

RESEARCH

Open Access

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



Michael H. Guo^{1,2*}, Prashanth Sama^{2,3,4,5}, Brenna A. LaBarre^{2,3,4,5}, Hrishikesh Lokhande^{6,7}, John Balibalos⁸, Ci Chu⁸, Xiaomi Du⁸, Pouya Kheradpour⁸, Charles C. Kim⁸, Taylor Oniskey⁸, Thomas Snyder⁸, Damien Z. Soghoian⁸, Howard L. Weiner^{6,7}, Tanuja Chitnis^{6,7} and Nikolaos A. Patsopoulos^{2,3,4,5*}

* Correspondence: mhguo1@gmail.com; npatsop@gmail.com

⁵Harvard Medical School, Boston, MA 02115, USA

Full list of author information is available at the end of the article

Abstract

Background: 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.

Results: We utilize chromatin accessibility data across hematopoietic cells to identify cell type-specific enrichments of MS genetic signals. We find that CD4 T and B cells are independently enriched for MS genetics and further refine the driver subsets to T_h17 and memory B cells, respectively. We replicate our findings in data from untreated and treated MS patients and find that immunomodulatory treatments suppress chromatin accessibility at driver cell types. Integration of statistical fine-mapping and chromatin interactions nominate numerous putative causal genes, illustrating complex interplay between shared and cell-specific genes.

Conclusions: Overall, our study finds that open chromatin regions in CD4 T cells and B cells independently drive MS genetic signals. 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.

Background

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



© The Author(s). 2022 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

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.

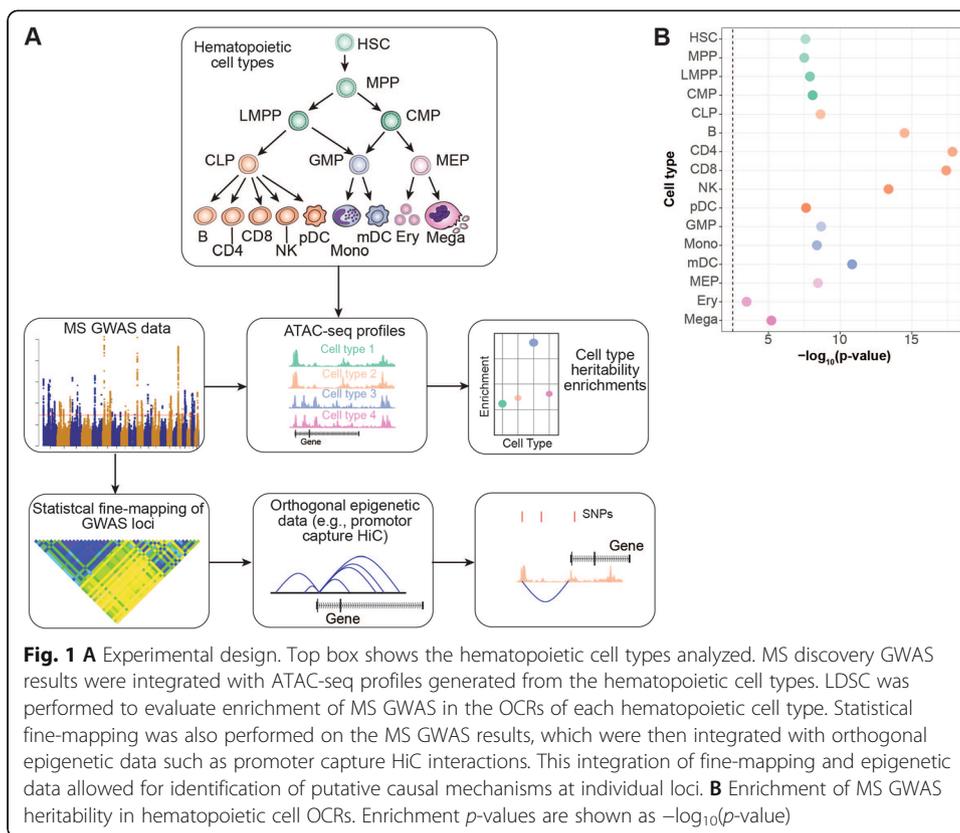
Human genetics has emerged as a powerful tool for probing the underlying biology of a disease [10]. The identification of genes and pathways prioritized by GWAS associations is not constrained by our prior knowledge of disease mechanisms and can therefore identify novel biological mechanisms. However, 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 (Fig. 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) [16, 22] 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.

