Title: Dissection of multiple sclerosis genetics identifies B and CD4+ T cells as driver cell subsets

Michael H. Guo, MD, PhD[1,2]*; Prashanth Sama[2-5], MSc; Brenna A. LaBarre, PhD[2-5], Hrishikesh[6-7] Lokhande, MSc, John Balibalos[8], Ci Chu, PhD[8], Xiaomi Du[8], Pouya Kheradpour, PhD[8], Charles C Kim, PhD[8], Taylor Oniskey[8], Thomas Snyder, PhD[8], Damien Z Soghoian, PhD[8], Howard L. Weiner, MD[6-7], Tanuja Chitnis, MD[6-7], Nikolaos A. Patsopoulos, MD, PhD [2-5] *

1. Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA.
2. Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA, USA
3. Systems Biology and Computer Science Program, Ann Romney Center for Neurological Diseases, Department of Neurology, Brigham & Women's Hospital, Boston, 02115 MA, USA
4. Division of Genetics, Department of Medicine, Brigham & Women's Hospital, Harvard Medical School, Boston, MA, USA.
5. Harvard Medical School, Boston, MA 02115, USA.
6. Brigham Multiple Sclerosis Center, Brigham and Women's Hospital, Boston, Massachusetts, USA.
7. Ann Romney Center for Neurologic Diseases, Brigham and Women's Hospital, Boston, Massachusetts, USA.
8. Verily Life Sciences, South San Francisco, 94080

*Corresponding authors

Correspondence should be addressed to:

Nikolaos Patsopoulos, npatsopoulos@rics.bwh.harvard.edu

Michael Guo, michael.guo@pennmedicine.upenn.edu

Target Journal: Nature Communications

Word Count (without Methods): 5174

Word count, Methods: 2305

Figures: 8

Supplementary material: 29 Supplementary Tables, 17 Supplementary Figures

Abstract: 151

References: 57

Abstract

Multiple sclerosis (MS) is an autoimmune condition of the central nervous system with a well-characterized genetic background. Prior analyses of MS genetics have identified broad enrichments across peripheral immune cells, yet the driver immune subsets are unclear. We utilized chromatin accessibility data across hematopoietic cells to identify cell type-specific enrichments of MS genetic signals. We found that CD4 T and B cells were independently enriched for MS genetics and further refined the driver subsets to T_h17 and memory B cells, respectively. We replicated our findings in data from untreated and treated MS patients and found that immunomodulatory treatments suppress chromatin accessibility at driver cell types. Integration of statistical fine-mapping and chromatin interactions nominated numerous putative causal genes, illustrating complex interplay between shared and cell-specific genes. Our study highlights how careful integration of genetics and epigenetics can provide fine-scale insights into causal cell types and nominate new genes and pathways for disease.

Introduction

Multiple sclerosis (MS) is an immune-mediated neurodegenerative disease characterized by demyelinating focal lesions in the central nervous system (CNS)[1]. Despite the CNS being the target of autoimmunity, there is extensive evidence from basic science models and human studies that dysregulation of the peripheral immune compartment is key for disease manifestation and progression [2]. MS has long been regarded as a T cell-mediated disease, with several T cell subpopulations implicated [3,4]. More recently, other peripheral immune cell populations, most notably B cells, have also been shown to drive disease pathogenesis[3,5]. Moreover, immune modulating therapies targeting B cells have been demonstrated to be remarkably effective in treating patients with MS[6,7].

MS has a strong genetic component and is characterized by a polygenic architecture. To date, over 200 independent genetic variants have been associated with MS risk, the vast majority of which are common variants with small effect sizes on disease risk[8,9]. Prior studies have shown enrichment of GWAS target genes in the peripheral immune system, but it is unclear exactly which cell types within the peripheral immune system drive these observed enrichments of genetic signals.

Human genetics has emerged as a powerful tool for probing the underlying biology of a disease[10]. The identification of genes and pathways prioritized by GWAS associations is not constrained by our prior knowledge of disease mechanisms and can therefore identify novel biological mechanisms. However, a key challenge for translating GWAS findings into biological insights is that most associations are noncoding in nature and likely act by modulating regulatory elements to mediate gene expression[11,12]. Identifying the causal gene at these GWAS signals can be challenging since it is usually unclear which gene a given regulatory element regulates [10,13,14]. A key step in translating genetic associations into biological

mechanisms is identifying the cell types in which GWAS variants act and the genes they modulate. To help understand the function of disease-associated variants, genetic associations can be intersected with orthogonal epigenetic and gene expression data[12]. As the epigenetic and gene expression landscape differ from cell type to cell type, examining the enrichments of GWAS data on these orthogonal datasets can identify specific cell types that may be implicated in disease pathogenesis[15,16].

We and others have previously reported a strong enrichment of MS GWAS variants in regulatory regions of multiple cell types of the peripheral immune system[8,9,15,17]. However, it has yet to be determined if these enrichments are driven by shared regulatory mechanisms common to many immune cell types, or whether different mechanisms are present in distinct immune cell populations. To address this gap, we performed detailed analyses of the enrichment of MS GWAS variants in the peripheral immune system to identify cell populations that independently mediate the effects of MS GWAS variants on disease risk.

Results

MS GWAS associations are enriched in progenitor and terminal peripheral immune cells

To identify the causal cell types that uniquely and independently contribute to MS pathogenesis via mediation of genetic effects, we leveraged bulk ATAC-seq data from 16 flow-sorted hematopoietic progenitor and terminal cell populations isolated from human peripheral blood or bone marrow[18–20]. These cells represent progenitor and terminal populations from across the hematopoietic tree, enabling investigation of MS GWAS enrichments broadly and across stem, progenitor, and mature cell populations (**Figure 1A**). ATAC-seq data were processed and open chromatin regions (OCRs, i.e. ATAC-seq peaks), were identified as previously described[18,21]. We applied stratified LD SCore regression (LDSC)[22,23] to estimate the enrichment of MS GWAS

in OCRs from each of the 16 hematopoietic cell types. LDSC has the distinct advantage in that it leverages the genome-wide polygenic signal in the GWAS summary statistics rather than selecting variants based on p-value thresholds or fine-mapping posterior probabilities.

Applying LDSC, we observed strong statistically significant enrichments across all hematopoietic cell populations, even after correcting for multiple hypothesis testing (Bonferroni-corrected p-value threshold of 3.13×10^{-3}) (**Figure 1B; Supplemental Table 1**). The strongest enrichments for MS GWAS were observed in OCRs from CD4 T cells (enrichment p-value = 1.47×10^{-18}), CD8 T cells (p-value = 4.00×10^{-18}) and B cells (p-value = 3.27×10^{-15}); reflecting their known and emerging roles in MS pathogenesis and as targets of treatment[1,3,24]. We also detected strong enrichment in OCRs from natural killer (NK) cells (p-value =4.23×10^{-14}) which have a less well-established role in MS[24]. Interestingly, we observed enrichments across all progenitor cells, suggesting that many MS genetic associations are located in regulatory regions involved in core cellular processes in immune cell populations.

CD4 T and B cell regulatory regions independently mediate MS genetics

Many of the studied cell populations share common cellular regulatory signatures, which is reflected in the substantial correlation of the OCR profiles across cell populations (**Figure S1**). Hence, we examined whether the strong enrichment observed across cell populations is a result of truly independent cell type-specific enrichments or whether it is due to shared regulatory landscape across immune cell types. To address this question, we applied a joint model in LDSC to measure the contribution of OCRs from a given cell type, stratified on all other cell types in the model along with a set of baseline annotations. We report the p-value of the coefficient τ_c, which reflects the SNP heritability of a given annotation stratified on all other annotations in the model. In this joint model where OCRs from all 16 cell types were included,

122 we observed that B cells and CD4 T cells contributed significantly to SNP heritability (coefficient
123 p-value = 3.99×10^{-5} and 3.49×10^{-4}, respectively), suggesting independent contributions of B and
124 CD4 T cell OCRs to MS GWAS heritability (**Figure 2A, Supplemental Table 2**).
125
126 To further delineate the cell types with OCRs that specifically mediate the effect of MS GWAS
127 results, we performed a series of pairwise LDSC analyses. In brief, OCRs of a given
128 hematopoietic cell type were stratified against the OCRs of each of the other 15 cell types, as
129 well as the LDSC baseline annotations (**Figure 2B; Supplemental Table 3**). As above with the
130 joint model, we report the p-value of the coefficient τ_c. We observed that B cells remained
131 significant even after stratifying on OCRs of any of the other 15 cell populations (coefficient p-
132 value ranging from 1.47×10^{-12} when stratifying against HSCs, to 8.33×10^{-5} when stratifying
133 against CD4 T cells). This was also the case for CD4 T cells, which remained significant after
134 stratifying on OCRs from any of the other 15 cell populations (coefficient p-values ranging from
135 3.82×10^{-17} when stratifying against HSCs, to 2.39×10^{-4} when stratifying against CD8 T cells). In
136 contrast, CD8 T cell OCRs were no longer significant after stratifying against OCRs from $CD4^+$
137 T cells (coefficient p-value = 0.21), though they were significant when stratifying on any of the
138 other cell populations. NK cells were also no longer significant after stratifying against either
139 CD4 T cells (coefficient p-value = 0.165) or against CD8 T cells (coefficient p-value: 0.356).
140 These results indicate that the enrichment of both CD8 T cells and NK cells can be largely
141 explained by shared regulatory landscapes that are also present in CD4 T cells. Prior studies
142 have also suggested a role for monocytes in MS[24,25]. OCRs from monocytes had an enrichment
143 p-value of 4.17×10^{-9}, but we observed that stratifying on OCRs from CD4 T cells or B cells
144 ameliorated this monocyte heritability enrichment (coefficient p-values 0.166 and 0.266,
145 respectively).
146

147 We also performed a separate analysis examining whether OCRs specific to a given cell type
148 mediate MS heritability enrichments. For each of the mature hematopoietic cell populations, we
149 performed LDSC on only the cell type-specific OCRs, i.e. ATAC-seq peaks present in only that
150 cell type. B cells exhibited a statistically significant enrichment of cell-specific OCRs for MS
151 GWAS (enrichment p-value = 3.27×10^{-4}). CD4 T cell-specific peaks were nominally significant at
152 a p-value of 0.013 but did not survive correction for multiple hypothesis testing (Bonferroni
153 corrected p-value threshold of 5.6×10^{-3}). Cell type-specific peaks for all other terminal
154 hematopoietic cell types had enrichment p-values > 0.05 (**Figure S2; Supplementary Table 4**).
155

156 ***MS genetic associations are mediated in terminal immune cell populations***

157 In the lymphoid lineage, we observed stronger enrichments in terminal cell populations than we
158 did for the progenitor populations (**Figure 1B**). For example, the strongest enrichment in
159 progenitor cells was observed for common lymphoid progenitor cells (CLP, enrichment p-value:
160 2.32×10^{-4}), and it was orders of magnitude less statistically significant compared to the
161 enrichment observed for CD4 T or B cells (**Figure 1B).** For each of the terminal populations, the
162 significance remained even after stratifying against OCRs from CLP; however, the converse
163 was not true (**Figure 2B**). The enrichment in CLP was completely ameliorated by stratifying
164 against B cell or CD4 T cell OCRs (coefficient p-value 0.98 and 0.966, respectively, **Figure 2B**).
165 Together, these results suggest that terminal cells of the lymphoid compartment retain cellular
166 regulatory features from their progenitor populations that are important for MS pathogenesis, but
167 have also developed specific regulatory features of additional importance to MS susceptibility.
168

169 ***Comparison of immune cell enrichment with neuropsychiatric and autoimmune disorders***

170 We next sought to understand how the immune cell enrichments in MS might be similar or
171 different from those of other autoimmune or neuropsychiatric disorders. To test this, we

172 calculated heritability enrichment within these 16 hematopoietic OCRs for GWAS of various
173 other neuropsychiatric disorders and autoimmune disorders: Alzheimer disease (AD)[26],
174 schizophrenia (SCZ)[27], bipolar disorder (BPD)[28], type 1 diabetes (T1D)[29], Crohn's disease
175 (CD)[30], ulcerative colitis (UC)[30], systemic lupus erythematosus (SLE)[31], rheumatoid arthritis
176 (RA)[32], and primary biliary cirrhosis (PBC)[33] (**Figure 3A; Supplemental Table 5**). We identified
177 previously recognized cell-type enrichments across these other diseases, such as enrichments
178 in OCRs from B cells (enrichment p-value: 1.93×10^{-7}), CD4 T cells (p-value: 7.34×10^{-8}) and CD8
179 T cells (p-value: 2.09×10^{-6}) for RA[15,32]. However, the heritability enrichments for OCRs from
180 these hematopoietic cell populations tended to be much stronger for MS, despite similar sample
181 sizes, e.g. 41,505 disease cases for MS and 38,242 disease cases for RA. To parse out cell
182 type-specific enrichments in these other disorders, we also tested heritability enrichments under
183 the joint model in LDSC by including OCRs from all 16 cell types (**Figure 3B, Supplemental**
184 **Table 6**). In a similar fashion to the MS GWAS joint analyses above, we measured the
185 contribution to heritability for a given set of OCRs of interest, controlling for the effects of a set of
186 baseline annotations and OCRs from all other hematopoietic cell types. These analyses were
187 performed separately for each disease. Across the nine other comparator diseases, the only
188 other statistically significant stratified enrichments were in OCRs from B cells in SLE GWAS
189 (coefficient p-value: 8.53×10^{-5}). We also identified other enrichments in this joint model that were
190 nominally significant, including signals for several diseases in either B or CD4 T cells. This is in
191 contrast to our MS findings, which show strong enrichments in both CD4 T cells and B cells,
192 highlighting a dual role of these cell types in MS and distinguishing it from other diseases where
193 the cell types are more restricted to specific lineages.
194

The CD4 T cell MS GWAS enrichment is driven by the Th17 subset

To further focus in on the MS GWAS enrichment observed in OCRs from CD4 T cells, we utilized ATAC-seq data from various subsets of human CD4 T cells (**Figure 4A**)[21]. We examined data from effector CD4 T cells (naïve effector CD4 T cells, T_h1, T_h2, T_h17, and follicular T_h) as well as, regulatory CD4 T cells (naïve T_{regs} and memory T_{regs}). We observed strong enrichments for OCRs from both effector and regulatory CD4 T cell populations (**Figure 4B; Supplementary Table 7**). Next, to identify the independent contribution of a given cell type, we applied the joint model in LDSC by including OCRs from all CD4 T cell populations together. This joint LDSC analysis revealed that OCRs from T_h17 cells independently contributed to heritability (coefficient p-value = 4.69×10^{-4}; **Figure 4C; Supplemental Table 8**). Performing pairwise stratified LDSC confirmed the independent contribution of OCRs in T_h17 cells to MS GWAS heritability **(Figure 4D; Supplemental Table 9**). OCRs from T_h17 remained enriched in MS GWAS even after stratifying against any of the other T_{eff} cell populations or T_{reg} cell populations. Conversely, the enrichments for all other T_{eff} cell populations or T_{reg} cell populations were ameliorated when stratifying against T_h17 cells.

Based on the LDSC analyses in the joint model, we note that enrichments tended to be stronger in OCRs from the memory effector CD4 T cell populations than from naïve effector cells. For each of these memory effector CD4 T cell populations, OCRs retained statistical significance even after stratifying against OCRs from naïve T_{eff} cells: T_h1 (coefficient p-value after stratifying against naïve T_{eff} cells: 8.22×10^{-6}), T_h2 (p-value = 1.45×10^{-4}), T_h17 (p-value = 7.27×10^{-7}) or follicular T_h cells (p-value = 1.77×10^{-7}) (**Figure 4D**). The converse was not true; the statistical significance of the naïve T_{eff} cell OCRs was completely lost when stratifying against OCRs from any of the memory T_{eff} cell populations. These results suggest that among CD4 T cells, OCRs from T_h17 cells drive the signal for enrichment in MS GWAS.

