson, M. D. Bias, robustness and scalability in single-cell differential
918 expression analysis. *Nat Methods* **15**, 255-261 (2018).
919 <https://doi.org/10.1038/nmeth.4612>
- 920 35 Park, Y. P. & Kellis, M. CoCoA-diff: counterfactual inference for single-cell gene
921 expression analysis. *Genome Biol* **22**, 228 (2021). [https://doi.org/10.1186/s13059-021-](https://doi.org/10.1186/s13059-021-02438-4)
922 [02438-4](https://doi.org/10.1186/s13059-021-02438-4)
- 923 36 Chapman, N. M. *et al.* mTOR coordinates transcriptional programs and mitochondrial
924 metabolism of activated Treg subsets to protect tissue homeostasis. *Nat Commun* **9**,
925 2095 (2018). <https://doi.org/10.1038/s41467-018-04392-5>
- 926 37 Zhang, F., Liu, G., Li, D., Wei, C. & Hao, J. DDIT4 and Associated IncDDIT4 Modulate
927 Th17 Differentiation through the DDIT4/TSC/mTOR Pathway. *J Immunol* **200**, 1618-
928 1626 (2018). <https://doi.org/10.4049/jimmunol.1601689>

- 929 38 Loo, C. S. *et al.* A Genome-wide CRISPR Screen Reveals a Role for the Non-canonical
930 Nucleosome-Remodeling BAF Complex in Foxp3 Expression and Regulatory T Cell
931 Function. *Immunity* **53**, 143-157 e148 (2020).
932 <https://doi.org/10.1016/j.immuni.2020.06.011>
- 933 39 Yu, X. *et al.* Metabolic control of regulatory T cell stability and function by TRAF3IP3 at
934 the lysosome. *J Exp Med* **215**, 2463-2476 (2018). <https://doi.org/10.1084/jem.20180397>
- 935 40 Paulos, C. M. *et al.* The inducible costimulator (ICOS) is critical for the development of
936 human T(H)17 cells. *Sci Transl Med* **2**, 55ra78 (2010).
937 <https://doi.org/10.1126/scitranslmed.3000448>
- 938 41 Hasan, Z. *et al.* JunB is essential for IL-23-dependent pathogenicity of Th17 cells. *Nat*
939 *Commun* **8**, 15628 (2017). <https://doi.org/10.1038/ncomms15628>
- 940 42 Ma, C. *et al.* Critical Role of CD6highCD4+ T Cells in Driving Th1/Th17 Cell Immune
941 Responses and Mucosal Inflammation in IBD. *J Crohns Colitis* **13**, 510-524 (2019).
942 <https://doi.org/10.1093/ecco-icc/jjy179>
- 943 43 Wang, N. *et al.* CD226 Attenuates Treg Proliferation via Akt and Erk Signaling in an EAE
944 Model. *Front Immunol* **11**, 1883 (2020). <https://doi.org/10.3389/fimmu.2020.01883>
- 945 44 Keller, A. D. & Maniatis, T. Identification and characterization of a novel repressor of
946 beta-interferon gene expression. *Genes Dev* **5**, 868-879 (1991).
947 <https://doi.org/10.1101/gad.5.5.868>
- 948 45 Martins, G. A. *et al.* Transcriptional repressor Blimp-1 regulates T cell homeostasis and
949 function. *Nat Immunol* **7**, 457-465 (2006). <https://doi.org/10.1038/ni1320>
- 950 46 Lin, M. H. *et al.* T cell-specific BLIMP-1 deficiency exacerbates experimental
951 autoimmune encephalomyelitis in nonobese diabetic mice by increasing Th1 and Th17
952 cells. *Clin Immunol* **151**, 101-113 (2014). <https://doi.org/10.1016/j.clim.2014.02.006>
- 953 47 Jain, R. *et al.* Interleukin-23-Induced Transcription Factor Blimp-1 Promotes