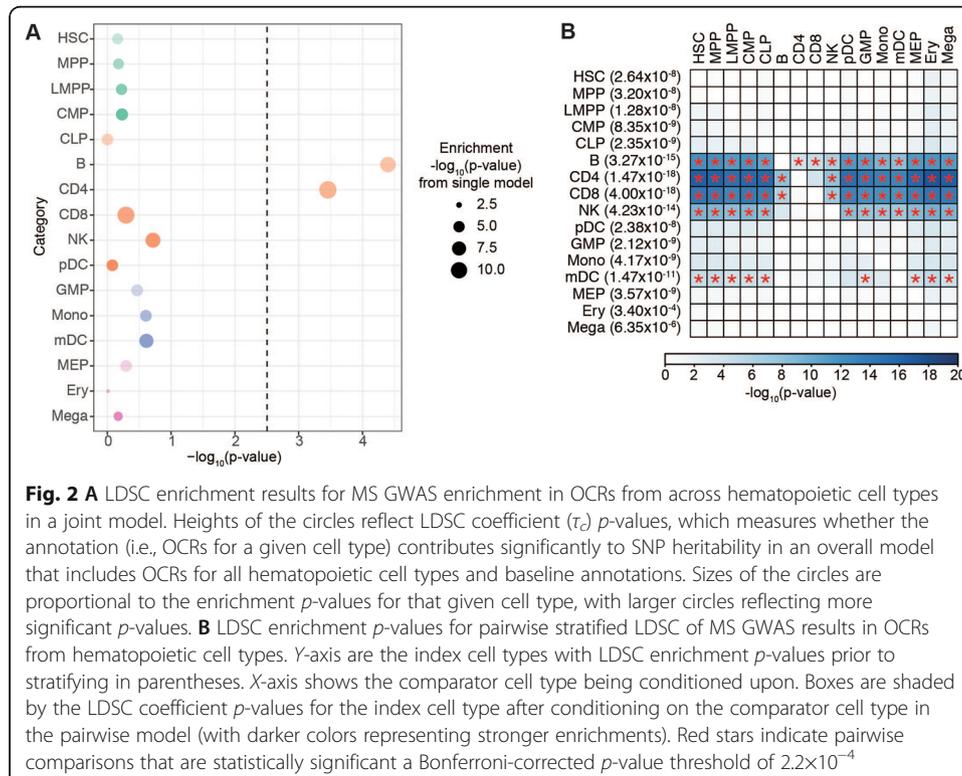
Applying LDSC, we observed strong statistically significant enrichments across all hematopoietic cell populations, even after correcting for multiple hypothesis testing (Bonferroni-corrected p -value threshold of 3.13×10^{-3}) (Fig. 1B; Additional file 1, Table S1). The strongest enrichments for MS GWAS 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, 23]. 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 [23]. 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 (Additional file 2, Fig. 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, we observed that B cells and CD4 T cells contributed significantly to SNP heritability (coefficient p -value = 3.99×10^{-5} and 3.49×10^{-4} , respectively), suggesting independent contributions of B and CD4 T cell OCRs to MS GWAS heritability (Fig. 2A, Additional file 1, Table S2).

To further delineate the cell types with OCRs that specifically mediate the effect of MS GWAS results, we performed a series of pairwise LDSC analyses. In brief, OCRs of a given hematopoietic cell type were stratified against the OCRs of each of the other 15 cell types, as well as the LDSC baseline annotations (Fig. 2B; Additional file 1, Table S3). As above with the joint model, we report the p -value of the coefficient τ_c . We observed that B cells remained significant even after stratifying on OCRs of any of the other 15 cell populations (coefficient p -value ranging from 1.47×10^{-12} when stratifying against HSCs, to 8.33×10^{-5} when stratifying against CD4 T cells). This was also the case for CD4 T cells, which remained significant after stratifying on OCRs from any of the other 15 cell populations (coefficient p -values ranging from 3.82×10^{-17} when stratifying against HSCs, to 2.39×10^{-4} when stratifying against CD8 T cells). In contrast, CD8 T cell OCRs were no longer significant after stratifying against OCRs from CD4⁺ T cells (coefficient p -value = 0.21), though they were significant when stratifying on any of the other cell populations. NK cells were also no longer significant after stratifying against either CD4 T cells (coefficient p -value = 0.165) or against CD8 T cells (coefficient p -value 0.356). These results indicate that the enrichment of both CD8 T cells and NK cells can be largely explained by shared regulatory landscapes that are also



present in CD4 T cells. Prior studies have also suggested a role for monocytes in MS [23, 24]. OCRs from monocytes had an enrichment p -value of 4.17×10^{-9} , but we observed that stratifying on OCRs from CD4 T cells or B cells ameliorated this monocyte heritability enrichment (coefficient p -values 0.166 and 0.266, respectively).

We also performed a separate analysis examining whether OCRs specific to a given cell type mediate MS heritability enrichments. For each of the mature hematopoietic cell populations, we performed LDSC on only the cell type-specific OCRs, i.e., ATAC-seq peaks present in only that cell type. B cells exhibited a statistically significant enrichment of cell-specific OCRs for MS GWAS (enrichment p -value = 3.27×10^{-4}). CD4 T cell-specific peaks were nominally significant at a p -value of 0.013 but did not survive correction for multiple hypothesis testing (Bonferroni-corrected p -value threshold of 5.6×10^{-3}). Cell type-specific peaks for all other terminal hematopoietic cell types had enrichment p -values > 0.05 (Additional file 2, Fig. S2; Additional file 1, Table S4).

MS genetic associations are mediated in terminal immune cell populations

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

Comparison of immune cell enrichment with neuropsychiatric and autoimmune disorders

We next sought to understand how the immune cell enrichments in MS might be similar or different from those of other autoimmune or neuropsychiatric disorders. To test this, we calculated heritability enrichment within these 16 hematopoietic OCRs for GWAS of various other neuropsychiatric disorders and autoimmune disorders: Alzheimer disease (AD) [25], schizophrenia (SCZ) [26], bipolar disorder (BPD) [27], type 1 diabetes (T1D) [28], Crohn's disease (CD) [29], ulcerative colitis (UC) [29], systemic lupus erythematosus (SLE) [30], rheumatoid arthritis (RA) [31], and primary biliary cirrhosis (PBC) [32] (Additional file 2, Fig. S3A; Additional file 1, Table S5). We identified previously recognized cell-type enrichments across these other diseases, such as enrichments in OCRs from B cells (enrichment p -value 1.93×10^{-7}), CD4 T cells (p -value 7.34×10^{-8}), and CD8 T cells (p -value 2.09×10^{-6}) for RA [15, 31]. However, the heritability enrichments for OCRs from these hematopoietic cell populations tended to be much stronger for MS, despite similar sample sizes, e.g., 41,505 disease cases for MS and 38,242 disease cases for RA. To parse out cell type-specific enrichments in these other disorders, we also tested heritability enrichments under the joint model in LDSC by including OCRs from all 16 cell types (Additional file 2, Fig. S3B, Additional file 1,