221 ***Memory subpopulations explain the enrichment of MS GWAS in B cells***

222 We next examined enrichments of MS GWAS data in ATAC-seq from the B cell lineage
223 including naïve B cells, memory B cells, and plasmablasts (**Figure 5A**) [21]. We found that OCRs
224 from all B cell lineage cell types were significantly enriched for MS heritability (**Figure 5B;**
225 **Supplemental Table 10**). LDSC under a joint model including all B cell lineage cell types
226 revealed an independent contribution from memory B cells (coefficient p-value = 1.10×10^{-3}),
227 but not naïve B cells or plasmablasts **Figure 5C; Supplemental Table 11**). Pairwise stratified
228 LDSC confirmed the independent enrichment of OCRs from memory B cells. Memory B cell
229 OCRs remained statistically significant even after stratifying on naïve B cells (coefficient p-value
230 = 1.26×10^{-5}) or plasmablasts (coefficient p-value: 3.29×10^{-4}; **Figure 5D; Supplemental Table**
231 **12**). In contrast, OCRs from naïve B cells and plasmablasts were no longer statistically
232 significant when stratifying on memory B cells OCRs (coefficient p-value: 0.71 and 0.27,
233 respectively). These results suggest that in the B cell lineage, the MS GWAS enrichment signal
234 is driven by OCRs in memory B cells.
235

236 ***Immune cells from MS patients reinforce independent CD4 T and B cell enrichments***

237 Next, we tested whether the MS GWAS enrichment in CD4+ T and B cells were also present in
238 OCRs from the respective immune cell types derived from individuals with MS. We utilized
239 ATAC-seq data performed in flow-sorted bulk CD4 T and B cell subsets (n=6) derived from six
240 patients with MS who were not treated with immunomodulatory therapy within at least 4 months
241 of sample collection (see **Supplementary Table 13** for clinical details). LDSC showed
242 statistically significant enrichments of transitional B cells (traB; enrichment p-value=2.08×10^{-3}),
243 class switched classical memory B cells (cMBc; p-value = 2.58×10^{-4}), effector memory CD4 T
244 cells ($T4_{em}$; p-value = 1.87×10^{-4}), and CD45RA+ effector memory CD4 T cells (T4ra; p-value =
245 3.02×10^{-4}). Central memory CD4 T cells had an enrichment p-value of 1.53×10^{-3}, which was not

246 significant after correcting for multiple hypothesis testing (**Figure 6A; Supplementary Table**
247 **14**).
248
249 To further identify independent cell type enrichments, we performed joint models in LDSC as
250 described above. We first tested a model that included CD4 T cell subsets (T4nv, T4cm, T4em,
251 and T4ra). In this joint model, only T4em was statistically significant (coefficient p-value 7.45 x
252 10^{-3}), reflecting the independent contribution of effector CD4 T cells in MS that we observed
253 above using cells from healthy individuals (**Figure 6B; Supplementary Table 15**). We also ran
254 a model that included B cell subsets (traB and cMBc). In this joint comparison, neither cell type
255 was statistically significant when correcting for multiple hypothesis testing. However, cMBc had
256 a coefficient p-value that was nominally significant (p-value=0.037), reiterating the independent
257 contributions of mature B cell types (**Figure 6C; Supplementary Table 16**).
258
259 ***Immunomodulatory treatments suppress mediation of MS genetics in cell-specific fashion***
260 Next, we tested whether immunomodulatory treatments alter the cell-specific mediation of MS
261 genetic associations by utilizing data for the same immune subsets sorted from patients with MS
262 (n=3) under treatment with either natalizumab, interferon, or glatiramer acetate. Following
263 treatment with any of the agents still resulted in statistically significant enrichments of cMBc and
264 T4em (**Figure 6D; Supplementary Table 17)**, though the magnitude of the enrichments were
265 attenuated relative to the signals observed from cells from untreated patients (compare **Figure**
266 **6A and 6D).** To better understand this attenuation of enrichments, we ran joint models in LDSC
267 for T4em and cBMc cells in which we included OCRs from untreated patients and treated
268 patients. In a joint model with T4em OCRs from treated and untreated MS patients, only OCRs
269 from untreated patients were statistically significant (**Figure S3A; Supplementary Tables 18**).
270 Similarly, in a joint model with cMBc OCRs from treated and untreated MS patients, only OCRs

from untreated patients were statistically significant (**Figure S3B; Supplementary Table 19**). Together, these results suggest that immune-modulating therapies may attenuate the chromatin accessibility signals at MS GWAS.

MS GWAS signals in B and CD4 T cells driven by active enhancer and promoter regions

We next sought to gain further insight into the functional consequences of the B and CD4 T cell OCRs underlying MS GWAS signals. We examined enrichments for MS GWAS in chromatin immunoprecipitation sequencing (ChIP-seq) peaks from various histone modifications (H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me3, and H3K9me3) [34,35]. In T_h17 cells (**Figure S4A; Supplementary Table 20**), we detected statistically significant enrichments for H3K27ac (enrichment p-value = 6.54×10^{-9}), H3K4me1 (p-value = 3.96×10^{-19}) and H3K4me3 (p-value = 2.29×10^{-8}). Similarly, in B cells (**Figure S4B; Supplementary Table 21**), we also detected significant enrichments for H3K27ac (enrichment p-value = 1.18×10^{-11}), H3K4me1 (p-value = 4.96×10^{-16}) and H3K4me3 (p-value = 3.93×10^{-8}). These results suggest that MS genetic association are primarily enriched at active noncoding elements: primed enhancers (H3K4me1), active enhancers (H3K27ac and H3K4me1) and active promoters (H3K4me3).

To further delineate the chromatin states with the strongest MS genetic associations, we examined enrichments of MS GWAS results in the predicted chromatin states for B cells and T_h17 T cells as available in the RoadMap Epigenomics Project [36]. For Th17 CD4 T cells, the "EnhA2" chromatin state (Active Enhancer 2) were statistically enriched (enrichment p-value=1.19×10^{-3}) (**Figure S5A; Supplementary Table 22)**. For B cells, "Tx3" (Transcribed 3' preferential; enrichment p-value=1.80×10^{-3}) and "PromD1" (Promoter Downstream TSS 1; p-value=4.40×10^{-4}) were statistically enriched, again reflecting the strongest enrichments at active regulatory elements (**Figure S5B; Supplementary Table 23)**. These results demonstrate that

296 MS GWAS variants act through activating regulatory elements, consistent with the autoimmune
297 nature of MS.
298

299 ***Fine-mapping of MS GWAS loci in cell-specific OCRs***

300 We next sought to understand underlying mechanisms by nominating putative causal genes and
301 variants in a cell-specific fashion. First, we applied statistical fine-mapping to nominate likely
302 causal SNPs. Most statistical fine-mapping approaches require association information across
303 all SNPs in a given locus. In contrast, for MS GWAS, genome-wide results are based on
304 targeted replication analyses, which by design included only a select subset of SNPs within
305 each locus. To overcome this challenge, we applied PICS to perform statistical fine-mapping,
306 which as compared to other statistical fine-mapping approaches, does not require GWAS
307 summary statistics for all SNPs in a region[15]. We defined for each locus a 95% credible set such
308 that the sum of the posterior probabilities for variants in that credible set is greater than or equal
309 to 95%. For fine-mapping, we used the joint analysis MS GWAS results, which include only
310 replicated genome-wide effects[8]. We included all 200 non-MHC loci where the MS GWAS joint
311 association p-value was less than 5×10^{-8}. Next, we prioritized SNPs if they had a PICS
312 posterior probability (PP) > 1% and were included in the 95% PICS credible set. This is a liberal
313 threshold for defining prioritized SNPs, aimed at increasing sensitivity for detecting possible
314 causal variants. Across the 200 loci, there were 3436 credible set variants (1-58 variants per
315 locus) (**Figure S6A**). At 19 loci, there was only one variant in the credible set, and at 37 loci,
316 there were four or fewer prioritized variants (**Figure S6B)**.
317

318 Next, we intersected the 3436 credible set variants with OCRs from the 16 hematopoietic cell
319 populations, which we chose to use as they cover a broad range of hematopoietic cell types[18].
320 Across the 200 loci, 870 of the prioritized variants overlapped an OCR in at least one cell type.

Remarkably, 163 out of the 200 loci had at least one prioritized variant overlapping an OCR in any cell type. B and CD4 T cells were the cell types with the greatest number of loci with a credible set SNP overlapping an OCR,126 and 125 loci respectively **(Figure S7**).

Integration of genetic and epigenetic data identifies putative causal genes

The vast majority of GWAS-associated variants are noncoding and regulate genes that may be far from the variant in terms of linear distance on the genome. To address this challenge and nominate putative causal genes that are regulated by the MS-associated OCRs, we leveraged promoter capture Hi-C data (PCHiC), which identifies chromatin looping interactions between regulatory elements and target gene promoters. We utilized PCHiC data performed on 17 hematopoietic cell populations, which partially overlap with the cell types for which we have ATAC-seq data[37]. These cell populations include naïve B cells, total B cells, activated CD4 T cells, non-activated CD4 T cells, and total T cells. We considered only GWAS loci where a credible set MS GWAS SNP overlapped both an OCR and a PCHiC interaction.

Through these analyses, we nominated 261 genes within 86 MS loci in B cells and 364 genes within 115 MS loci in CD4 T cells (**Supplemental Table 24).** We note that these genes are putative causal genes, representing a list of genes that could be linked with the MS loci via a regulatory mechanism in the respective cell types. The majority of these genes were shared between B and CD4 T cells (n=178; 68.2% and 48.9% respectively; **Figure 7A**). We have previously suggested a list of putative causal genes (n=551) based on an ensemble of methods that did not include ATAC-seq or chromatin interactions[8]. Of these 551 previously nominated genes, 67 (12.2%) overlapped with our B cell prioritized genes and 111 (20.1%) with our CD4 T cell prioritized genes, highlighting that our current mechanism-specific gene prioritization is

capturing a large number of potentially causal genes that have not been previously implicated in MS genetic studies.

Next, we utilized the list of putative causal genes to identify enriched canonical pathways. Starting with CD4 T and B cell lists, we created additional lists for the common genes (shared between CD4 T and B cells), and finally genes unique to CD4 T cells and B cells. We observed widespread pathway enrichment for the putative causal genes of the CD4 T cells (n=294) and B cells (n=236) at FDR<5%. The common set of genes was enriched in 85 pathways (**Supplemental Table 25)**. The B unique gene list (n=83) was enriched in 22 canonical pathways, including lipoprotein and cholesterol pathways, the CD40 pathway, and JAK-STAT pathway (**Figure 7B-C).** The unique genes in CD4 T cells (n=186) were enriched in 99 pathways, including TCR pathways, various interleukin pathways, and MAPK/ERK signaling pathways (**Figure 7B-C)**.

Pathway analyses utilize known biological connections for a given set of genes but many of underlying mechanisms could be still uncharacterized. Thus, we leveraged protein-protein interaction (PPI) data to test whether the respective putative causal gene lists exhibit a high degree of connectivity[38]. A similar percent of the mapped CD4 T cell and B cell prioritized genes were directly connected, 48.9% and 43.7% respectively (**Supplementary Table 26; Figures S8-12)**. Only the CD4 T and CD4 T unique gene lists demonstrated a higher degree of connectivity than expected (p-value<0.05), although all gene lists had communities of genes with high connectivity (p-value<0.05; **Supplementary Table 26**). These results are consistent with the pathway analyses, implying that the MS genetics are mediated by several different mechanisms in both cell types, some of which are shared and some of which are cell-type specific.

370

371 ***A fine interplay between shared and cell-specific B and CD4 T cell putative causal genes***

372 The results of the pathways and PPI analyses suggest that the MS genes common and unique
373 to B and CD4 T cells do not act independently but rather have shared cellular mechanisms. We
374 illustrate this by studying a locus on chromosome 19 (**Figure 8**) where the lead SNP rs1465697
375 (chr19:49837246 C>T) has a MS GWAS association p-value of 3.02×10^{-18}. The lead SNP is the
376 most highly prioritized variant by statistical fine-mapping out of 21 overall SNPs in the 95%
377 credible set (PP=15% for rs1465697; next highest PP 9%**).** This SNP overlies an OCR present
378 in all lymphoid lineages, including B cells, CD4 T cells, and CD8 T cells. We have previously
379 suggested five putative causal genes for this locus: *DKKL1*, *CCDC155, CD37*, *TEAD2*, and
380 *SLC6A16* [8]. Using PCHiC data, this OCR forms a chromatin loop interaction with *TEAD2* and
381 *DKKL1*, but not the other three genes. This chromatin loop interaction is observed in activated
382 CD4 T cells, naïve B cells, total B cells, naïve CD8 T cells, and fetal thymus, but none of the
383 other hematopoietic cell types [37]. Furthermore, this SNP is an eQTL for *TEAD2* in B cells, but is
384 not an eQTL for any other gene in this locus in any of the available hematopoietic cell types
385 (**Figure 8**)[39]. Together, these lines of evidence support *TEAD2* as the causal gene at this locus.

386

387 Interestingly, *TEAD2* is a transcription factor with 1459 predicted regulated genes [40], including
388 38 putatively causal CD4 T cell genes and 25 B cell genes as nominated above (FDR < 1 %,
389 FDR < 1%, respectively; **Figures S13-14**). The majority of these genes are common in both
390 CD4 T and B cells (n=23; FDR < 1%; **Figure S15**). To identify genes whose expression is
391 modulated by T*EAD2* in CD4 T and B cells, we examined *TEAD2* knock-down (KD) and over-
392 expression (OE) in cancer cell lines (n=8, respectively) from the Library of Integrated Network-
393 Based Cellular Signatures (LINCS) Program [41]. Although these cell lines do not represent an
394 ideal experimental model to study the effect of *TEAD2* in immune cells, they can still be used to
395 understand mechanisms reflecting core cellular functions. Within each cell line we identified

396 genes whose expression changed in the opposite direction (top or bottom 10%) in the KD
397 versus OE models. These ranged from 7 to 24 for the CD4 T cell prioritized genes
398 (**Supplementary Table 27; Figure S16**) and from 3 to 16 for the B cell genes (**Supplementary**
399 **Table 28; Figure S17)**. Of these, 17 genes and 9 genes changed expression in at least 2 cell
400 lines, respectively (**Supplementary Tables 27-28)**. These data demonstrate that perturbation of
401 *TEAD2*, a key immune cell transcription factor, results in indirect changes of putative MS causal
402 genes in B and CD4 T cells.
403
404 **Discussion**
405 In this paper, we integrated MS GWAS with chromatin accessibility data from a broad array of
406 peripheral immune cells in order to identify putative causal cell types. Our analyses identified
407 regulatory regions in B cells and CD4 T cells as each being independently enriched for MS
408 genetics. Within the CD4 T cell and B cell populations, we further identified OCRs from T_h17
409 cells and memory B cells as specifically driving their respective enrichments. Chromatin data
410 from MS patients reiterated these findings and further suggested that immunomodulatory
411 treatments alter the chromatin accessibility overlying MS-associated GWAS variants. Integration
412 of PCHiC data led to prioritization of putative causal genes in B and CD4 T cells, identifying
413 target genes that are both shared and specific to each of these cell populations. The putative
414 causal genes implicate several known signaling pathways, mostly due to the cell-specific MS-
415 associated genes, despite these representing a smaller percentage compared to genes shared
416 between B and CD4 T cells. Finally, we illustrate that the B and CD4 T cells mechanisms are
417 intertwined by describing how *TEAD2,* a putative causal gene shared between B and CD4 T
418 cells, contributes to disease susceptibility directly and indirectly by targeting both shared and
419 cell-specific MS genes.
420

Our study provides genetic evidence for the independent involvement of both CD4 T and B cells in the pathogenesis of MS [1–3]. It further supports the long-debated causal role of memory B cells in MS[24,42,43], shown by the highly effective therapies targeting B cell receptors, most notably ocrelizumab[7]. Through our analyses, we also corroborate decades of research that have demonstrated a primary role of CD4 T cells in MS, including the importance of T_h17 T cells [24,44,45]. For both the B cell and T_h17 T cell populations, we find that active enhancers and promoters drive the enrichment signal, consistent with activating roles of these cell types in MS as an autoimmune condition. The complex interplay between the B and CD4 cells in MS pathogenesis [2,3,42] is also reflected by our finding of some OCRs present in both cell types, while other OCRs are present only in specific cell populations.