954 Pathogenicity of T Helper 17 Cells. *Immunity* **44**, 131-142 (2016).
955 <https://doi.org/10.1016/j.immuni.2015.11.009>
- 956 48 Bankoti, R. *et al.* Differential regulation of Effector and Regulatory T cell function by
957 Blimp1. *Sci Rep* **7**, 12078 (2017). <https://doi.org/10.1038/s41598-017-12171-3>
- 958 49 Ogawa, C. *et al.* Blimp-1 Functions as a Molecular Switch to Prevent Inflammatory
959 Activity in Foxp3(+)RORgammat(+) Regulatory T Cells. *Cell Rep* **25**, 19-28 e15 (2018).
960 <https://doi.org/10.1016/j.celrep.2018.09.016>
- 961 50 Garg, G. *et al.* Blimp1 Prevents Methylation of Foxp3 and Loss of Regulatory T Cell
962 Identity at Sites of Inflammation. *Cell Rep* **26**, 1854-1868 e1855 (2019).
963 <https://doi.org/10.1016/j.celrep.2019.01.070>
- 964 51 Gyory, I., Fejer, G., Ghosh, N., Seto, E. & Wright, K. L. Identification of a functionally
965 impaired positive regulatory domain I binding factor 1 transcription repressor in myeloma
966 cell lines. *J Immunol* **170**, 3125-3133 (2003).
967 <https://doi.org/10.4049/jimmunol.170.6.3125>
- 968 52 Boi, M., Zucca, E., Inghirami, G. & Bertoni, F. PRDM1/BLIMP1: a tumor suppressor
969 gene in B and T cell lymphomas. *Leuk Lymphoma* **56**, 1223-1228 (2015).
970 <https://doi.org/10.3109/10428194.2014.953155>
- 971 53 Wang, X. *et al.* High-resolution genome-wide functional dissection of transcriptional
972 regulatory regions and nucleotides in human. *Nat Commun* **9**, 5380 (2018).
973 <https://doi.org/10.1038/s41467-018-07746-1>
- 974 54 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome.
975 *Nature* **489**, 57-74 (2012). <https://doi.org/10.1038/nature11247>
- 976 55 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human
977 epigenomes. *Nature* **518**, 317-330 (2015). <https://doi.org/10.1038/nature14248>
- 978 56 Smith, M. A. *et al.* PRDM1/Blimp-1 controls effector cytokine production in human NK
979 cells. *J Immunol* **185**, 6058-6067 (2010). <https://doi.org/10.4049/jimmunol.1001682>

- 980 57 Fragale, A. *et al.* IFN regulatory factor-1 negatively regulates CD4+ CD25+ regulatory T
981 cell differentiation by repressing Foxp3 expression. *J Immunol* **181**, 1673-1682 (2008).
982 <https://doi.org/10.4049/jimmunol.181.3.1673>
- 983 58 Hayatsu, N. *et al.* Analyses of a Mutant Foxp3 Allele Reveal BATF as a Critical
984 Transcription Factor in the Differentiation and Accumulation of Tissue Regulatory T
985 Cells. *Immunity* **47**, 268-283 e269 (2017). <https://doi.org/10.1016/j.immuni.2017.07.008>
- 986 59 Renoux, F. *et al.* The AP1 Transcription Factor Fosl2 Promotes Systemic Autoimmunity
987 and Inflammation by Repressing Treg Development. *Cell Rep* **31**, 107826 (2020).
988 <https://doi.org/10.1016/j.celrep.2020.107826>
- 989 60 Wu, C. *et al.* SGK1 Governs the Reciprocal Development of Th17 and Regulatory T
990 Cells. *Cell Rep* **22**, 653-665 (2018). <https://doi.org/10.1016/j.celrep.2017.12.068>
- 991 61 Li, Z. *et al.* Identification of transcription factor binding sites using ATAC-seq. *Genome*
992 *Biology* **20**, 45 (2019). <https://doi.org/10.1186/s13059-019-1642-2>
- 993 62 Bentsen, M. *et al.* ATAC-seq footprinting unravels kinetics of transcription factor binding
994 during zygotic genome activation. *Nat Commun* **11**, 4267 (2020).