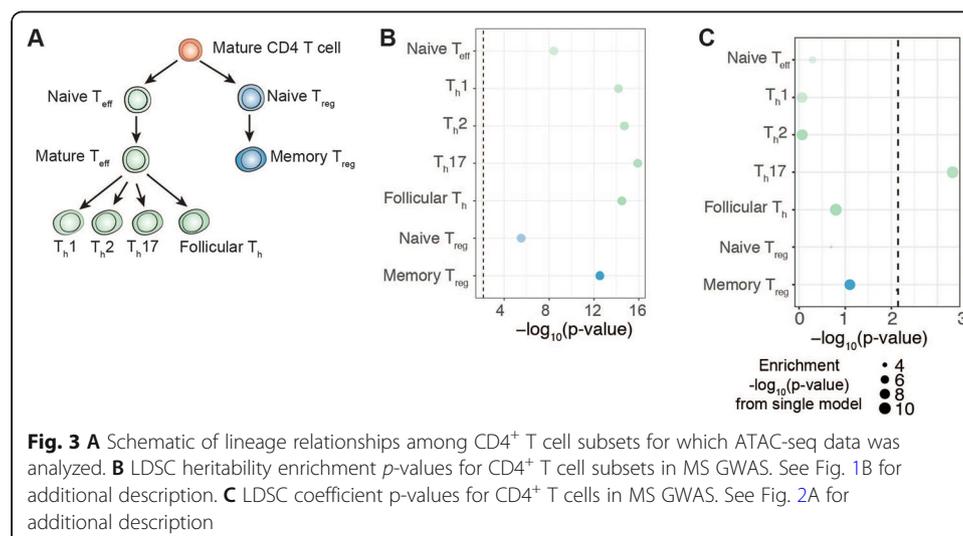
Table S6). In a similar fashion to the MS GWAS joint analyses above, we measured the contribution to heritability for a given set of OCRs of interest, controlling for the effects of a set of baseline annotations and OCRs from all other hematopoietic cell types. Across the nine other comparator diseases, the only other statistically significant stratified enrichments were in OCRs from B cells in SLE GWAS (coefficient p -value 8.53×10^{-5}). These results highlight the striking heritability enrichments in OCRs from CD4 T and B cells in the MS GWAS.

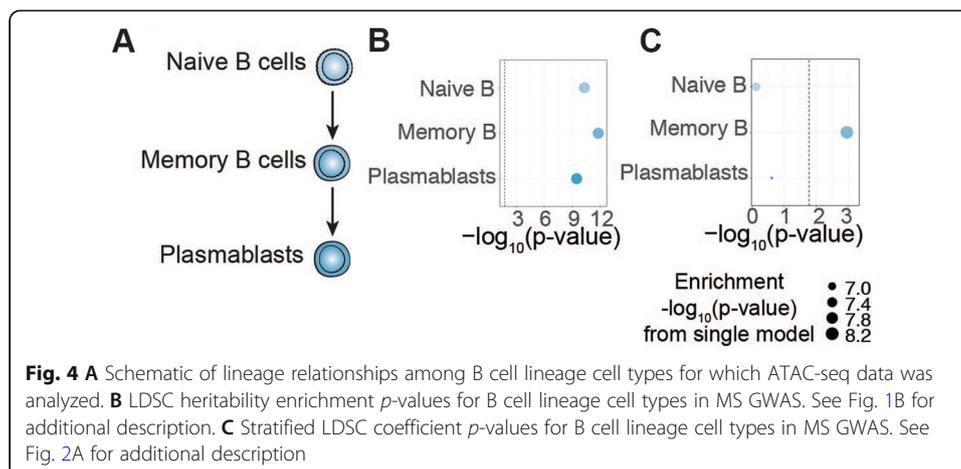
The CD4 T cell MS GWAS enrichment is driven by the Th17 subset

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 (Fig. 3A) [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_{reg} s and memory T_{reg} s). We observed strong enrichments for OCRs from both effector and regulatory CD4 T cell populations (Fig. 3B; Additional file 1, Table S7). 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} ; Fig. 3C; Additional file 1, Table S8). These results suggest that among CD4 T cells, OCRs from T_{h17} cells drive the signal for enrichment in MS GWAS.

Memory subpopulations explain the enrichment of MS GWAS in B cells

We next examined enrichments of MS GWAS data in ATAC-seq from the B cell lineage including naïve B cells, memory B cells, and plasmablasts (Fig. 4A) [21]. We found that OCRs from all B cell lineage cell types were significantly enriched for MS heritability (Fig. 4B; Additional file 1, Table S10). LDSC under a joint model including all B cell lineage cell types revealed an independent contribution from memory B cells (coefficient p -value = 1.10×10^{-3}), but not naïve B cells or plasmablasts (Fig. 4C;





Additional file 1, Table S11). These results suggest that in the B cell lineage, the MS GWAS enrichment signal is driven by OCRs in memory B cells.