Although other peripheral immune cell types have been shown to be involved in MS, including monocytes, CD8 T cells, and mDCs[24], we did not identify independent enrichments for these cell types. One possible explanation for this apparent contradiction is that these cell types might work secondarily to memory B and T_h17 cells, which are more directly under influence from MS GWAS-associated variants. Non-genetic effects, e.g. environment-specific response, could also explain their role in MS above and beyond any shared mechanisms with B and CD4 T cells. Further, context-specific studies, such as under various cell activation conditions, would be necessary to unravel any potential independent influence of these cells by MS genetic variants. Lastly, while our analyses do not identify an independent genome-wide enrichment for cell types other than B and CD4 T cells, GWAS variants may still act at individual loci in these other cell types.

One of the key challenges of GWAS is moving from genetic association to biological mechanisms[10,13,14]. This is driven by three main challenges. First, linkage disequilibrium, while highly advantageous to discovery of genetic associations, limits our ability to identify the causal

447 variant. Second, as most GWAS variants are noncoding, identifying the gene(s) that are
448 affected by the causal variant can be difficult. Third, the cell type(s) in which a given associated
449 variant acts can be unclear. Our study demonstrates how we can use statistical fine-mapping to
450 help solve the first challenge, though this is not without multiple caveats[46]. We integrated
451 orthogonal datasets (ATAC-seq, PCHiC) to help delineate the likely causal genes and cell types
452 and overcome the latter two challenges. We document how shared and cell-specific genes
453 affect putative causal pathways. We further illustrate the complex interplay between shared and
454 cell-specific putative causal MS genes by studying the *TEAD2* locus, a transcription factor
455 recently implicated in immune regulation[47]. Together, our study generates important insights into
456 the driver subpopulations of peripheral immune cells in MS, reinforcing how MS genetics act
457 primarily through B and CD4 T cells. Our study also demonstrates the need for in-depth context-
458 specific cellular data to carefully delineate the causal role of each immune cell subset in MS.
459
460 **Methods**
461 *GWAS summary statistics*
462 We utilized available MS GWAS summary statistics, which included data from 8,278,136
463 variants across 14,802 individuals with MS (cases) and 26,703 individuals without MS
464 (controls) [8]. For enrichment analyses, we included only the 6,773,531 variants that were
465 analyzed in all 15 cohorts of the discovery stage meta-analysis; this resulted in 6,773,531
466 variants carried forward. We additionally utilized GWAS summary statistics from various
467 neuropsychiatric disorders or autoimmune disorders: Alzheimer disease (AD)[26], schizophrenia
468 (SCZ)[27], bipolar disorder (BPD)[28], type 1 diabetes (T1D)[29], Crohn's disease (CD)[30], ulcerative
469 colitis (UC)[30], systemic lupus erythematosus (SLE)[31], rheumatoid arthritis (RA)[32], and primary
470 biliary cirrhosis (PBC)[33]. All GWAS data were converted to the ".sumstats" format as required by
471 LDSC using the "munge.py" function in LDSC with default parameters.

Epigenetic and eQTL datasets

Hematopoietic progenitor and terminal ATAC-seq data: ATAC-seq data for 16 different human hematopoietic progenitor and terminal populations was obtained from Corces et al.[18] and Buenrostro et al.[19]. These ATAC-seq profiles were generated on bulk FACS-sorted cells from human peripheral blood or bone marrow cells. Alignment of the ATAC-seq data and peak-calling were performed as previously described[20]. To identify cell-type specific peaks for each cell type, we used ATAC-seq peaks for that cell type and removed any peaks that overlapped with a peak present in any one of the other 15 cell types. A single base pair overlap was considered to be overlapping.

Immune cell ATAC-seq data: We used publicly available immune cell ATAC-seq data (NCBI GEO GSE 118189) derived from flow-sorted peripheral blood cells [21]. As each cell type had between 1-4 human donors, we merged the raw ATAC-seq data from the individual donors for a given cell type. We aligned ATAC-seq reads using bowtie2 version 2.2.1[48] with default parameters and a maximum paired-end insert distance of 2000 base pairs. The bowtie2 index was constructed with the default parameters for the hg19 reference genome. We filtered out reads that mapped to the mitochondria and used samtools version 1.10[49] to filter out reads with MAPQ < 30 and with the flags '- F 1804' and '-f 2'. Additionally, duplicate reads were discarded using picard version 2.20.6 (http://broadinstitute.org.github.io/picard). Finally, chromatin accessibility peaks were identified with MACS2 version 2.1.1[50] under default parameters and '--nomodel --nolambda --keep-dup all --call-summits'.

ChIP-seq data: We downloaded available pre-processed ChIP-seq peak calls from ENCODE for B cells[34]. Where replicates were available, the bed files for the replicates were merged to create a composite set of peaks for each histone mark. Data for T_h17 histone ChIP-seq were

498 downloaded from the Roadmap Epigenomics Project[35] (NCBI GEO GSM997225); we used pre-
499 processed ChIP-seq peak calls generated in Amariuta et al[51].
500
501 *chromHMM:* We used a 25 chromatin state model[36], which are imputed based on 12 epigenetic
502 marks from across 127 epigenomes generated as part of the Roadmap Epigenomics Project[35].
503 We used the chromatin states from B cells and Th17 CD4+ T cells. Chromatin states were
504 downloaded from https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html. We
505 excluded the "Quiescent/Low" cell state, as it encompasses a large proportion of the genome,
506 resulting in unstable estimates of heritability.
507
508 *Verily ATAC-seq data:* We utilized immune cell ATAC-seq data generated from Verily as part of
509 the SysteMS collaboration with Brigham & Women's Hospital (PIs: Dr. Chitnis and Dr. Weiner).
510 Cell sorting: Frozen cryovials in liquid nitrogen were thawed in a 37℃ bead bath and centrifuged
511 for 5 min at 600 x g, 4℃. The cell pellet was washed with 1 mL of FACS buffer, and the wash
512 repeated. The cell pellet was resuspended in the residual volume with 2.5 uL of 0.33mg/mL S7
513 DNAse. 50 uL of staining cocktail was added for the respective flow cytometry panels to be
514 analyzed (T cell, B cell, myeloid panel) and incubated for 25 min on ice and in the dark. Cells
515 were washed in FACS buffer, resuspended in a final volume of 400 uL FACS buffer, and passed
516 through a 35 µm cell strainer cap. Stained samples were sorted on a FACSAria Fusion (BD
517 Biosciences, San Jose, CA). Using FACSDiva v8.0.1 software, the samples were gated first by
518 forward and side scatter properties, then FSC-H vs FSC-A for singlet discrimination, and finally,
519 with their respective markers for each cell type (**Supplementary Table 29**). For each cell type
520 of interest, up to 500 cells were sorted into the tagmentation buffer.
521 ATAC-seq library preparation and sequencing: Cells were sorted directly into 20 uL of cold
522 tagmentation buffer (10 uL TD, 2 uL 2% IGEPAL CA-630, 6 uL nuclease-free H2O, 2 uL TDE1
523 per sample), followed by incubation at 37℃ for 30 min with shaking at 500 RPM. Samples were

stored at -20℃ until further processing. DNA was extracted with the QIAGEN MinElute PCR purification kit according to the manufacturer's protocol, and samples were amplified with KAPA HiFi kits and Illumina Nextera indices. The amplified material was cleaned with the QIAGEN MinElute PCR purification kit and quantified using KAPA library quantification kits. Samples were normalized and pooled for sequencing on the NextSeq (Illumina).

Processing: Paired-end raw ATAC-seq reads were trimmed using NGmerge [60] using the default parameters. The reads were then aligned to GRCh38 using Bowtie2 version 2.3.5 [49]. The resulting SAM files were converted and sorted into BAM format using samtools version 1.5 [50]. We filtered out the reads with MAPQ<10 and reads that were aligned to mitochondria using samtools. In addition, duplicate reads were removed using picard version 2.20.6 (http://broadinstitute.org.github.io/picard). Finally, peaks were called using MACS2 version 2.2.5 [51] with default parameters and --keep-dup all --nomodel –nolambda. To obtain peaks for each cell type, we merged the peak files from all samples for that specific cell type using bedtools version 2.29 [61].

Enrichment of GWAS results within ATAC-seq peaks

To calculate enrichments of the MS GWAS data within annotations (e.g., ATAC-seq or ChIP-seq peaks), we applied stratified LD SCore regression (LDSC)[16,23]. LDSC was performed using LDSC v1.0.0 (https://github.com/bulik/ldsc), which was run on the discovery summary statistics from the MS GWAS discovery stage summary statistics[8]. The human MHC locus was excluded given its complex LD patterns as recommended by Finucane et al.[16]

To run LDSC, we used precomputed LD scores based on the European ancestry samples of the 1000 Genomes Project Phase 1[52] which was restricted to HapMap3 SNPs[53], and we generated partitioned LD scores for each set of annotations. To perform LDSC, we regressed the summary statistics ($\chi 2$) from a given GWAS on to annotation-specific LD scores, with baseline scores

(original 53 annotation model), regression weights and allele frequencies based on 1000 Genome Project Phase 1 data as precomputed by software authors. We applied partitioned heritability analyses using LDSC under three different models:

1) To ask how much a given annotation contributes to trait heritability, we used a LDSC model that includes baseline annotations and an annotation of interest. The heritability enrichment of the annotation was defined as the proportion of SNP heritability in the category divided by the proportion of SNPs in that category; we report statistical significance of this enrichment as p-values.
2) When comparing multiple annotations (e.g., ATAC-seq peaks from different cell types), we ran a LDSC model that includes the baseline model and annotations from all cell types. In this scenario, we calculate for each annotation the coefficient τ_c which measures the contribution to SNP heritability for a given annotation to heritability in this overall model, stratified on other annotations in the model. Z-scores for the coefficient τ_c were converted to a one-sided p-value, which we report as a measure of statistical significance.
3) We also performed LDSC on pairs of annotations, which we term pairwise stratified LDSC. In these models, we include the baseline model, an index annotation of interest, and a comparator annotation of interest. To run LDSC, we used a previously described extension of LDSC [54].

Throughout, all default LDSC parameters were used.

Statistical fine-mapping

Statistical fine-mapping was performed using the marginal p-values from the replication (joint analysis) summary statistics from the MS GWAS [8]. The 200 genome-wide significant loci at $p<5x10^{-8}$ were used. LD was calculated between each lead variant and all variants with $r^2>0.2$ and within a 2 Mb window based on the 1000 Genomes Phase 1 (European subset) reference

575 panel[52]. PLINK v1.90b3.32 was used to perform LD calculations[55,56] with parameters of '--r2 --ld-
576 window-kb 2000 --ld-window 999999 --ld-window-r2 0.2'.
577
578 We then applied PICS to each locus[15]. Briefly, PICS uses the lead association p-value and LD
579 structure of the locus to calculate the most likely causal SNPs given the observed lead
580 association signal. PICS probabilities represent the probability of a given SNP in a locus being
581 the causal SNP. Default PICS parameters were used. From the PICS probabilities, we
582 calculated 95% credible sets (CS). We defined the 95% CS as a set of variants such that the
583 true causal variant has a 95% chance of being in the credible set. To calculate credible sets, for
584 each locus, we ranked variants in descending order by their PICS probabilities. We then
585 iteratively added variants to the credible set for that locus until the sum of their PICS
586 probabilities was greater than or equal to 0.95. For CS inclusion, we also required the variant to
587 have a PICS probability > 0.1.
588
589 *Identification of target genes*
590 We leveraged promoter capture Hi-C (PCHiC) data from 17 hematopoietic cell populations to
591 link genetic associations with genes that they may regulate [37]. We filtered the PCHiC dataset for
592 looping interactions with a CHiCAGO score > 5[57]. An overlap between a GWAS variant and a
593 PCHiC looping interaction was considered if the GWAS variant overlapped any position in the
594 non-promoter ("other end") of the PCHiC interaction. For $CD4^+$ T cells, we considered only
595 GWAS SNPs that overlapped an ATAC-seq peak in bulk $CD4^+$ T cells or any of the $CD4^+$ T cell
596 subsets (naïve effector $CD4^+$ T cells, T_h1, T_h2, T_h17, follicular T_h, naïve T_{regs} and memory T_{regs}),
597 and which overlapped a PCHiC interaction in naïve $CD4^+$ T cells (nCD4), total $CD4^+$ T cells
598 (tCD4), non-activated total $CD4^+$ T cells (naCD4), or activated total $CD4^+$ T cells (aCD4). For B
599 cells, we considered only GWAS SNPs that overlapped an ATAC-seq peak in bulk B cells or

any of the B cell subsets (naïve B cells, memory B cells, or plasmablasts), and which overlapped a PCHiC interaction in naïve B cells (nB) or total B cells (tB).

Correction for multiple hypothesis testing

Throughout our manuscript, we use Bonferroni corrections when testing multiple hypotheses. To generate a Bonferroni-corrected p-value threshold, we used a traditional p-value threshold of 0.05 divided by the number of tests being performed in a given analysis. We note that as many of the tests are correlated (since the underlying annotations are often highly correlated with each other), the effective number of independent tests being performed is fewer than the number of tests actually performed. As such, our analyses are overly conservative. We decided on this approach of using Bonferroni corrections as opposed to false discovery rate (FDR) approaches, as the number of tests being performed is often small, leading to unstable estimates of FDR.

Gene set enrichment analyses

We performed pathway analyses utilizing the canonical pathways (CP) of the Molecular Signatures Database (MSigDB v7.2), as it is available from the Gene Set Enrichment Analysis website (http://software.broadinstitute.org/gsea/msigdb). We ran the Canonical Pathways, Biocarta, KEGG, and Reactome gene sets categories together in the same model. We estimated statistical significance using the hypergenometric distribution and applied false discovery correction, as previously described[8]. The same model was applied for the enrichment of prioritized gene sets with Gene Transcription Regulation Database (GTRD) transcription factor targets gene sets[40]. Significant enrichment level was set to a false discovery rate < 5%.

Protein-protein interaction networks

We utilized GeNets (https://apps.broadinstitute.org/genets)[38] to leverage known protein-protein interactions (PPI) of our prioritized gene sets. GeNets uses a random forest classified, trained in PPI data with 18 parameters that capture information about centrality and clustering. It creates communities of genes, sets of genes (nodes) that are connected to each other more than genes outside this community. Furthermore, it uses the random forest classifier and the connectivity to the tested gene set to propose candidate genes. For each described network the p-value is estimated by testing whether the number of observed edges divided by the numbers of possible edges using permutations. We ran GeNets via the web interface with the GeNets Metanetwork v1.0 and utilizing the InWeb model ("Override network the analysis model was trained on" option).

TEAD2 knockdown and over-expression in LINCS cell lines

Robust z-scores, "level 5 data", from knock-down (KD) or over-expression (OE) of *TEAD2* in cancer cell lines[41] were downloaded from clue.io (https://clue.io/command?q=/sig%20%22TEAD2%22). The robust z scores represent differential expression for each genetic perturbagen, adjusted for the gene expression of all other perturbagens on the same physical plate. For knockdown and over-expression experiments the differential expression comparator were samples using a vector control, which are negative genetic controls that either lack a gene-specific sequence or target a non-human gene (like GFP).