995 <https://doi.org/10.1038/s41467-020-18035-1>
- 996 63 Li, P. *et al.* BATF–JUN is critical for IRF4-mediated transcription in T cells. *Nature* **490**,
997 543-546 (2012). <https://doi.org/10.1038/nature11530>
- 998 64 Murphy, T. L., Tussiwand, R. & Murphy, K. M. Specificity through cooperation: BATF-IRF
999 interactions control immune-regulatory networks. *Nat Rev Immunol* **13**, 499-509 (2013).
1000 <https://doi.org/10.1038/nri3470>
- 1001 65 Gate, R. E. *et al.* Genetic determinants of co-accessible chromatin regions in activated T
1002 cells across humans. *Nat Genet* **50**, 1140-1150 (2018). [https://doi.org/10.1038/s41588-](https://doi.org/10.1038/s41588-018-0156-2)
1003 [018-0156-2](https://doi.org/10.1038/s41588-018-0156-2)
- 1004 66 Pham, D. *et al.* Batf Pioneers the Reorganization of Chromatin in Developing Effector T
1005 Cells via Ets1-Dependent Recruitment of Ctcf. *Cell Rep* **29**, 1203-1220 e1207 (2019).
1006 <https://doi.org/10.1016/j.celrep.2019.09.064>
- 1007 67 Lee, S. M., Gao, B. & Fang, D. FoxP3 maintains Treg unresponsiveness by selectively
1008 inhibiting the promoter DNA-binding activity of AP-1. *Blood* **111**, 3599-3606 (2008).
1009 <https://doi.org/10.1182/blood-2007-09-115014>
- 1010 68 Samstein, R. M. *et al.* Foxp3 exploits a pre-existent enhancer landscape for regulatory T
1011 cell lineage specification. *Cell* **151**, 153-166 (2012).
1012 <https://doi.org/10.1016/j.cell.2012.06.053>
- 1013 69 Lam, A. J. *et al.* Optimized CRISPR-mediated gene knockin reveals FOXP3-
1014 independent maintenance of human Treg identity. *Cell Rep* **36**, 109494 (2021).
1015 <https://doi.org/10.1016/j.celrep.2021.109494>
- 1016 70 Costantino, C. M., Baecher-Allan, C. & Hafler, D. A. Multiple sclerosis and regulatory T
1017 cells. *J Clin Immunol* **28**, 697-706 (2008). <https://doi.org/10.1007/s10875-008-9236-x>
- 1018 71 Sambucci, M. *et al.* FoxP3 isoforms and PD-1 expression by T regulatory cells in
1019 multiple sclerosis. *Scientific Reports* **8**, 3674 (2018). [https://doi.org/10.1038/s41598-018-](https://doi.org/10.1038/s41598-018-21861-5)
1020 [21861-5](https://doi.org/10.1038/s41598-018-21861-5)
- 1021 72 Xi, H. & Blanck, G. The IRF-2 DNA binding domain facilitates the activation of the class
1022 II transactivator (CIITA) type IV promoter by IRF-1. *Mol Immunol* **39**, 677-684 (2003).
1023 [https://doi.org/10.1016/s0161-5890\(02\)00214-6](https://doi.org/10.1016/s0161-5890(02)00214-6)
- 1024 73 Mould, A. W., Morgan, M. A., Nelson, A. C., Bikoff, E. K. & Robertson, E. J.
1025 Blimp1/Prdm1 Functions in Opposition to Irf1 to Maintain Neonatal Tolerance during
1026 Postnatal Intestinal Maturation. *PLoS Genet* **11**, e1005375 (2015).
1027 <https://doi.org/10.1371/journal.pgen.1005375>
- 1028 74 Lal, G. *et al.* Distinct inflammatory signals have physiologically divergent effects on
1029 epigenetic regulation of Foxp3 expression and Treg function. *Am J Transplant* **11**, 203-
1030 214 (2011). <https://doi.org/10.1111/j.1600-6143.2010.03389.x>

- 1031 75 He, L. *et al.* NEBULA is a fast negative binomial mixed model for differential or co-
1032 expression analysis of large-scale multi-subject single-cell data. *Commun Biol* **4**, 629
1033 (2021). [https://doi.org:10.1038/s42003-021-02146-6](https://doi.org/10.1038/s42003-021-02146-6)
- 1034 76 Hill, J. A. *et al.* Foxp3 transcription-factor-dependent and -independent regulation of the
1035 regulatory T cell transcriptional signature. *Immunity* **27**, 786-800 (2007).