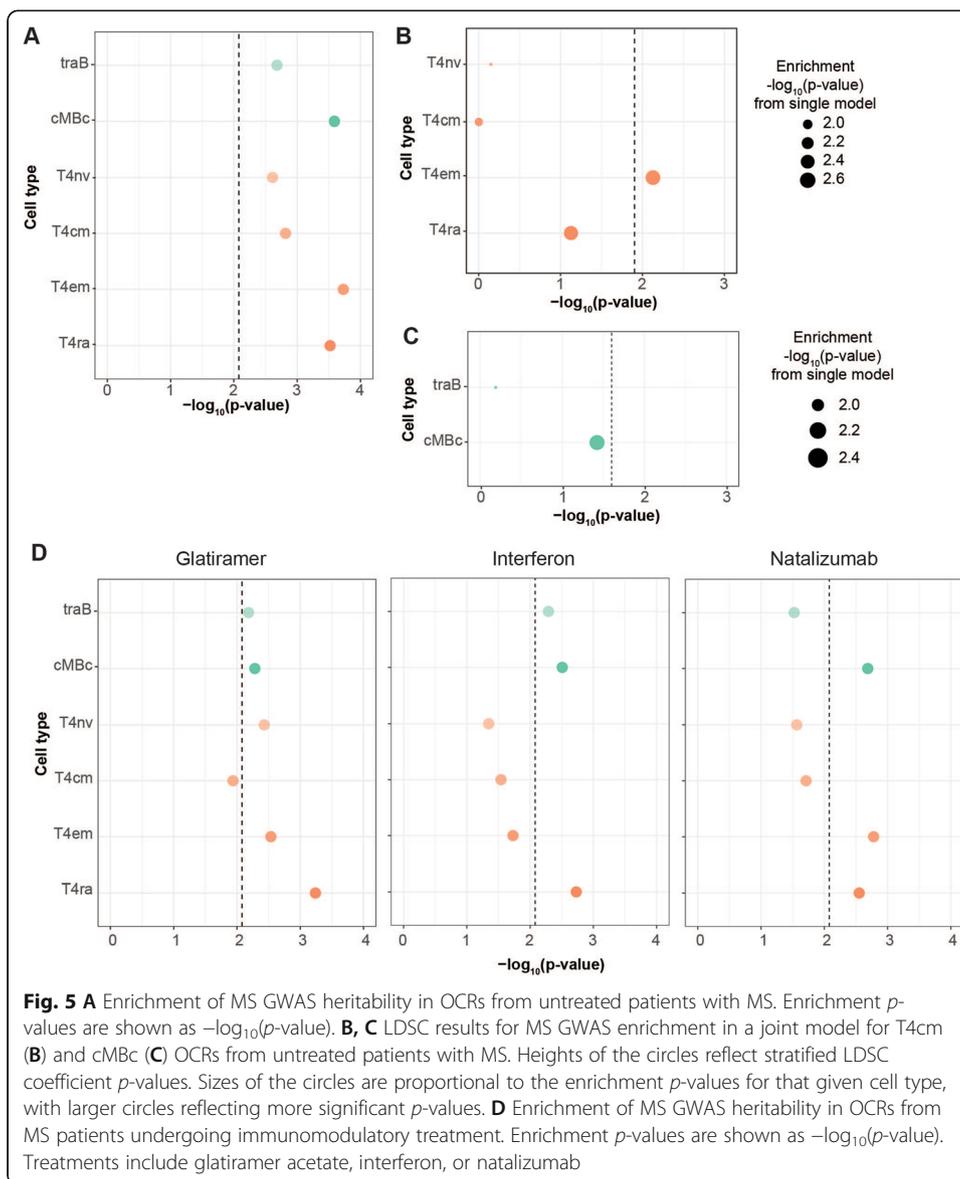
Immune cells from MS patients reinforce independent CD4 T and B cell enrichments

Next, we tested whether the MS GWAS enrichment in CD4⁺ T and B cells were also present in OCRs from the respective immune cell types derived from individuals with MS. We utilized ATAC-seq data performed in flow-sorted bulk CD4 T and B cell subsets ($n = 6$) derived from six patients with MS who were not treated with immunomodulatory therapy within at least 4 months of sample collection (see Additional file 1, Table S13 for clinical details). LDSC showed statistically significant enrichments of transitional B cells (traB; enrichment p -value = 2.08×10^{-3}), class switched classical memory B cells (cMBC; p -value = 2.58×10^{-4}), effector memory CD4 T cells (T4_{em}; p -value = 1.87×10^{-4}), and CD45RA⁺ effector memory CD4 T cells (T4ra; p -value = 3.02×10^{-4}). Central memory CD4 T cells had an enrichment p -value of 1.53×10^{-3} , which was not significant after correcting for multiple hypothesis testing (Fig. 5A; Additional file 1, Table S14).

To further identify independent cell type enrichments, we performed joint models in LDSC as described above. We first tested a model that included CD4 T cell subsets (T4nv, T4cm, T4em, and T4ra). In this joint model, only T4em was statistically significant (coefficient p -value 7.45×10^{-3}), reflecting the independent contribution of effector CD4 T cells in MS that we observed above using cells from healthy individuals (Fig. 5B; Additional file 1, Table S15). We also ran a model that included B cell subsets (traB and cMBC). In this joint comparison, neither cell type was statistically significant when correcting for multiple hypothesis testing. However, cMBC had a coefficient p -value that was nominally significant (p -value = 0.037), reiterating the independent contributions of mature B cell types (Fig. 5C; Additional file 1, Table S16).

Immunomodulatory treatments suppress mediation of MS genetics in cell-specific fashion

Next, we tested whether immunomodulatory treatments alter the cell-specific mediation of MS genetic associations by utilizing data for the same immune subsets sorted



from patients with MS ($n = 3$) under treatment with either natalizumab, interferon, or glatiramer acetate. Following treatment with any of the agents still resulted in statistically significant enrichments of cMBc and T4em (Fig. 5D; Additional file 1, Table S17), though the magnitude of the enrichments were attenuated relative to the signals observed from cells from untreated patients (compare Fig. 5A, D). To better understand this attenuation of enrichments, we ran joint models in LDSC for T4em and cMBc cells in which we included OCRs from untreated patients and treated patients. In a joint model with T4em OCRs from treated and untreated MS patients, only OCRs from untreated patients were statistically significant (Additional file 2, Fig. S4A; Additional file 1, Table S18). Similarly, in a joint model with cMBc OCRs from treated and untreated MS patients, only OCRs from untreated patients were statistically significant (Additional file 2, Fig. S4B; Additional file 1, Table S19). 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 T_h17 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) [33, 34]. In T_h17 cells (Additional file 2, Fig. S5A; Additional file 1, Table S20), 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 (Additional file 2, Fig. S5B; Additional file 1, Table S21), 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 associations 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 [35]. For T_h17 CD4 T cells, the “EnhA2” chromatin state (Active Enhancer 2) were statistically enriched (enrichment p -value = 1.19×10^{-3}) (Additional file 2, Fig. S6A; Additional file 1, Table S22). 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 (Additional file 2, Fig. S6B; Additional file 1, Table S23). These results demonstrate that MS GWAS variants act through activating regulatory elements, consistent with the autoimmune nature of MS.