Data availability

MS GWAS summary statistics are available via request to the IMSGC (https://imsgc.net/). LD score regression software and reference panels were obtained from software developers (https://github.com/bulik/LDSC). Processed GWAS summary statistics for diseases other than MS were obtained from https://alkesgroup.broadinstitute.org/LDSCORE/independent_sumstats/. 1000 Genomes Phase 1 reference panel was obtained from ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/. Hematopoietic ATAC-seq data for Buenrostro et al was obtained from NCBI GEO (accession GSE74912). Hematopoietic ATAC-seq data from Calderon et al was obtained from NCBI GEO (accession GSE118189). ChIP-seq data from Roadmap were obtained from NCBI GEO (accession GSM997225). chromHMM chromatin state partitions for ENCODE were obtained from https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html. PCHiC data were downloaded from Data S1 in the referenced manuscript (https://www.sciencedirect.com/science/article/pii/S0092867416313228). DICE eQTL data were obtained from https://dice-database.org/.

The Verily Life Sciences ATAC-seq from the SysteMS data can be requested by Charlie Kim (charliekim@verily.com).

The LINCS KD and OE TEAD2 data can be accessed through the Broad Connectivity Map portal at clue.io: https://clue.io/command?q=/sig%20%22TEAD2%22.

667 **References**

668 1. Filippi M, Bar-Or A, Piehl F, et al. Multiple sclerosis. *Nat Rev Dis Prim*. 2018;4(1):43.
669 doi:10.1038/s41572-018-0041-4
670 2. Baecher-Allan C, Kaskow BJ, Weiner HL. Multiple Sclerosis: Mechanisms and
671 Immunotherapy. *Neuron*. 2018;97(4):742-768. doi:10.1016/j.neuron.2018.01.021
672 3. van Langelaar J, Rijvers L, Smolders J, van Luijn MM. B and T Cells Driving Multiple
673 Sclerosis: Identity, Mechanisms and Potential Triggers. *Front Immunol*. 2020;11:760.
674 doi:10.3389/fimmu.2020.00760
675 4. Chihara N. Dysregulated T cells in multiple sclerosis. *Clin Exp Neuroimmunol*. 2018;9:20-
676 29. doi:10.1111/cen3.12438
677 5. Negron A, Robinson RR, Stüve O, Forsthuber TG. The role of B cells in multiple
678 sclerosis: Current and future therapies. *Cell Immunol*. 2019;339:10-23.
679 doi:10.1016/j.cellimm.2018.10.006
680 6. Greenfield AL, Hauser SL. B-cell Therapy for Multiple Sclerosis: Entering an era. *Ann*
681 *Neurol*. 2018;83(1):13-26. doi:10.1002/ana.25119
682 7. Hauser SL, Bar-Or A, Comi G, et al. Ocrelizumab versus Interferon Beta-1a in Relapsing
683 Multiple Sclerosis. *N Engl J Med*. 2017;376(3):221-234. doi:10.1056/NEJMoa1601277
684 8. Patsopoulos NA, Baranzini SE, Santaniello A, et al. Multiple sclerosis genomic map
685 implicates peripheral immune cells and microglia in susceptibility. *Science (80-)*. 2019.
686 doi:10.1126/science.aav7188
687 9. Consortium IMSG. Low-Frequency and Rare-Coding Variation Contributes to Multiple
688 Sclerosis Risk. *Cell*. 2018;175(6):1679-1687.e7. doi:10.1016/j.cell.2018.09.049
689 10. Visscher PM, Wray NR, Zhang Q, et al. 10 Years of GWAS Discovery: Biology, Function,
690 and Translation. *Am J Hum Genet*. 2017;101(1):5-22. doi:10.1016/j.ajhg.2017.06.005
691 11. Gusev A, Lee SH, Trynka G, et al. Partitioning heritability of regulatory and cell-type-
692 specific variants across 11 common diseases. *Am J Hum Genet*. 2014;95(5):535-552.

693 doi:10.1016/j.ajhg.2014.10.004

694 12. Cano-Gamez E, Trynka G. From GWAS to Function: Using Functional Genomics to
695 Identify the Mechanisms Underlying Complex Diseases. *Front Genet*. 2020;11:424.
696 doi:10.3389/fgene.2020.00424

697 13. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: Illuminating the Dark
698 Road from Association to Function. *Am J Hum Genet*. 2013;93(5):779-797.
699 doi:10.1016/j.ajhg.2013.10.012

700 14. Gallagher MD, Chen-Plotkin AS. The Post-GWAS Era: From Association to Function. *Am*
701 *J Hum Genet*. 2018;102(5):717-730. doi:10.1016/j.ajhg.2018.04.002

702 15. Farh KK-H, Marson A, Zhu J, et al. Genetic and epigenetic fine mapping of causal
703 autoimmune disease variants. *Nature*. 2015;518(7539):337-343.
704 doi:10.1038/nature13835

705 16. Finucane HK, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by functional
706 annotation using genome-wide association summary statistics. *Nat Genet*.
707 2015;47(11):1228-1235. doi:10.1038/ng.3404

708 17. International Multiple Sclerosis Genetics Consortium. A systems biology approach
709 uncovers cell-specific gene regulatory effects of genetic associations in multiple sclerosis.
710 *Nat Commun*. 2019;10(1):2236. doi:10.1038/s41467-019-09773-y

711 18. Corces MR, Buenrostro JD, Wu B, et al. Lineage-specific and single-cell chromatin
712 accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet*.
713 2016;48(10):1193-1203. doi:10.1038/ng.3646

714 19. Buenrostro JD, Corces MR, Lareau CA, et al. Integrated Single-Cell Analysis Maps the
715 Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell*.
716 2018;173(6):1535-1548.e16. doi:10.1016/j.cell.2018.03.074

717 20. Ulirsch JC, Lareau CA, Bao EL, et al. Interrogation of human hematopoiesis at single-cell
718 and single-variant resolution. *Nat Genet*. 2019;51(4):683-693. doi:10.1038/s41588-019-

0362-6

21. Calderon D, Nguyen MLT, Mezger A, et al. Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat Genet*. 2019. doi:10.1038/s41588-019-0505-9

22. Finucane HK, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47(11):1228-1235. doi:10.1038/ng.3404

23. Bulik-Sullivan BK, Loh P-R, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295. doi:10.1038/ng.3211

24. Dendrou CA, Fugger L, Friese MA. Immunopathology of multiple sclerosis. *Nat Rev Immunol*. 2015;15(9):545-558. doi:10.1038/nri3871

25. Fani Maleki A, Rivest S. Innate Immune Cells: Monocytes, Monocyte-Derived Macrophages and Microglia as Therapeutic Targets for Alzheimer's Disease and Multiple Sclerosis. *Front Cell Neurosci*. 2019;13. doi:10.3389/fncel.2019.00355

26. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*. 2013;45(12):1452-1458. doi:10.1038/ng.2802

27. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511(7510):421-427. doi:10.1038/nature13595

28. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet*. 2011;43(10):977-983. doi:10.1038/ng.943

29. Bradfield JP, Qu H-Q, Wang K, et al. A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. McCarthy MI, ed. *PLoS Genet*.

2011;7(9):e1002293. doi:10.1371/journal.pgen.1002293

30. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012;491(7422):119-124. doi:10.1038/nature11582

31. Bentham J, Morris DL, Graham DSC, et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat Genet*. 2015;47(12):1457-1464. doi:10.1038/ng.3434

32. Okada Y, Wu D, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*. 2014;506(7488):376-381. doi:10.1038/nature12873

33. Cordell HJ, Han Y, Mells GF, et al. International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. *Nat Commun*. 2015;6(1):8019. doi:10.1038/ncomms9019

34. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57-74. doi:10.1038/nature11247

35. Consortium RE, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518(7539):317-330. doi:10.1038/nature14248

36. Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc*. 2017. doi:10.1038/nprot.2017.124

37. Javierre BM, Burren OS, Wilder SP, et al. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell*. 2016;167(5):1369-1384.e19. doi:10.1016/j.cell.2016.09.037

38. Li T, Kim A, Rosenbluh J, et al. GeNets: A unified web platform for network-based genomic analyses. *Nat Methods*. 2018;15(7):543-546. doi:10.1038/s41592-018-0039-6

39. Schmiedel BJ, Singh D, Madrigal A, et al. Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell*. 2018;175(6):1701-1715.e16.

771 doi:10.1016/j.cell.2018.10.022

772 40. Yevshin I, Sharipov R, Kolmykov S, Kondrakhin Y, Kolpakov F. GTRD: A database on
773 gene transcription regulation - 2019 update. *Nucleic Acids Res*. 2019.
774 doi:10.1093/nar/gky1128

775 41. Subramanian A, Narayan R, Corsello SM, et al. A Next Generation Connectivity Map:
776 L1000 Platform and the First 1,000,000 Profiles. *Cell*. 2017.
777 doi:10.1016/j.cell.2017.10.049

778 42. Jelcic I, Al Nimer F, Wang J, et al. Memory B Cells Activate Brain-Homing, Autoreactive
779 CD4+ T Cells in Multiple Sclerosis. *Cell*. 2018;175(1):85-100.e23.
780 doi:10.1016/j.cell.2018.08.011

781 43. Li R, Rezk A, Miyazaki Y, et al. Proinflammatory GM-CSF-producing B cells in multiple
782 sclerosis and B cell depletion therapy. *Sci Transl Med*. 2015;7(310):310ra166.
783 doi:10.1126/scitranslmed.aab4176

784 44. Kebir H, Ifergan I, Alvarez JI, et al. Preferential recruitment of interferon-γ-expressing T_H
785 17 cells in multiple sclerosis. *Ann Neurol*. 2009;66(3):390-402. doi:10.1002/ana.21748

786 45. Kebir H, Kreymborg K, Ifergan I, et al. Human TH17 lymphocytes promote blood-brain
787 barrier disruption and central nervous system inflammation. *Nat Med*. 2007;13(10):1173-
788 1175. doi:10.1038/nm1651

789 46. Benner C, Havulinna AS, Järvelin M-R, Salomaa V, Ripatti S, Pirinen M. Prospects of
790 Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from
791 Genome-wide Association Studies. *Am J Hum Genet*. 2017;101(4):539-551.
792 doi:10.1016/j.ajhg.2017.08.012

793 47. Bai X, Huang L, Niu L, et al. Mst1 positively regulates B-cell receptor signaling via CD19
794 transcriptional levels. *Blood Adv*. 2016;1(3):219-230.
795 doi:10.1182/bloodadvances.2016000588

796 48. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*.

2012;9(4):357-359. doi:10.1038/nmeth.1923

49. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079. doi:10.1093/bioinformatics/btp352

50. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137. doi:10.1186/gb-2008-9-9-r137

51. Amariuta T, Luo Y, Gazal S, et al. IMPACT: Genomic Annotation of Cell-State-Specific Regulatory Elements Inferred from the Epigenome of Bound Transcription Factors. *Am J Hum Genet*. 2019;104(5):879-895. doi:10.1016/j.ajhg.2019.03.012

52. 1000 Genomes Project Consortium, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56-65. doi:10.1038/nature11632

53. Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010;467(7311):52-58. doi:10.1038/nature09298

54. Finucane HK, Reshef YA, Anttila V, et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet*. 2018;50(4):621-629. doi:10.1038/s41588-018-0081-4

55. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4(1):7. doi:10.1186/s13742-015-0047-8

56. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575. doi:10.1086/519795

57. Cairns J, Freire-Pritchett P, Wingett SW, et al. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol*. 2016;17(1):127. doi:10.1186/s13059-016-0992-2

Acknowledgements

NAP was supported in part by National Multiple Sclerosis Society (grants JF-1808-32223 and RG-1707-28657). This work was supported in part by the Water Cove Charitable Foundation. We thank Jacob Ulirsch for technical assistance with the use of LD score regression.

Author contributions.

MHG and NAP had the original idea and supervised the project. JB, CC, XD, PK, CCK, TO, TS, and DZS generated data. HLW and TC provided samples. MHG, PS, BAL, HL, and NAP performed analyses. MHG and NAP wrote a first draft of the paper. All authors reviewed results, contributed to writing and final approval of the paper.

Competing interests.

JB, CC, XD, PK, CCK, TO, TS, and DZS employment in Verily Life Sciences at time of study.

Correspondence.

Correspondence should be addressed to Nikolaos Patsopoulos (npatsopoulos@rics.bwh.harvard.edu) and Michael Guo (michael.guo@pennmedicine.upenn.edu)

Computer code

Code used in this paper can be accessed in bitbucket: https://bitbucket.org/patslab/pis_ms_enrichment/

Figures

Figure 1: A: Experimental Design. Top box shows the hematopoietic cell types analyzed. MS discovery GWAS results were integrated with ATAC-seq profiles generated from the hematopoietic cell types. LDSC was performed to evaluate enrichment of MS GWAS in the OCRs of each hematopoietic cell type. Statistical fine-mapping was also performed on the MS GWAS results, which were then integrated with orthogonal epigenetic data such as promotor capture HiC interactions. This integration of fine-mapping and epigenetic data allowed for identification of putative causal mechanisms at individual loci. **B:** Enrichment of MS GWAS heritability in hematopoietic cell OCRs. Enrichment p-values are shown as $-\log_{10}$(p-value).

Figure 1

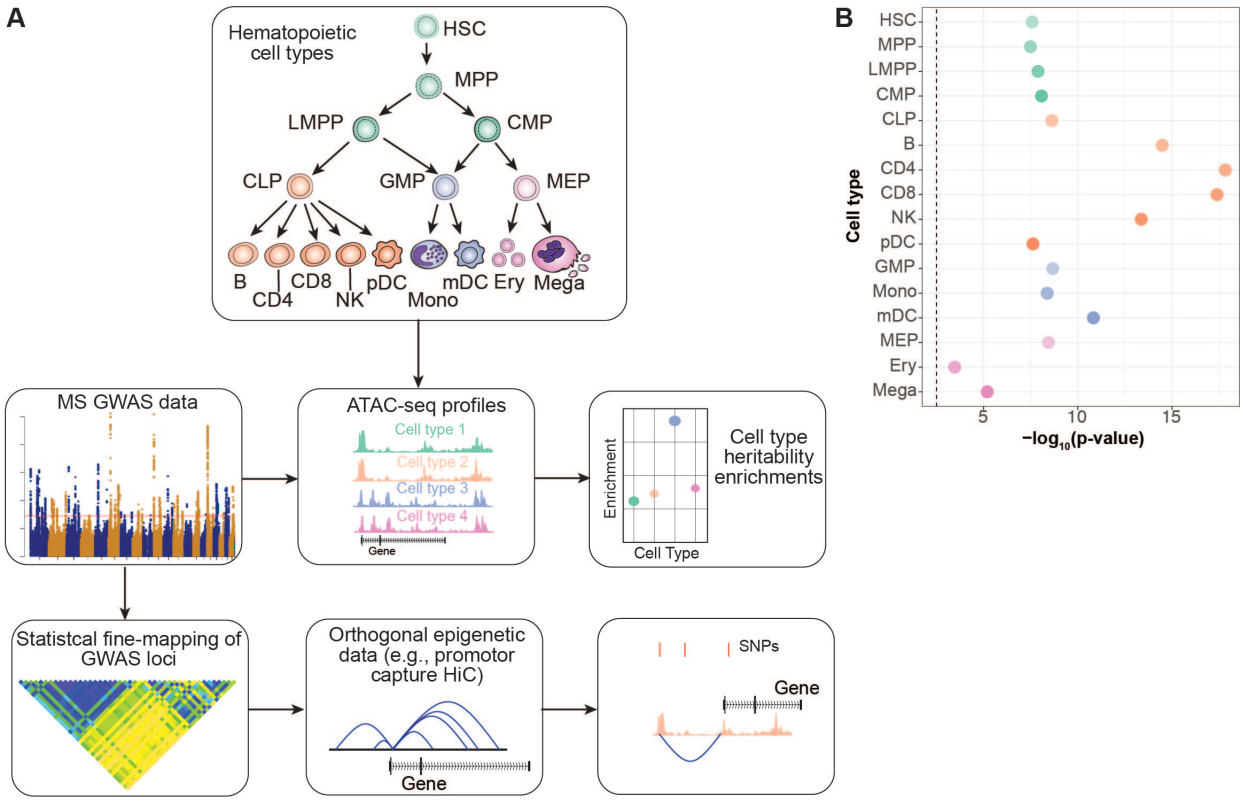


Figure 2: A: LDSC enrichment results for MS GWAS enrichment in OCRs from across hematopoietic cell types in a joint model. Heights of the circles reflect LDSC coefficient (τ_c) p-values, which measures whether the annotation (i.e., OCRs for a given cell type) contributes significantly to SNP heritability in an overall model that includes OCRs for all hematopoietic cell

types and baseline annotations. Sizes of the circles are proportional to the enrichment p-values for that given cell type, with larger circles reflecting more significant p-values. **B:** LDSC enrichment p-values for pairwise stratified LDSC of MS GWAS results in OCRs from hematopoietic cell types. Y-axis are the index cell types with LDSC enrichment p-values prior to stratifying in parentheses. X-axis shows the comparator cell type being conditioned upon. Boxes are shaded by the LDSC coefficient p-values for the index cell type after conditioning on the comparator cell type in the pairwise model (with darker colors representing stronger enrichments). Red stars indicate pair-wise comparisons that are statistical significant a Bonferroni corrected p-value threshold of 2.2×10^{-4}.