1036 [https://doi.org:10.1016/j.immuni.2007.09.010](https://doi.org/10.1016/j.immuni.2007.09.010)
- 1037 77 Ferraro, A. *et al.* Interindividual variation in human T regulatory cells. *Proceedings of the*
1038 *National Academy of Sciences* **111**, E1111-E1120 (2014).
1039 [https://doi.org:10.1073/pnas.1401343111](https://doi.org/10.1073/pnas.1401343111)
- 1040 78 Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered
1041 CRISPR-Cas9 complex. *Nature* **517**, 583-588 (2015).
1042 [https://doi.org:10.1038/nature14136](https://doi.org/10.1038/nature14136)
- 1043 79 Fishilevich, S. *et al.* GeneHancer: genome-wide integration of enhancers and target
1044 genes in GeneCards. *Database (Oxford)* **2017** (2017).
1045 [https://doi.org:10.1093/database/bax028](https://doi.org/10.1093/database/bax028)
- 1046 80 van Galen, P. *et al.* A Multiplexed System for Quantitative Comparisons of Chromatin
1047 Landscapes. *Mol Cell* **61**, 170-180 (2016). [https://doi.org:10.1016/j.molcel.2015.11.003](https://doi.org/10.1016/j.molcel.2015.11.003)
- 1048 81 Glasmacher, E. *et al.* A genomic regulatory element that directs assembly and function
1049 of immune-specific AP-1-IRF complexes. *Science* **338**, 975-980 (2012).
1050 [https://doi.org:10.1126/science.1228309](https://doi.org/10.1126/science.1228309)
- 1051 82 Ciofani, M. *et al.* A validated regulatory network for Th17 cell specification. *Cell* **151**,
1052 289-303 (2012). [https://doi.org:10.1016/j.cell.2012.09.016](https://doi.org/10.1016/j.cell.2012.09.016)
- 1053 83 Li, J., Comeau, H. Y., Zhang, Z. & Ren, X. Landscape of transcript isoforms in single T
1054 cells infiltrating in non-small-cell lung cancer. *J Genet Genomics* **47**, 373-388 (2020).
1055 [https://doi.org:10.1016/j.jgg.2020.06.006](https://doi.org/10.1016/j.jgg.2020.06.006)
- 1056 84 Yukawa, M. *et al.* AP-1 activity induced by co-stimulation is required for chromatin
1057 opening during T cell activation. *J Exp Med* **217** (2020).
1058 [https://doi.org:10.1084/jem.20182009](https://doi.org/10.1084/jem.20182009)
- 1059 85 Bossini-Castillo, L. *et al.* Immune disease variants modulate gene expression in
1060 regulatory CD4(+) T cells. *Cell Genom* **2**, None (2022).
1061 [https://doi.org:10.1016/j.xgen.2022.100117](https://doi.org/10.1016/j.xgen.2022.100117)
- 1062 86 Huhn, K. *et al.* Skin sodium is increased in male patients with multiple sclerosis and
1063 related animal models. *Proc Natl Acad Sci U S A* **118** (2021).
1064 [https://doi.org:10.1073/pnas.2102549118](https://doi.org/10.1073/pnas.2102549118)
- 1065 87 Lowther, D. E. *et al.* PD-1 marks dysfunctional regulatory T cells in malignant gliomas.