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

We next sought to understand underlying mechanisms by nominating putative causal genes and variants in a cell-specific fashion. First, we applied statistical fine-mapping to nominate likely causal SNPs. Most statistical fine-mapping approaches require association information across all SNPs in a given locus. In contrast, for MS GWAS, genome-wide results are based on targeted replication analyses, which by design included only a select subset of SNPs within each locus. To overcome this challenge, we applied PICS to perform statistical fine-mapping, which, as compared to other statistical fine-mapping approaches, does not require GWAS summary statistics for all SNPs in a region [15]. We defined for each locus a 95% credible set such that the sum of the posterior probabilities for variants in that credible set is greater than or equal to 95%. For fine-mapping, we used the joint analysis MS GWAS results, which include only replicated genome-wide effects [8]. We included all 200 non-MHC loci where the MS GWAS joint association p -value was less than 5×10^{-8} . Next, we prioritized SNPs if they had a PICS posterior probability (PP) > 1% and were included in the 95% PICS credible set. This is a liberal threshold for defining prioritized SNPs, aimed at increasing sensitivity for detecting possible causal variants. Across the 200 loci, there were 3436 credible set variants (1–58 variants per locus) (Additional file 2, Fig. S7A). At 19 loci, there was only one variant in the credible set, and at 37 loci, there were four or fewer prioritized variants (Additional file 2, Fig. S7B).

Next, we intersected the 3436 credible set variants with OCRs from the 16 hematopoietic cell populations, which we chose to use as they cover a broad range of hematopoietic cell types [18]. 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 (Additional file 2, Fig. S8). We further examined the statistical overlap of the PICS credible set variants with OCRs from CD4 T cell and B cell subpopulations. Among B cells, we found that memory B cells had the strongest enrichment (1.51 fold; p -value 4.33×10^{-16}) and remained statistically significant when examining only OCRs present specifically in memory B cells (Additional file 1, Table S24). Among CD4 T cells, we found that T_{H17} cells had the strongest enrichment (1.72 fold; p -value 2.21×10^{-26}) and also remained statistically significant when examining only OCRs present specifically in T_{H17} cells (Additional file 1, Table S25). These results confirm that statistical fine-mapped variants highlight memory B cell and T_{H17} T cell populations and that statistical fine-mapping results are useful for further functional prioritization.

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 [36]. 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. We examined the statistical enrichment of PICS credible set variants and PCHiC loop regions that overlap an OCR. We found that among all cell types, PCHiC loops from naïve B (enrichment p -value 2.09×10^{-11}) and activated CD4 T cells (p -value 1.77×10^{-14}) had the strongest statistical enrichments (Additional file 1, Table S26). This result confirms that PCHiC from B cells and CD4 T cells are highly enriched for statistically fine-mapped variants and may be useful for further prioritization of variants.

Having demonstrated that PCHiC loops and OCRs map onto MS GWAS variants, we next asked whether eQTLs would colocalize SNPs overlapping these epigenetic features. We examined whether the MS GWAS loci that overlap OCR and loops colocalize more often with cell-specific expression quantitative trait locus (eQTL) studies than expected by chance. We performed colocalization [37] using CD4 T and B cell-specific immune *cis*-eQTL results from DICE [38]. We identified a small number of colocalized loci with eQTL eGenes, ranging from 15 SNP-eGenes pairs for stimulated CD4 T cells to 34 for T_{H17} cells, of which about one third overlapped the previously identified OCR+loops per cell type (Additional file 2, Fig. S9). Naïve Tregs and T_{H17} exhibited the strongest enrichment of colocalized loci that overlap OCR and loops with follicular

T_h, memory Treg, and T_h1/17 also passing the Bonferroni-corrected enrichment threshold (Additional file 2, Fig. S10). We note that we used a less stringent threshold of a colocalization posterior probability of 0.6 for these enrichment analyses to increase the number of observations for examining OCR overlap.