Figure 2

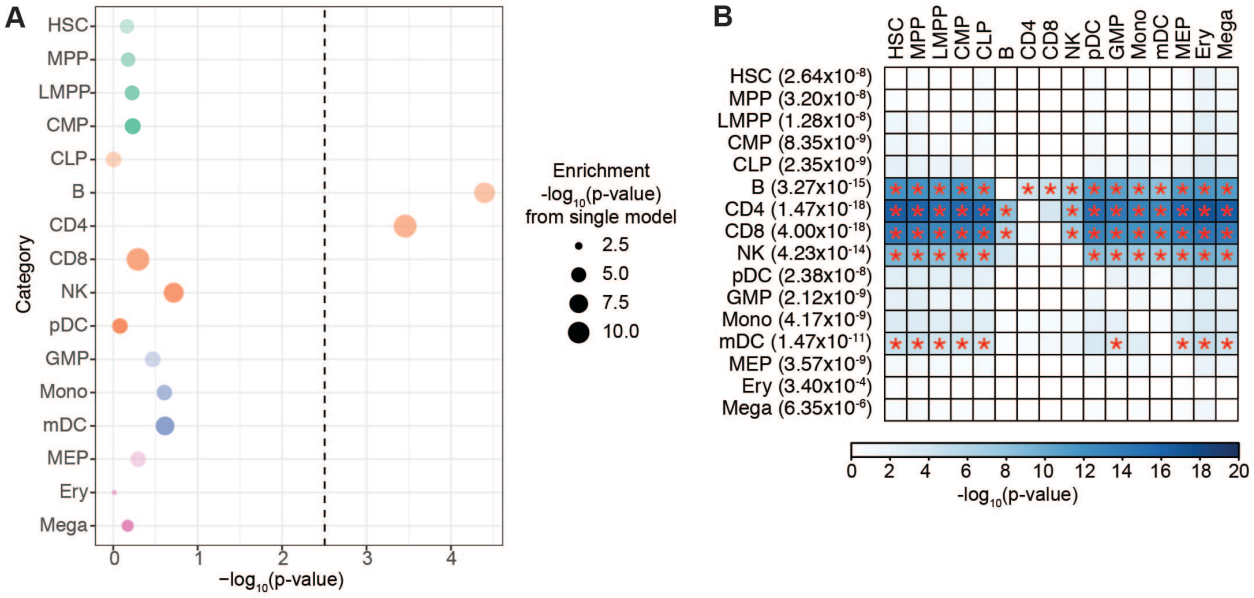
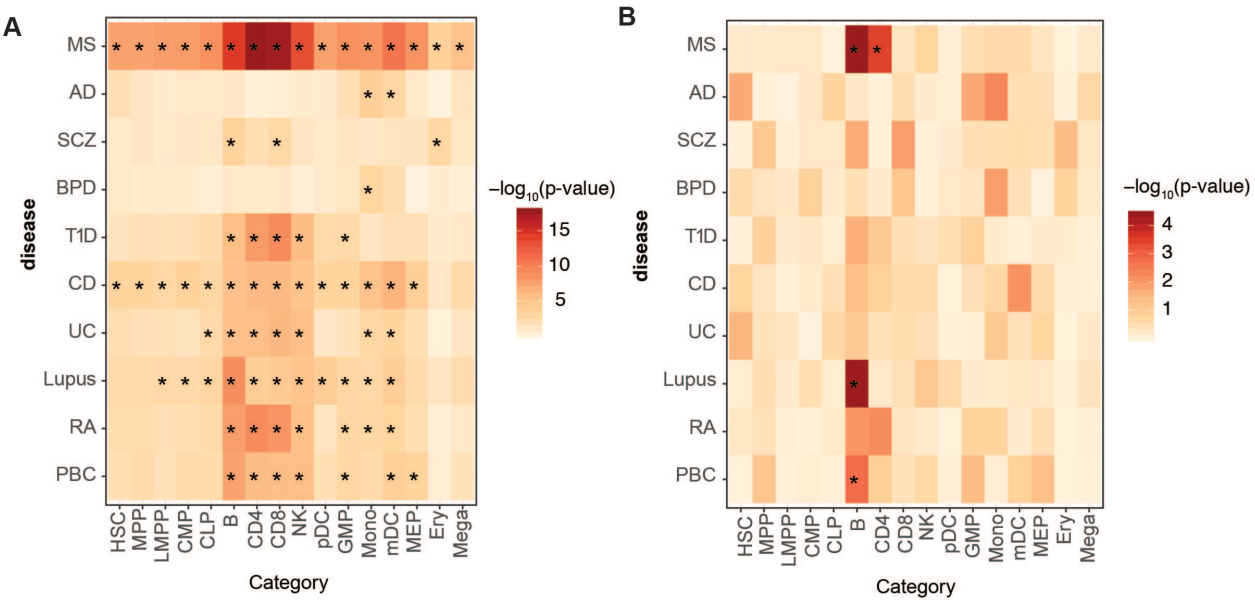


Figure 3: A: Enrichments of GWAS results from 10 neuropsychiatric or autoimmune conditions in OCRs across various hematopoietic cell types. **B:** LDSC coefficient p-values in the joint model across hematopoietic cell ATAC-seq in 10 neuropsychiatric or autoimmune conditions. For **A** and **B**, boxes are shaded by $-\log_{10}$(p-value), with darker shading reflecting more statistical significance, and statistically significant p-values (p-values<3.13×10^{-3}) are starred.

Figure 3



875

876

877 **Figure 4: A:** Schematic of lineage relationships among CD4$^+$ T cell subsets for which ATAC-seq

878 data was analyzed. **B:** LDSC heritability enrichment p-values for CD4$^+$ T cell subsets in MS

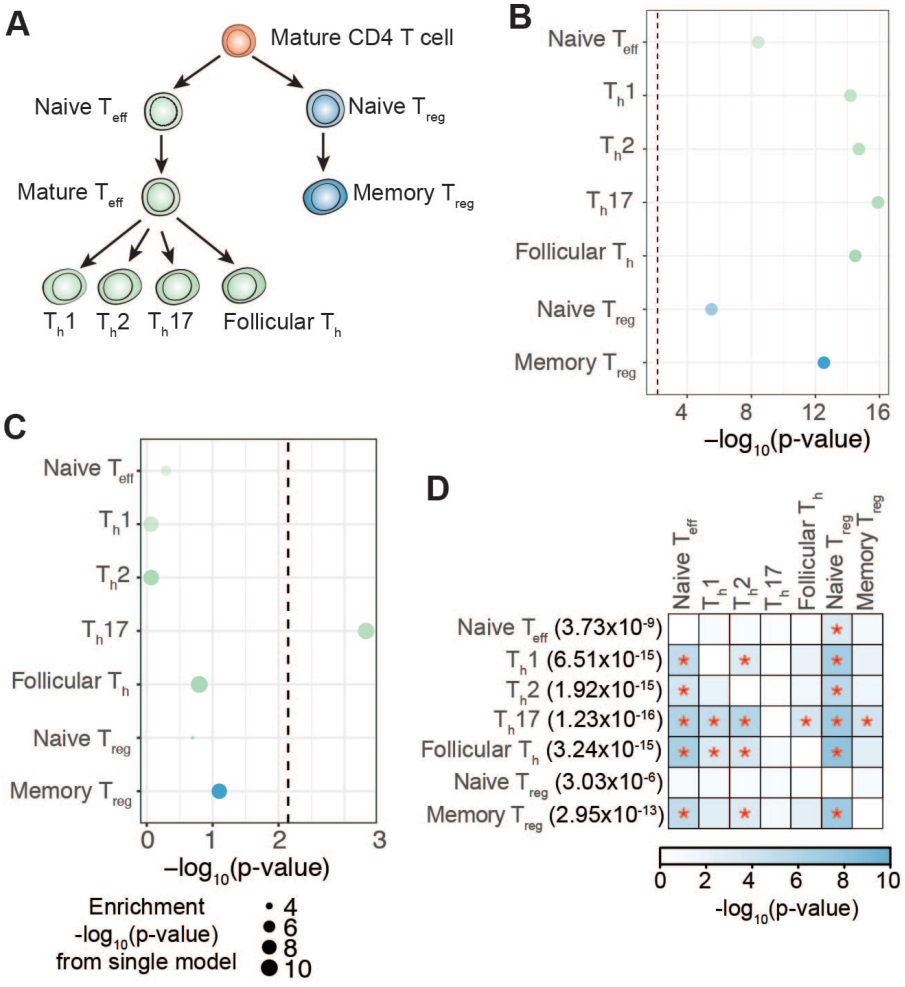
879 GWAS. See Figure 1B for additional description. **C:** LDSC coefficient p-values for CD4$^+$ T cells

880 in MS GWAS. See Figure 2A for additional description. **D:** LDSC coefficient p-values for

881 pairwise stratified analyses of MS GWAS results in ATAC-seq data from CD4$^+$ T cell subsets.

882 See Figure 2B legend for additional desc

Figure 4



883

884 ription.

885

886

887 **Figure 5: A:** Schematic of lineage relationships among B cell lineage cell types for which ATAC-

888 seq data was analyzed. **B:** LDSC heritability enrichment p-values for B cell lineage cell types in

889 MS GWAS. See Figure 1B for additional description. **C:** Stratified LDSC coefficient p-values for

890 B cell lineage cell types in MS GWAS. See Figure 2A for additional description. **D:** LDSC

891 coefficient p-values for pairwise stratified analyses of MS GWAS results in ATAC-seq data from

892 B cell lineage cell types. See Figure 2B legend for additional description.

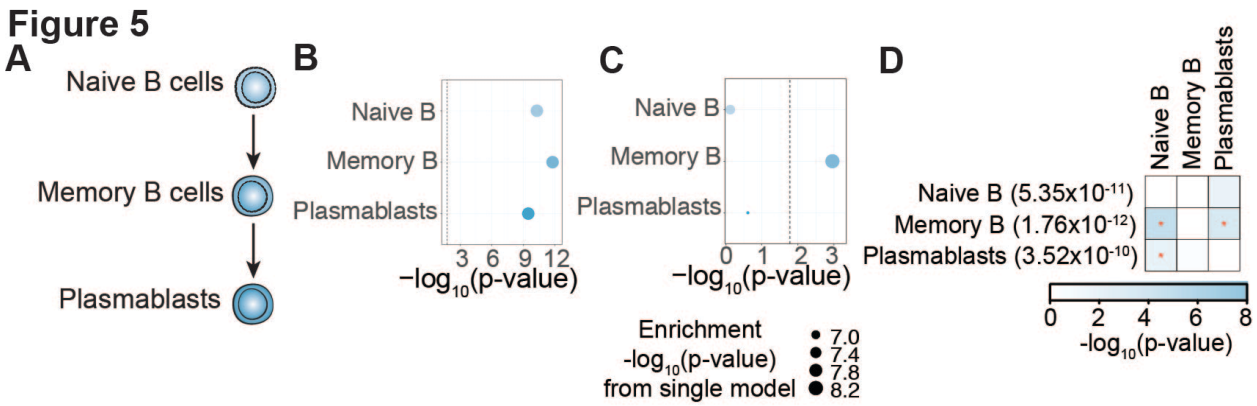


Figure 6: A: Enrichment of MS GWAS heritability in OCRs from untreated patients with MS. Enrichment p-values are shown as $-\log_{10}$(p-value). **B, C:** LDSC results for MS GWAS enrichment in a joint model for T4cm (**B**) and cMBc (**C**) OCRs from untreated patients with MS. Heights of the circles reflect stratified LDSC coefficient p-values. Sizes of the circles are proportional to the enrichment p-values for that given cell type, with larger circles reflecting more significant p-values. **D:** Enrichment of MS GWAS heritability in OCRs from MS patients undergoing immunomodulatory treatment. Enrichment p-values are shown as $-\log_{10}$(p-value). Treatments include glatiramer acetate, interferon, or natalizumab.