1066 *JCI Insight* **1** (2016). [https://doi.org:10.1172/jci.insight.85935](https://doi.org/10.1172/jci.insight.85935)
- 1067 88 Lucca, L. E. *et al.* TIGIT signaling restores suppressor function of Th1 Tregs. *JCI Insight*
1068 **4** (2019). [https://doi.org:10.1172/jci.insight.124427](https://doi.org/10.1172/jci.insight.124427)
- 1069 89 Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for
1070 Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21 29 21-21
1071 29 29 (2015). [https://doi.org:10.1002/0471142727.mb2129s109](https://doi.org/10.1002/0471142727.mb2129s109)
- 1072 90 Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory
1073 variation. *Nature* **523**, 486-490 (2015). [https://doi.org:10.1038/nature14590](https://doi.org/10.1038/nature14590)
- 1074 91 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21
1075 (2013). [https://doi.org:10.1093/bioinformatics/bts635](https://doi.org/10.1093/bioinformatics/bts635)
- 1076 92 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with
1077 or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
1078 [https://doi.org:10.1186/1471-2105-12-323](https://doi.org/10.1186/1471-2105-12-323)
- 1079 93 Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE
1080 Project. *Genome Res* **22**, 1760-1774 (2012). [https://doi.org:10.1101/gr.135350.111](https://doi.org/10.1101/gr.135350.111)

- 1081 94 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for
1082 differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-
1083 140 (2010). [https://doi.org:10.1093/bioinformatics/btp616](https://doi.org/10.1093/bioinformatics/btp616)
1084 95 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing
1085 and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
1086 [https://doi.org:10.1093/nar/gkv007](https://doi.org/10.1093/nar/gkv007)
1087 96 Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear
1088 model analysis tools for RNA-seq read counts. *Genome Biology* **15**, R29 (2014).
1089 [https://doi.org:10.1186/gb-2014-15-2-r29](https://doi.org/10.1186/gb-2014-15-2-r29)
1090 97 Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. Normalization of RNA-seq data using factor
1091 analysis of control genes or samples. *Nat Biotechnol* **32**, 896-902 (2014).
1092 [https://doi.org:10.1038/nbt.2931](https://doi.org/10.1038/nbt.2931)
1093 98 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing
1094 reads. *2011* **17**, 3 (2011). [https://doi.org:10.14806/ej.17.1.200](https://doi.org/10.14806/ej.17.1.200)
1095 99 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods*
1096 **9**, 357-359 (2012). [https://doi.org:10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923)
1097 100 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137
1098 (2008). [https://doi.org:10.1186/gb-2008-9-9-r137](https://doi.org/10.1186/gb-2008-9-9-r137)
1099 101 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic
1100 features. *Bioinformatics* **26**, 841-842 (2010).
1101 [https://doi.org:10.1093/bioinformatics/btq033](https://doi.org/10.1093/bioinformatics/btq033)
1102 102 Gusmao, E. G., Allhoff, M., Zenke, M. & Costa, I. G. Analysis of computational
1103 footprinting methods for DNase sequencing experiments. *Nat Methods* **13**, 303-309
1104 (2016). [https://doi.org:10.1038/nmeth.3772](https://doi.org/10.1038/nmeth.3772)
1105 103 Fisher, R. A. Dispersion on a sphere. *Proceedings of the Royal Society of London.*
1106 *Series A. Mathematical and Physical Sciences* **217**, 295 - 305 (1953).
1107 104 Banerjee, A., Dhillon, I. S., Ghosh, J. & Sra, S. Clustering on the Unit Hypersphere using
1108 von Mises-Fisher Distributions. *J. Mach. Learn. Res.* **6**, 1345-1382 (2005).
1109 105 Polański, K. *et al.* BBKNN: fast batch alignment of single cell transcriptomes.
1110 *Bioinformatics* **36**, 964-965 (2019). [https://doi.org:10.1093/bioinformatics/btz625](https://doi.org/10.1093/bioinformatics/btz625)
1111 106 Hie, B., Bryson, B. & Berger, B. Efficient integration of heterogeneous single-cell
1112 transcriptomes using Scanorama. *Nature Biotechnology* **37**, 685-691 (2019).
1113 [https://doi.org:10.1038/s41587-019-0113-3](https://doi.org/10.1038/s41587-019-0113-3)
1114 107 Unterman, A. *et al.* Single-cell multi-omics reveals dyssynchrony of the innate and
1115 adaptive immune system in progressive COVID-19. *Nat Commun* **13**, 440 (2022).
1116 [https://doi.org:10.1038/s41467-021-27716-4](https://doi.org/10.1038/s41467-021-27716-4)

1117