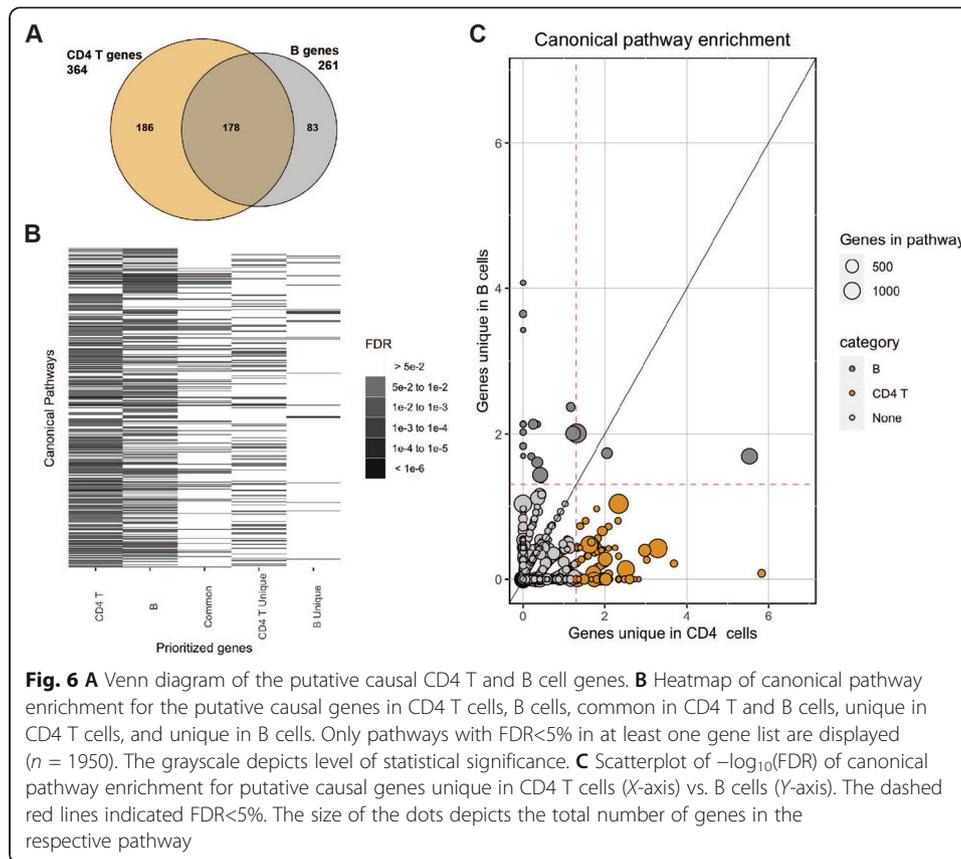
To identify putative causal genes in an unbiased fashion, we integrated the MS fine-mapping data with the OCR and PChIC loops to nominate 261 genes within 86 MS loci in B cells and 364 genes within 115 MS loci in CD4 T cells (Additional file 1, Table S27). We note that these genes are putative causal 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; Fig. 6A). 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 (Additional file 1, Table S28). 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 (Fig. 6B, 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 (Fig. 6B, 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 [39]. A similar percent of the mapped CD4 T cell and B cell prioritized genes were directly connected, 48.9% and 43.7% respectively (Additional file 1, Table S29; Additional file 2, Figs. S11-15). 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; Additional file 1, Table S29). 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.

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

The results of the pathways and PPI analyses suggest that the MS genes common and unique to B and CD4 T cells do not act independently but rather have shared cellular mechanisms. We illustrate this by studying a locus on chromosome 19 (Fig. 7) where the lead SNP rs1465697 (chr19:49837246 C>T) has a MS GWAS association p -value of



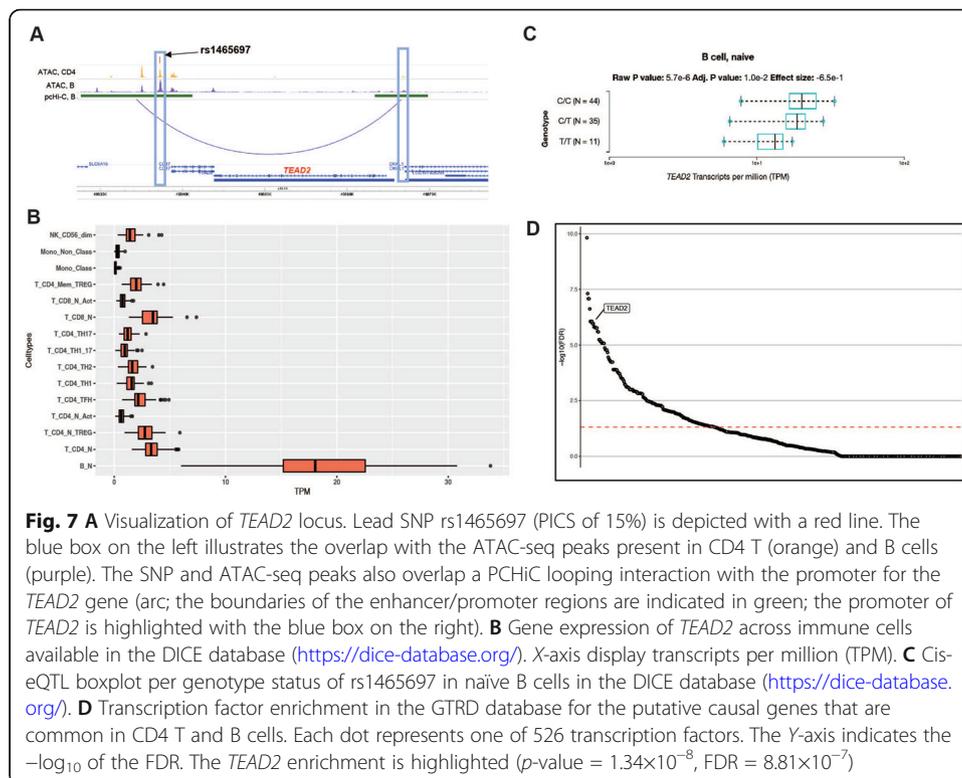
3.02×10^{-18} . The lead SNP is the most highly prioritized variant by statistical fine-mapping out of 21 overall SNPs in the 95% credible set (PP=15% for rs1465697; next highest PP 9%). This SNP overlies an OCR present in all lymphoid lineages, including B cells, CD4 T cells, and CD8 T cells. We have previously suggested five putative causal genes for this locus: *DKKL1*, *CCDC155*, *CD37*, *TEAD2*, and *SLC6A16* [8]. Using PCHiC data, this OCR forms a chromatin loop interaction with *TEAD2* and *DKKL1*, but not the other three genes. This chromatin loop interaction is observed in activated CD4 T cells, naïve B cells, total B cells, naïve CD8 T cells, and fetal thymus, but none of the other hematopoietic cell types [36]. Furthermore, this SNP is an eQTL for *TEAD2* in B cells, but is not an eQTL for any other gene in this locus in any of the available hematopoietic cell types (Fig. 7) [38]. Together, these lines of evidence support *TEAD2* as the causal gene at this locus.