Figure 6

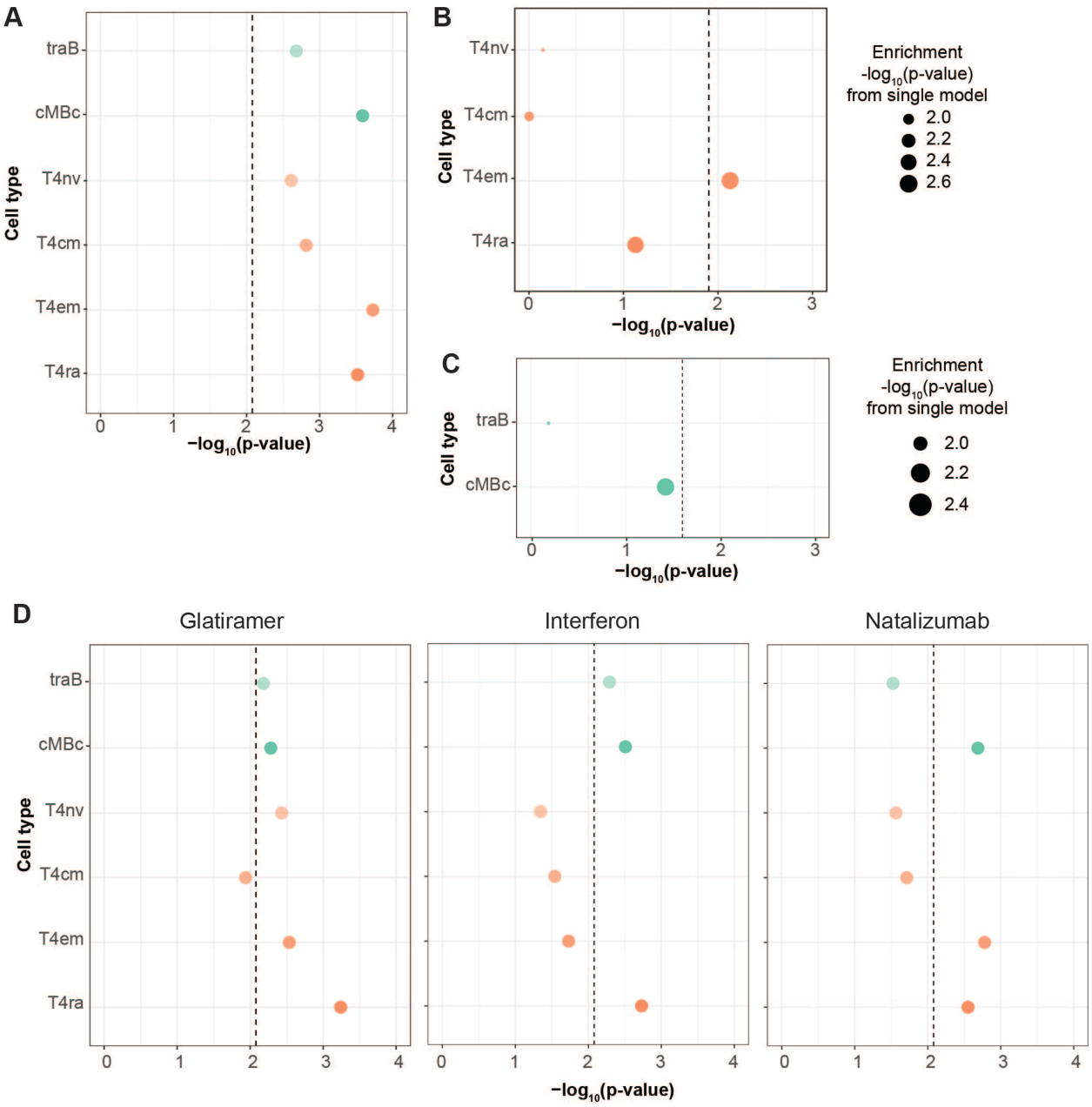
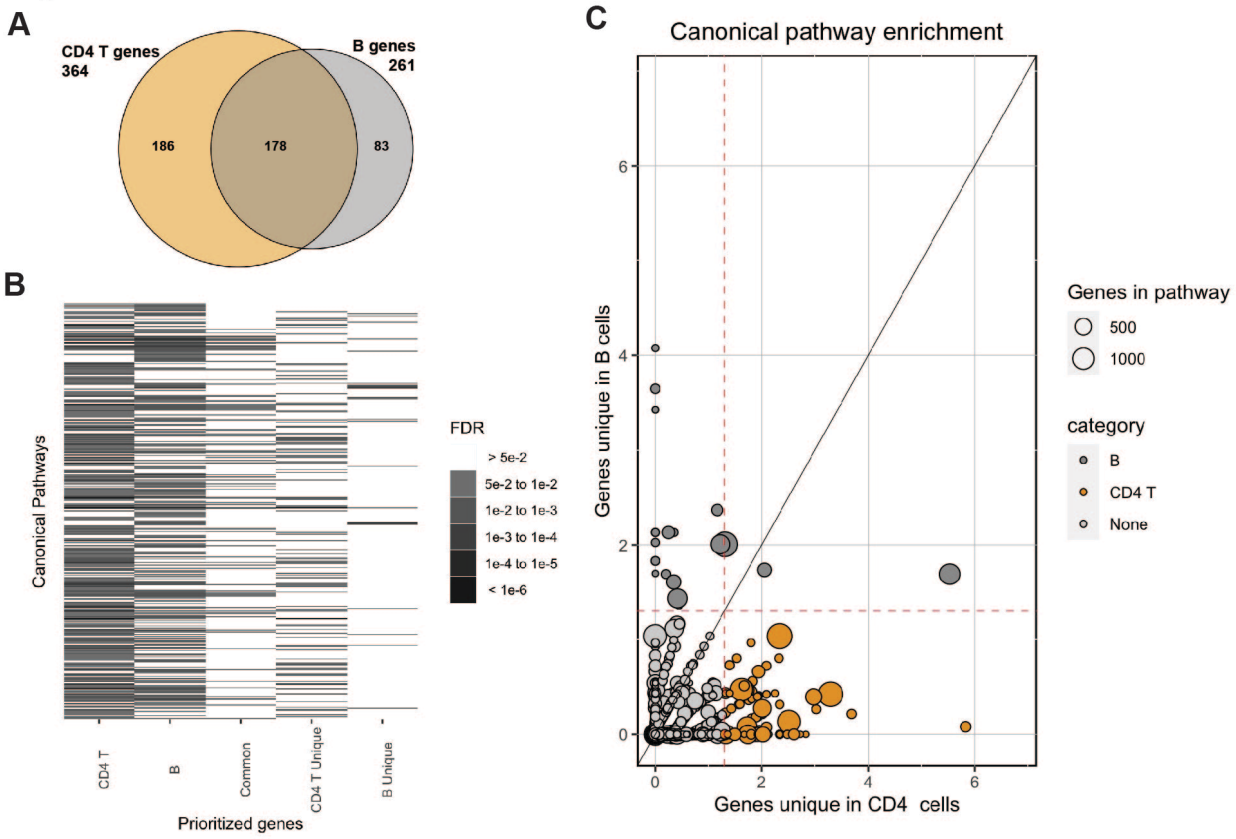


Figure 7: A. Venn diagram of the putative causal CD4 T and B cell genes. **B.** Heatmap of canonical pathway enrichment for the putative causal genes in CD4 T cells, B cells, common in CD4 T and B cells, unique in CD4 T cells, and unique in B cells. Only pathways with FDR<5% in at least one gene list are displayed (n=1950). The grayscale depicts level of statistical significance. **C.** Scatterplot of -log10(FDR) of canonical pathway enrichment for putative causal

909 genes unique in CD4 T cells (X axis) vs. B cells (Y axis). The dashed red lines indicated

910 FDR<5%. The size of the dots depicts the total number of genes in the respective pathway.

Figure 7



911

912

913 **Figure 8: A:** Visualization of *TEAD2* locus. Lead SNP rs1465697 (PICS of 15%) is depicted

914 with a red line. The blue box on the left illustrates the overlap with the ATAC-seq peaks present

915 in CD4 T (orange) and B cells (purple). The SNP and ATAC-seq peaks also overlap a PCHiC

916 looping interaction with the promoter for the *TEAD2* gene (arc; the boundaries of the

917 enhancer/promoter regions are indicated in green; the promoter of *TEAD2* is highlighted with

918 the blue box on the right). **B**. Gene expression of *TEAD2* across immune cells available in the

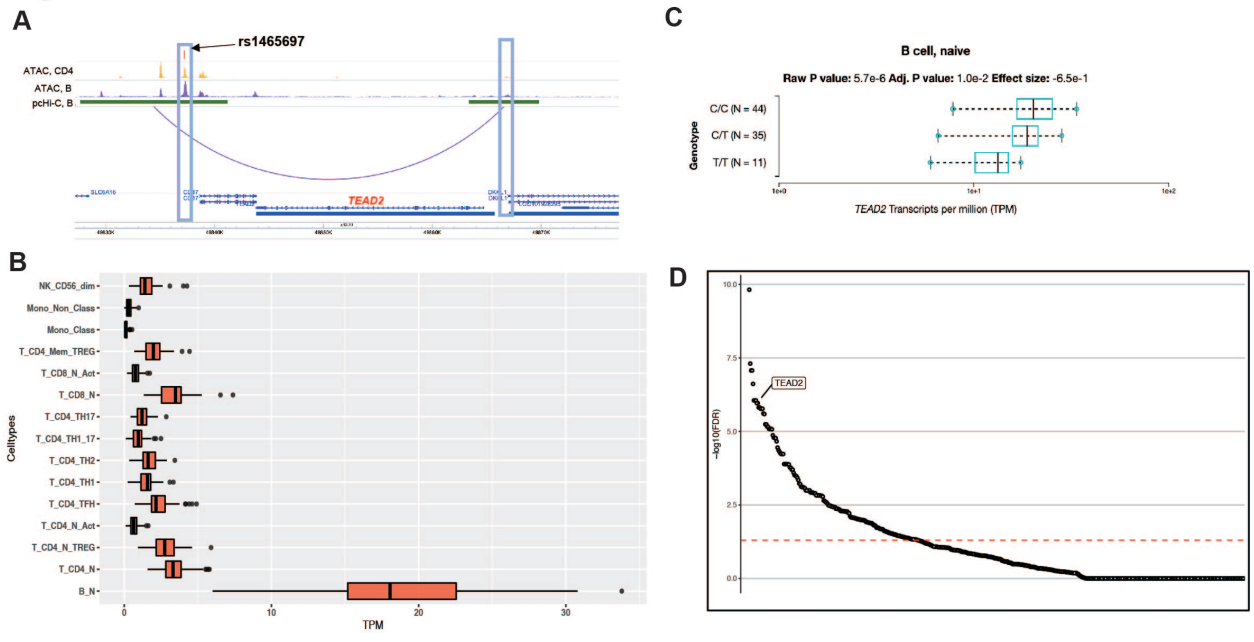
919 DICE database (https://dice-database.org/). X axis display transcripts per million (TPM). **C.** Cis-

920 eQTL boxplot per genotype status of rs1465697 in naïve B cells in the DICE database

921 (https://dice-database.org/). **D.** Transcription factor enrichment in the GTRD database for the

putative causal genes that are common in CD4 T and B cells. Each dot represents one of 526 transcription factors. The Y axis indicates the -log10 of the FDR. The *TEAD2* enrichment is highlighted (p-value = 1.34×10^{-8}, FDR = 8.81×10^{-7}).

Figure 8



Supplemental Figure Legends

Figure S1: Correlation (Pearson's r^2) in ATAC-seq profiles across hematopoietic cell types.

Figure S1

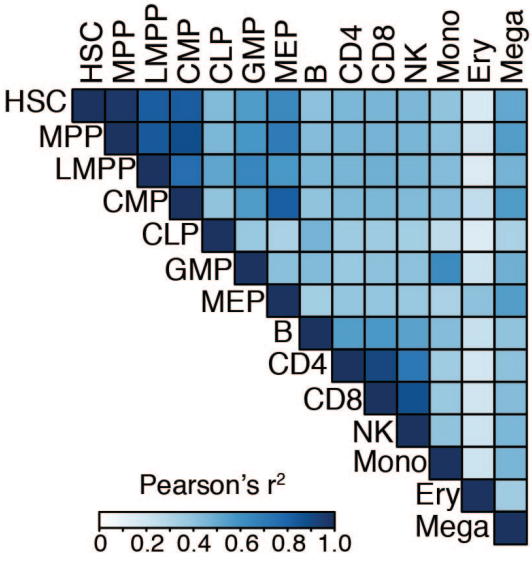


Figure S2: LDSC enrichments for MS GWAS in cell-type specific ATAC-seq peaks in each of the nine mature hematopoietic cell type. Y-axis shows $-\log_{10}$(p-value) of the LDSC heritability enrichment.

Figure S2

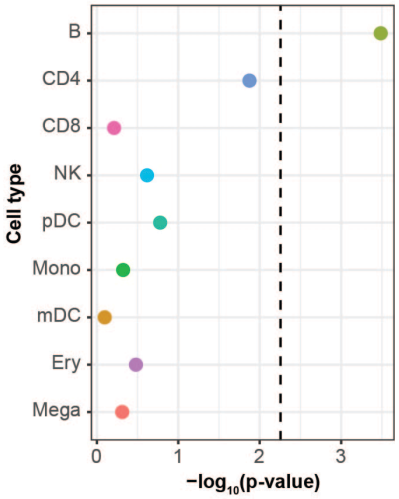


Figure S3: A. LDSC enrichment results for MS GWAS enrichment in T4cm OCRs from treated and MS treated patients in a joint model. **B.** Stratified LDSC enrichment results for MS GWAS enrichment in cMBc OCRs from treated and MS treated patients in a joint model. Heights of the circles reflect stratified LDSC coefficient p-values. Sizes of the circles are proportional to the enrichment p-values for that given cell type, with larger circles reflecting more significant p-values.

Figure S3

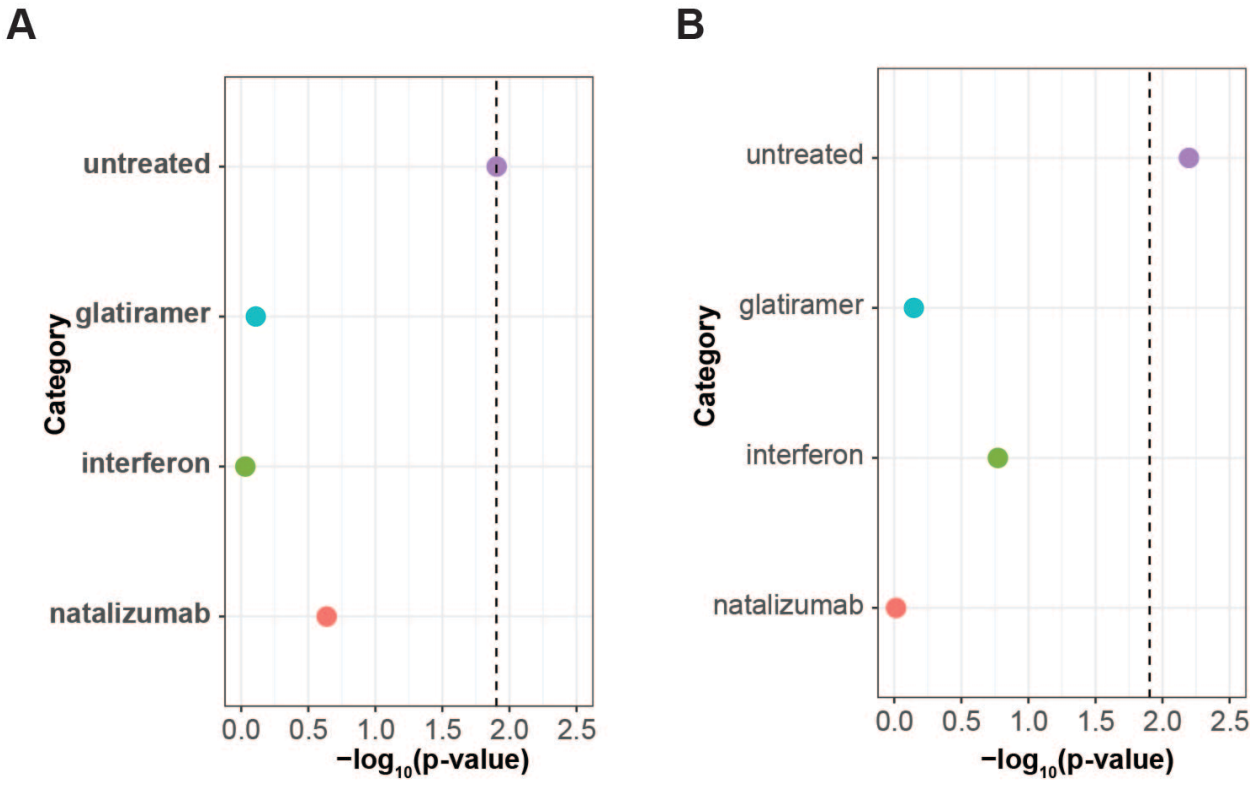


Figure S4: A: LDSC enrichment p-values for MS GWAS data in $CD4^+$ T cell ChIP-seq peaks of various histone markers. Y-axis shown as $-\log_{10}$(p-value). **B:** Same as Figure S3A, except performed for B cell ChIP-seq histone markers.

Figure S4

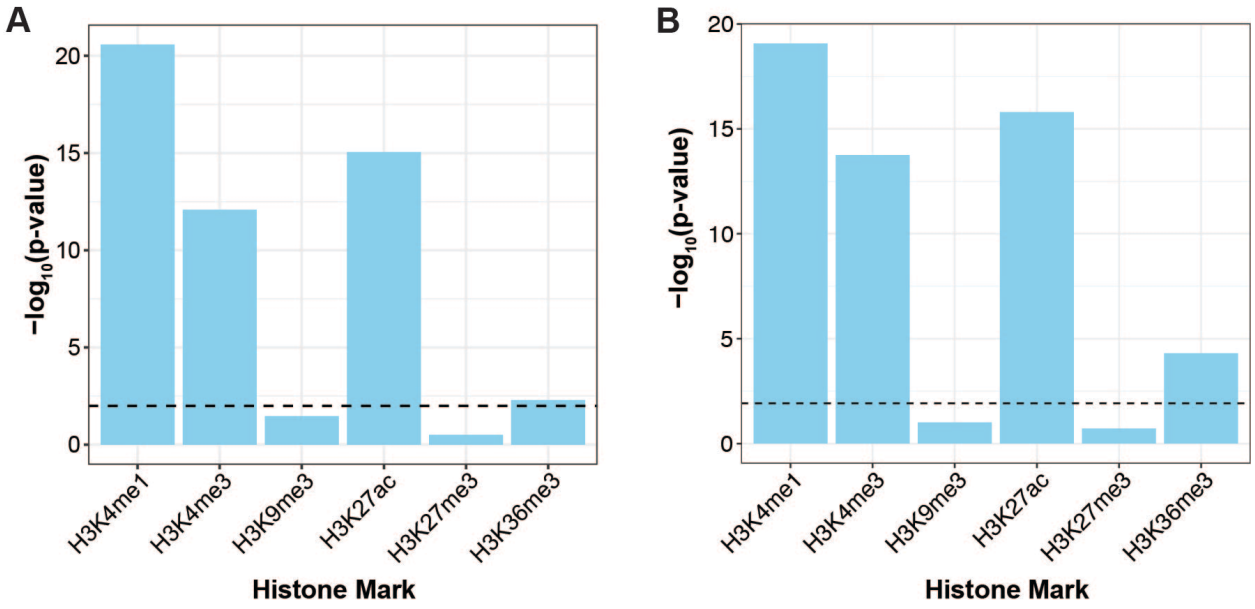


Figure S5: A: LDSC enrichment p-values for chromHMM chromatin states in CD4+ T cells from ENCODE. **B:** Same as Figure S5A, except for B cells from ENCODE. TssA: Active TSS; PromU: Promoter Upstream TSS; PromD1: Promoter Downstream TSS 1; PromD2: Promoter Downstream TSS 2; Tx5: Transcribed - 5' preferential; Tx: Strong transcription; Tx3: Transcribed - 3' preferential; TxWk: Weak transcription; TxReg: Transcribed & regulatory (Prom/Enh); TxEnh5: Transcribed 5' preferential and Enh; TxEnh3: Transcribed 3' preferential and Enh; TxEnhW: Transcribed and Weak Enhancer; EnhA1: Active Enhancer 1; EnhA2: Active Enhancer 2; EnhAF: Active Enhancer Flank; EnhW1: Weak Enhancer 1; EnhW2: Weak Enhancer 2; EnhAc: Primary H3K27ac possible Enhancer; DNase: Primary DNase; ZNF: ZNF genes & repeats; Het: Heterochromatin; PromP: Poised Promoter; PromBiv: Bivalent Promoter; ReprPC: Repressed Polycomb.

Figure S5

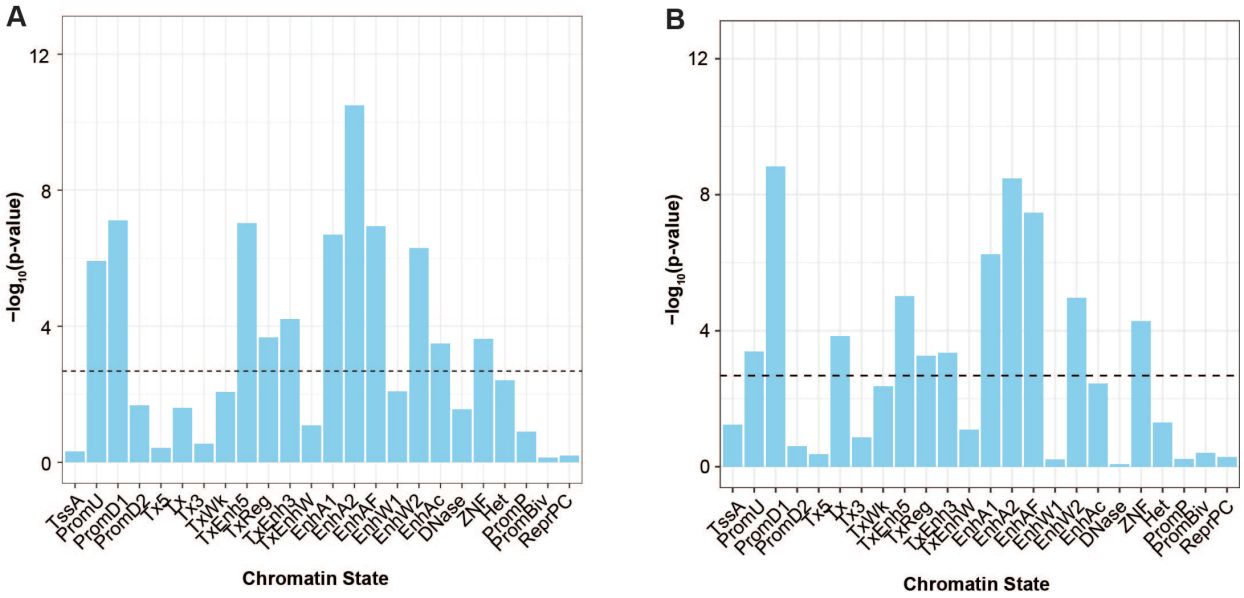


Figure S6: A: Histogram of the number of credible set (CS) variants across loci. **B:** Histogram of the PICS probability of the top variant in each locus.

Figure S6

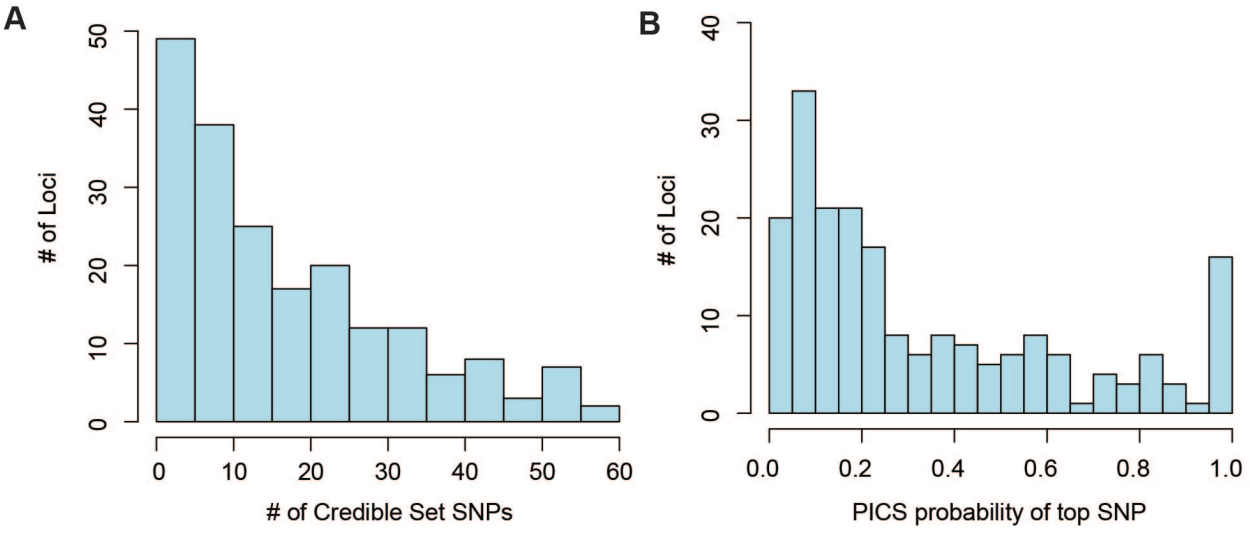


Figure S7: Number of MS GWAS loci (out of 200) with at least one CS SNP overlapping an ATAC-seq peak in the listed cell types.

Figure S7

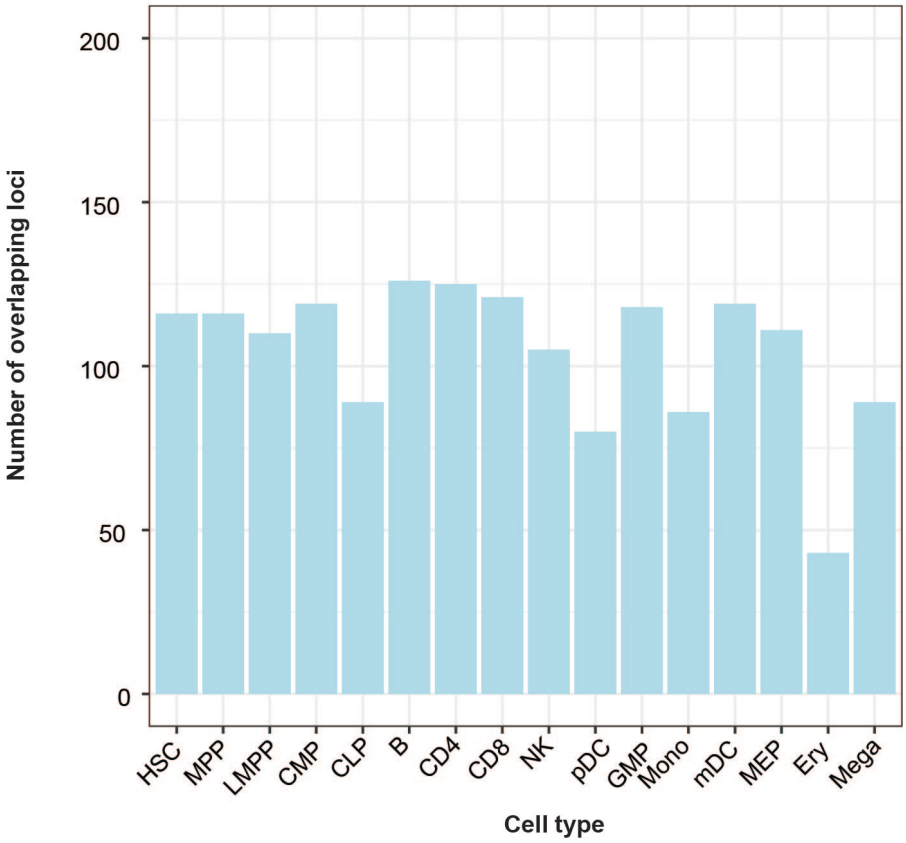


Figure S8: Protein-protein interaction communities of putative causal genes in CD4 T cells.

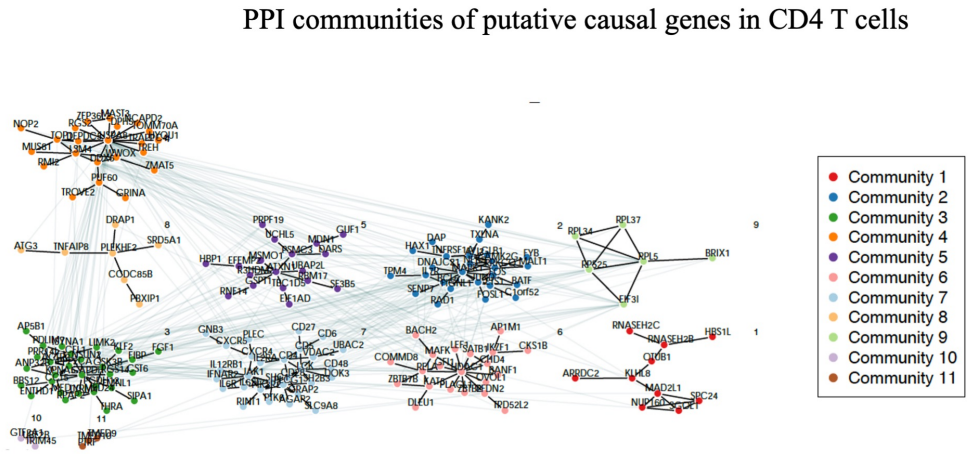


Figure S9: Protein-protein interaction communities of putative causal genes in B cells.

PPI communities of putative causal genes in B cells

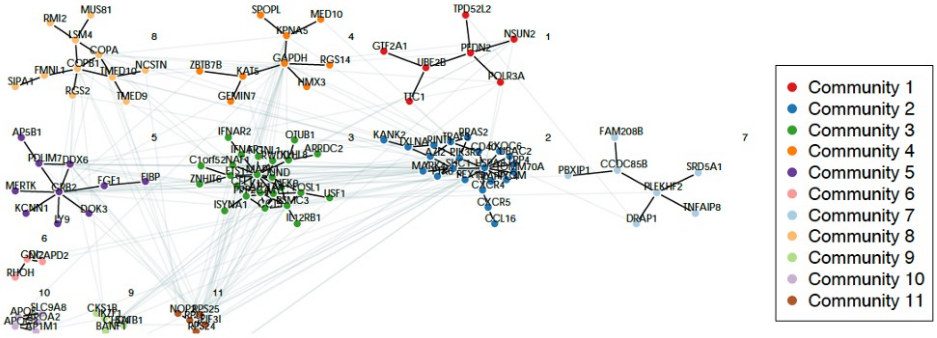


Figure S10: Protein-protein interaction communities of putative causal genes shared in CD4 T and B cells.

PPI communities of putative causal genes common in CD4 T and B cells

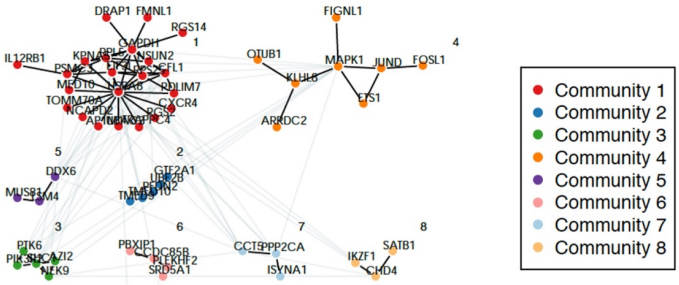


Figure S11: Protein-protein interaction communities of putative causal genes unique in CD4 T cells.

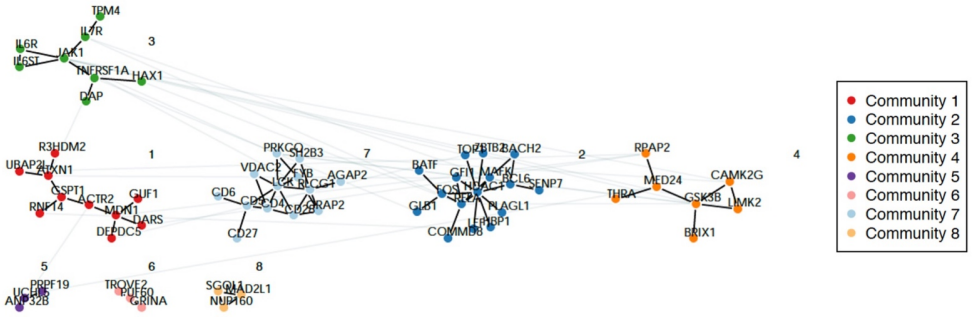


Figure S12: Protein-protein interaction communities of putative causal genes unique in B cells.

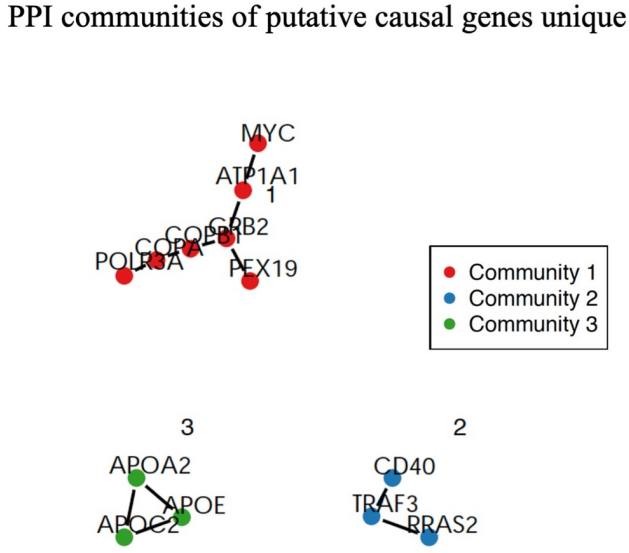


Figure S13: Enrichment of CD4 T cell putative causal genes in GTRD database. Each dot represents one transcription factor. The Y axis displays -log10 of false discovery rate (FDR). The dashed red line indicates the threshold of 1% FDR. The enrichment for TEAD2 predicted target genes is labeled.