Interestingly, *TEAD2* is a transcription factor with 1459 predicted regulated genes [40], including 38 putatively causal CD4 T cell genes and 25 B cell genes as nominated above (FDR < 1 %, FDR < 1%, respectively; Additional file 2, Figs. S16-17). The majority of these genes are common in both CD4 T and B cells ($n = 23$; FDR < 1%; Additional file 2, Fig. S18). To identify genes whose expression is modulated by *TEAD2* in CD4 T and B cells, we examined *TEAD2* knockdown (KD) and over-expression (OE) in cancer cell lines ($n = 8$, respectively) from the Library of Integrated Network-Based Cellular Signatures (LINCS) Program [41]. Although these cell lines do not represent an ideal experimental model to study the effect of *TEAD2* in immune cells, they can still be used to understand mechanisms reflecting core cellular functions. Within each cell line,

we identified genes whose expression changed in the opposite direction (top or bottom 10%) in the KD versus OE models. These ranged from 7 to 24 for the CD4 T cell prioritized genes (Additional file 1, Table S30; Additional file 2, Fig. S19) and from 3 to 16 for the B cell genes (Additional file 1, Table S31; Additional file 2, Fig. S23). Of these, 17 genes and 9 genes changed expression in at least 2 cell lines, respectively (Additional file 1, Table S30-31). These data demonstrate that perturbation of *TEAD2*, a key immune cell transcription factor, results in indirect changes of putative MS causal genes in B and CD4 T cells.

Discussion

In this paper, we integrated MS GWAS with chromatin accessibility data from a broad array of peripheral immune cells in order to identify putative causal cell types. Our analyses identified regulatory regions in B cells and CD4 T cells as each being independently enriched for MS genetics. Within the CD4 T cell and B cell populations, we further identified OCRs from T_H17 cells and memory B cells as specifically driving their respective enrichments. Chromatin data from MS patients reiterated these findings and further suggested that immunomodulatory treatments alter the chromatin accessibility overlying MS-associated GWAS variants. Integration of PCHiC data led to prioritization of putative causal genes in B and CD4 T cells, identifying target genes that are both shared and specific to each of these cell populations. The putative causal genes implicate several known signaling pathways, mostly due to the cell-specific MS-associated genes, despite these representing a smaller percentage compared to genes shared between B and CD4 T cells. Finally, we illustrate that the B and CD4 T cell mechanisms are intertwined by describing how *TEAD2*, a putative causal gene shared



between B and CD4 T cells, contributes to disease susceptibility directly and indirectly by targeting both shared and cell-specific MS genes.

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 [23, 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 [23, 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 auto-immune 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 [23], 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. Another potential caveat is that we utilized bulk sorted cell populations, which may miss subpopulations of cells that bear chromatin signatures that mediate independent effects of MS genetics. One approach to overcome this is using single-cell epigenetic approaches, which will be an important area of future exploration. 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 variant. Second, as most GWAS variants are noncoding, identifying the gene(s) that are affected by the causal variant can be difficult. Third, the cell type(s) in which a given associated variant acts can be unclear. Our study demonstrates how we can use statistical fine-mapping to help solve the first challenge, though this is not without multiple caveats [46]. We integrated orthogonal datasets (ATAC-seq, PChIC, and eQTL data) to help delineate the likely causal genes and cell types and overcome the latter two challenges. We document how shared and cell-specific genes affect putative causal pathways. We further illustrate the complex interplay between shared and cell-specific putative causal MS genes by studying the *TEAD2* locus, a transcription factor recently implicated in immune regulation [47].

Conclusions

Together, our study generates important insights into the driver subpopulations of peripheral immune cells in MS, reinforcing how MS genetics act primarily through B and

CD4 T cells. Our study also demonstrates the need for in-depth context-specific cellular data to carefully delineate the causal role of each immune cell subset in MS.

Methods

GWAS summary statistics

We utilized available MS GWAS summary statistics, which included data from 8,278,136 variants across 14,802 individuals with MS (cases) and 26,703 individuals without MS (controls) [8]. For enrichment analyses, we included only the 6,773,531 variants that were analyzed in all 15 cohorts of the discovery stage meta-analysis; this resulted in 6,773,531 variants carried forward. We additionally utilized GWAS summary statistics from various neuropsychiatric disorders or autoimmune disorders: Alzheimer disease (AD) [25], schizophrenia (SCZ) [26], bipolar disorder (BPD) [27], type 1 diabetes (T1D) [28], Crohn's disease (CD) [29], ulcerative colitis (UC) [29], systemic lupus erythematosus (SLE) [30], rheumatoid arthritis (RA) [31], and primary biliary cirrhosis (PBC) [32]. All GWAS data were converted to the ".sumstats" format as required by 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 Buenroostro 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 [33]. 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 downloaded from the Roadmap Epigenomics Project [34] (NCBI GEO GSM997225); we used pre-processed ChIP-seq peak calls generated in Amariuta et al. [51].

chromHMM

We used a 25 chromatin state model [35], which are imputed based on 12 epigenetic marks from across 127 epigenomes generated as part of the Roadmap Epigenomics Project [34]. We used the chromatin states from B cells and Th17 CD4+ T cells. Chromatin states were downloaded from https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html. We excluded the “Quiescent/Low” cell state, as it encompasses a large proportion of the genome, resulting in unstable estimates of heritability.

Verily ATAC-seq data

We utilized immune cell ATAC-seq data generated from Verily as part of the SystemS collaboration with Brigham & Women’s Hospital (PIs: Dr. Chitnis and Dr. Weiner).