Figure S13

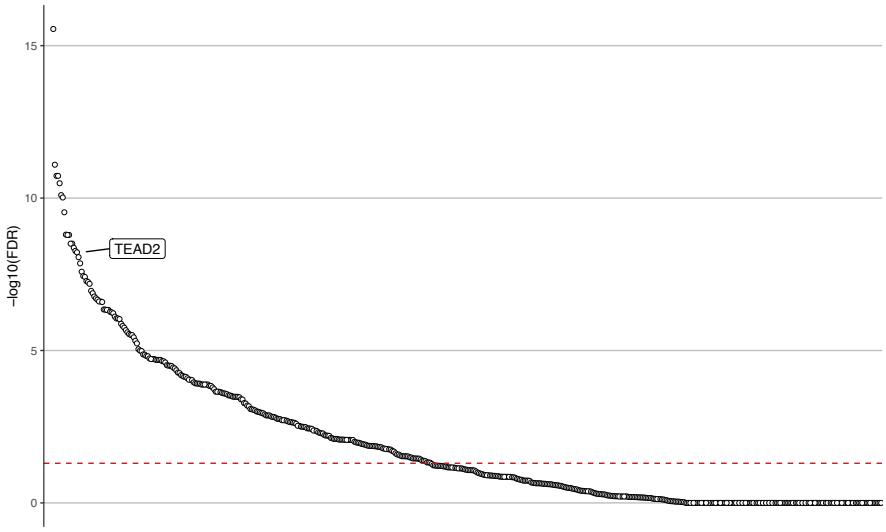
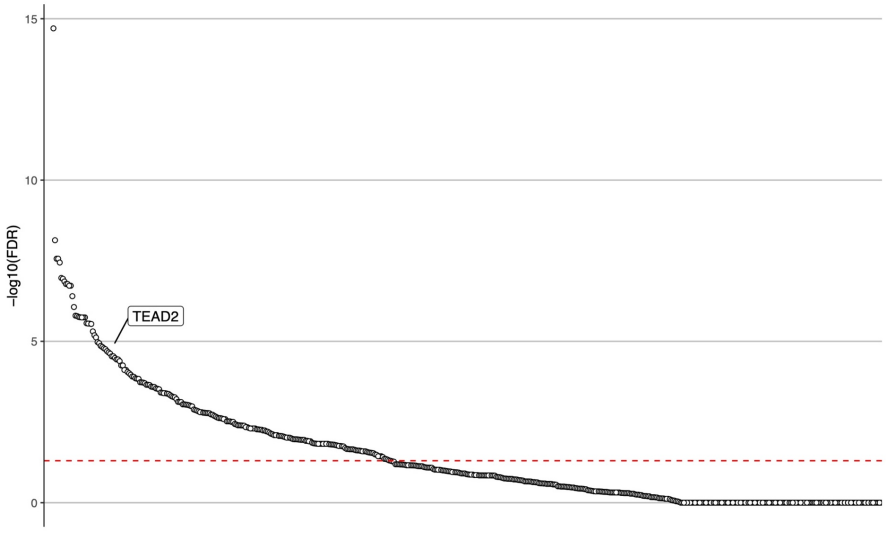


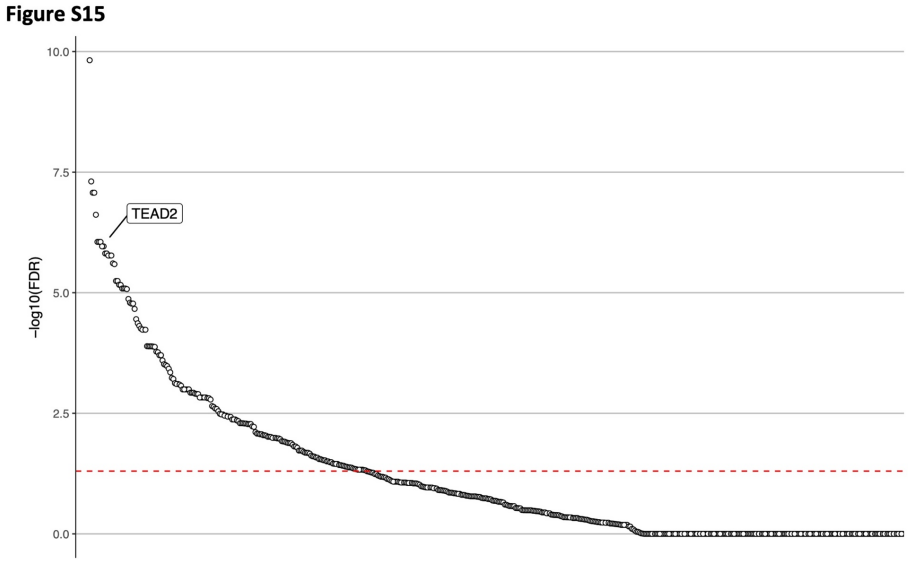
Figure S14: Enrichment of B cell putative causal genes in GTRD database. Each dot represents one transcription factor. The Y axis displays -log10 of false discovery rate (FDR). The dashed red line indicates the threshold of 1% FDR. The enrichment for TEAD2 predicted target genes is labeled.

Figure S14



994

995 **Figure S15:** Enrichment of shared between CD4 T and B cells putative causal genes in GTRD
996 database. Each dot represents one transcription factor. The Y axis displays -log10 of false
997 discovery rate (FDR). The dashed red line indicates the threshold of 1% FDR. The enrichment
998 for TEAD2 predicted target genes is labeled.



999

1000

1001 **Figure S16:** Change of gene expression of CD4 T cell putative causal genes in knock-down
1002 (KD) and over-expression (OE) models in cancer cell lines. Eight cancer cell lines are displayed:
1003 A375, A549, HA1E, HEPG2, HT29, MCF7, PC3, and VCAP. Putative causal genes are
1004 represented with lines connecting the ranked KD gene expression data (left column) with the
1005 ranked OE gene expression data (right column). Genes that are in the extreme 10% in opposite
1006 directions are indicated with green solid lines or red solid lines if these are also a predicted gene
1007 target for TEAD2. The light grey lines display all over putative causal genes.

Figure S16

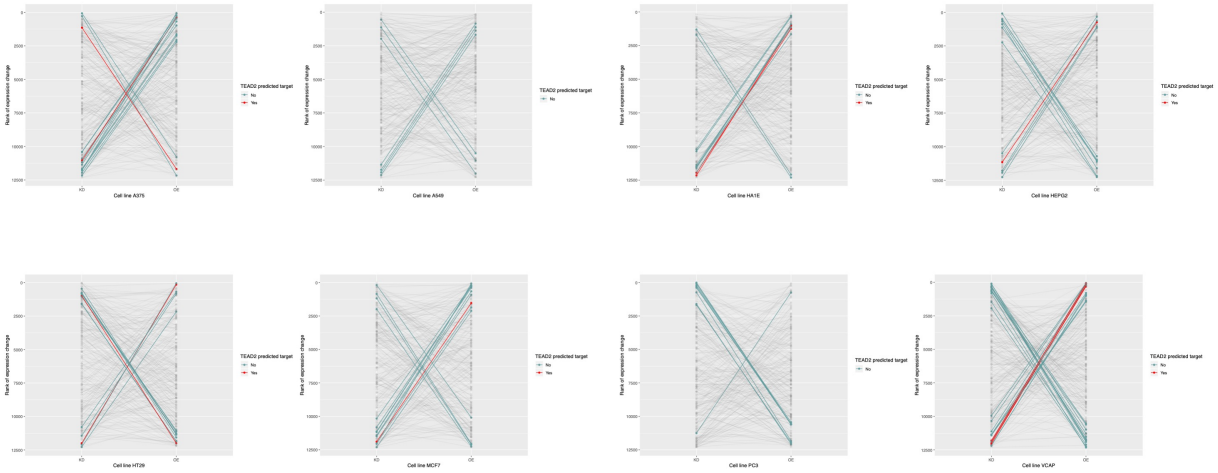
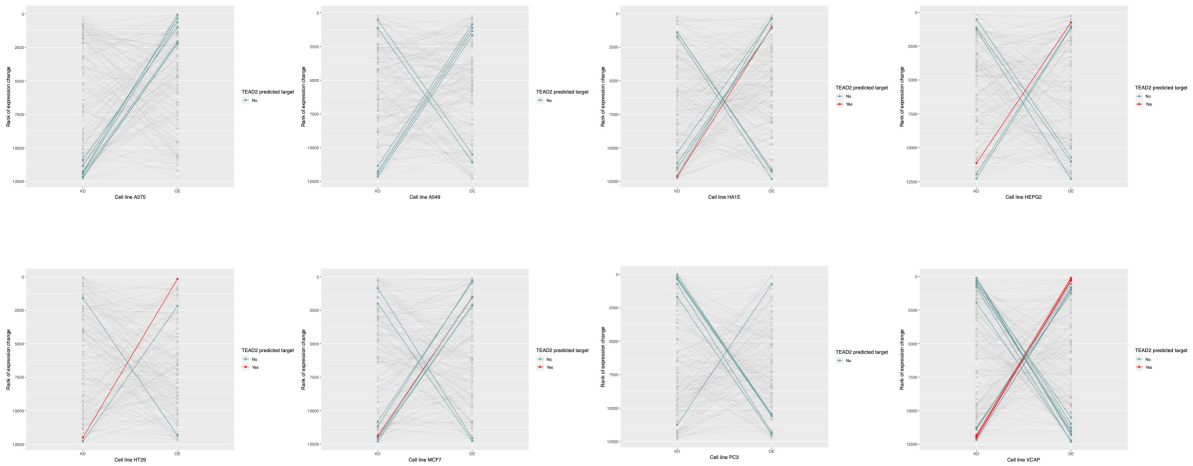


Figure S17: Change of gene expression of B cell putative causal genes in knock-down (KD) and over-expression (OE) models in cancer cell lines. Eight cancer cell lines are displayed: A375, A549, HA1E, HEPG2, HT29, MCF7, PC3, and VCAP. Putative causal genes are represented with lines connecting the ranked KD gene expression data (left column) with the ranked OE gene expression data (right column). Genes that are in the extreme 10% in opposite directions are indicated with green solid lines or red solid lines if these are also a predicted gene target for TEAD2. The light grey lines display all over putative causal genes.

Figure S17



Supplementary Tables

Supplementary Table 1: MS GWAS enrichment in 16 hematopoietic cell types. Results from LDSC from a model including annotation of interest and set of baseline annotations. Column 2 shows the Proportion of SNPs in OCRs of that cell type. Column 3 shows the SNP heritability (h^2_g) for the annotation. Column 4 shows the standard error for h^2g. Column 5 shows enrichment of SNP heritability, defined as proportion of SNP heritability in the annotation divided by the proportion of SNPs in that annotation. Column 6 shows the standard error of the enrichment of SNP heritability. Column 7 shows the p-value of the enrichment of SNP heritability.

Supplementary Table 2: MS GWAS enrichment of hematopoietic cell types in joint model: Results from LDSC from a joint model including annotations from all cell types and set of

baseline annotations. Column 1 shows the coefficient τ_c which measures the contribution for a given annotation to heritability in this overall model, stratified on other annotations in the model. Column 2 shows the standard error for the coefficient τ_c. Column 3 shows the z-score for coefficient τ_c and column 4 shows the p-value for coefficient τ_c.

Supplementary Table 3. MS GWAS pairwise enrichment in 16 hematopoietic cell types. Index cell types are shown on the left. The comparator cell types are shown on top. Coefficient p-values are shown for the index cell type in a model that includes the index cell type, comparator cell type, and the set of baseline annotations.

Supplementary Table 4: MS GWAS cell-specific enrichment within terminal hematopoietic cell types. Columns are the same as in Supplementary Table 1.

Supplementary Table 5: Heritability enrichments (single model) for other disorders. Heritability enrichment p-values for each cell types across 10 autoimmune or neuropsychiatric disorders.

Supplementary Table 6: Heritability enrichments (joint model) for other disorders. Heritability enrichments (p-values for coefficient τ_c) for each cell type under joint model across 10 autoimmune or neuropsychiatric disorders.

Supplementary Table 7. MS GWAS enrichment in CD4+ T cell subpopulations. Columns are the same as in Supplementary Table 1.

Supplementary Table 8. Stratified LDSC in CD4+ T cell subpopulations. Columns are the same as in Supplementary Table 2.

Supplementary Table 9. MS GWAS pairwise enrichment in CD4+ T subpopulations. Columns are the same as in Supplementary Table 3.

Supplementary Table 10. MS GWAS enrichment in B cell subpopulations. Columns are the same as in Supplementary Table 1.

Supplementary Table 11. Stratified LDSC in B cell subpopulations. Columns are the same as in Supplementary Table 2.

Supplementary Table 12. MS GWAS pairwise enrichment in B cell subpopulations. Columns are the same as in Supplementary Table 3.

Supplementary Table 13. Clinical characteristics of MS subjects. EDSS: Expanded Disability Status Scale. * for the Untreated subjects it indicates number of months since last treatment.

Supplementary Table 14. MS GWAS enrichment across OCRs in 6 cell types isolated from untreated MS patients. Columns are the same as in Supplementary Table 1.

Supplementary Table 15. MS GWAS enrichment across OCRs in joint model in CD4 T cell types isolated from untreated MS patients. Columns are the same as in Supplementary Table 2.

Supplementary Table 16. MS GWAS enrichment across OCRs in joint model in B cell types isolated from untreated MS patients. Columns are the same as in Supplementary Table 2.

Supplementary Table 17 MS GWAS enrichment across OCRs in 6 cell types isolated from treated MS patients. Columns are the same as in Supplementary Table 1, except column added for the immune-modulating treatment

Supplementary Table 18. MS GWAS enrichment across OCRs in joint model in T4em isolated from untreated and treated MS patients. Columns are the same as in Supplementary Table 2

Supplementary Table 19. MS GWAS enrichment across OCRs in joint model in cMBc isolated from untreated and treated MS patients. Columns are the same as in Supplementary Table 2

Supplementary Table 20. MS GWAS enrichments in histone ChIP-seq from T_h17 cells. Columns are the same as in Supplementary Table 1.

Supplementary Table 21. MS GWAS enrichments in histone ChIP-seq from B cells. Columns are the same as in Supplementary Table 1.

Supplementary Table 20. MS GWAS enrichments in chromatin states from T_h17 cells. Columns are the same as in Supplementary Table 1.

Supplementary Table 21. MS GWAS enrichments in chromatin states from B cells. Columns are the same as in Supplementary Table 1.

Supplementary Table 24: Summary of gene prioritizations. For each GWAS locus, the lead SNP ("Effect SNP"), region annotation according to Patsopoulos et al., lead SNP chromosome, lead SNP position (in hg19), A1 allele, A2 allele, odd's ratio (OR) and lead SNP p-value are shown. Previous gene prioritizations from Patsopolous et al., based on various criteria are listed. For columns "B Cell ATAC (Buenrostro)", "B Cell ATAC (Calderon)", "Memory B Cell ATAC (Calderon)", "CD4 Cell ATAC (Buenrostro)", "CD4 Cell ATAC (Calderon)", "Th17 Cell ATAC (Calderon)", a "1" indicates that a credible set SNP in the locus directly intersects an OCR from the indicated ATAC-seq dataset ("0" if no credible set SNP in the locus directly intersects. "PCHiC CD4 genes" and "PCHiC B genes" columns list genes that have a promoter capture HiC looping interaction to a credible set SNP in the indicated cell type. "PCHiC + ATAC CD4 genes" indicates a credible set SNP intersect an OCR in CD4 T cells and forms a PCHiC looping interaction in CD4 T cells to that gene. "PCHiC + ATAC B genes" indicates a credible set SNP intersect an OCR in B cells and forms a PCHiC looping interaction in B cells to that gene. The number of credible set SNPs in each locus is shown in the last column.

Supplementary Table 25: Canonical pathway enrichment for prioritized MS genes.

Supplementary Table 26: Protein-protein interaction connectivity summaries for prioritized gene lists. Detailed outputs from GeNets are provide.

Supplementary Table 27: CD4 T prioritized genes ranked in the opposite extreme 10% of gene expression changes in KD and OE cell lines.

Supplementary Table 28: B prioritized genes ranked in the opposite extreme 10% of gene expression changes in KD and OE cell lines.

Supplementary Table 29: Markers for FACS sorting strategy of Verily ATAC-seq data