Cell sorting is done as follows: Frozen cryovials in liquid nitrogen were thawed in a 37 °C bead bath and centrifuged for 5 min at 600×g, 4 °C. The cell pellet was washed with 1 mL of FACS buffer, and the wash repeated. The cell pellet was resuspended in the residual volume with 2.5 µL of 0.33 mg/mL S7 DNase. Fifty microliters of staining cocktail was added for the respective flow cytometry panels to be analyzed (T cell, B cell, myeloid panel) and incubated for 25 min on ice and in the dark. Cells were washed in FACS buffer, resuspended in a final volume of 400 µL FACS buffer, and passed through a 35-µm cell strainer cap. Stained samples were sorted on a FACSAria Fusion (BD Biosciences, San Jose, CA). Using FACSDiva v8.0.1 software, the samples were gated first by forward and side scatter properties, then FSC-H vs FSC-A for singlet discrimination, and finally, with their respective markers for each cell type (Additional file 1, Table S32). For each cell type of interest, up to 500 cells were sorted into the tagmentation buffer.

ATAC-seq library preparation and sequencing

Cells were sorted directly into 20 µL of cold tagmentation buffer (10 µL TD, 2 µL 2% IGEPAL CA-630, 6 µL nuclease-free H₂O, 2 µL TDE1 per sample), followed by incubation at 37 °C for 30 min with shaking at 500 RPM. Samples were stored at –20 °C 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 NGmerge⁶⁰ using the default parameters. The reads were then aligned to GRCh38 using Bowtie2 version 2.3.5⁴⁹. The resulting SAM files were converted and sorted into BAM format using samtools version

1.5⁵⁰. 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 2.2.5⁵¹ 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 2.29⁶¹.

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, 22]. 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 (χ^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 pairwise stratified 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 $p < 5 \times 10^{-8}$ 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 panel [52]. PLINK v1.90b3.32 was used to perform LD calculations [55, 56] with parameters of “--r2 --ld-window-kb 2000 --ld-window 999999 --ld-window-r2 0.2.”

We then applied PICS to each locus [15]. Briefly, PICS uses the lead association p -value and LD structure of the locus to calculate the most likely causal SNPs given the observed lead association signal. PICS probabilities represent the probability of a given SNP in a locus being the causal SNP. Default PICS parameters were used. From the PICS probabilities, we calculated 95% credible sets (CS). We defined the 95% CS as a set of variants such that the true causal variant has a 95% chance of being in the credible set. To calculate credible sets, for each locus, we ranked variants in descending order by their PICS probabilities. We then iteratively added variants to the credible set for that locus until the sum of their PICS probabilities was greater than or equal to 0.95. For CS inclusion, we also required the variant to have a PICS probability > 0.1 .

Identification of target genes

We leveraged promoter capture Hi-C (PCHiC) data from 17 hematopoietic cell populations to link genetic associations with genes that they may regulate [36]. We filtered the PCHiC dataset for looping interactions with a CHiCAGO score > 5 [57]. An overlap between a GWAS variant and a PCHiC looping interaction was considered if the GWAS variant overlapped any position in the non-promoter (“other end”) of the PCHiC interaction. For CD4⁺ T cells, we considered only GWAS SNPs that overlapped an ATAC-seq peak in bulk CD4⁺ T cells or any of the CD4⁺ T cell subsets (naïve effector CD4⁺ T cells, T_h1, T_h2, T_h17, follicular T_h, naïve T_{regs}, and memory T_{regs}), and which overlapped a PCHiC interaction in naïve CD4⁺ T cells (nCD4), total CD4⁺ T cells (tCD4), non-activated total CD4⁺ T cells (naCD4), or activated total CD4⁺ T cells (aCD4). For B 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).

Colocalization of MS GWAS loci with eQTL studies

We performed colocalization analyses utilizing coloc [37], MS GWAS summary statistics [8], and cis-eQTL analyses of CD4 T and B cells from the DICE database [38]. We included the following subsets: T_h1, T_h2, T_h1/17 (labeled as THSTAR in the provided summary statistics), T_h17, naïve CD4, stimulated CD4, naïve Tregs, memory Tregs, and naïve B cells. Sorting details are available in the DICE original publication [38]. For the MS GWAS, we included per locus all variants within a ± 100 kb region. For the eQTL studies, we included variants with any p -value around ± 1 megabase of the transcription starting site (TSS), as provided by the DICE authors, and any eGene with a significant eQTL association at 10% FDR. Coloc was run on the common set of variants and we defined a priori as colocalized a locus-eGene pair that had a posterior probability of both traits being associated and share a single causal variant (H4) larger than 60%. Per cell subtype OCR+loop enrichment was estimated with Fisher’s exact for the 2×2 table of a locus-eGene being colocalized vs. overlapping OCR+loop within either CD4 T cells or B cells.

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 set categories together in the same model. We estimated statistical significance using the hypergeometric 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>) [39] to leverage known protein-protein interactions (PPI) of our prioritized gene sets. GeNets uses a random forest classifier, 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 exceeds 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 knockdown (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 perturbation, adjusted for the gene expression of all other perturbations on the same physical plate. For knockdown and over-expression experiments, the differential expression comparators 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).

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-022-02694-y>.

Additional file 1: Supplementary Tables. **Table S1.** MS GWAS enrichment in 16 hematopoietic cell types. **Table S2.** MS GWAS enrichment of hematopoietic cell types in joint model. **Table S3.** MS GWAS pairwise enrichment in 16 hematopoietic cell types. **Table S4.** MS GWAS cell-specific enrichment within terminal hematopoietic cell types. **Table S5.** Heritability enrichments (single model) for other disorders. **Table S7.** MS GWAS enrichment in CD4+ T cell subpopulations. **Table S8.** Stratified LDSC in CD4+ T cell subpopulations. **Table S9.** MS GWAS pairwise enrichment in CD4+ T subpopulations. **Table S10.** MS GWAS enrichment in B cell subpopulations. **Table S11.** Stratified LDSC in B cell subpopulations. **Table S12.** MS GWAS pairwise enrichment in B cell subpopulations. **Table S13.** Clinical characteristics of MS subjects. **Table S14.** MS GWAS enrichment across OCRs in 6 cell types isolated from untreated MS patients. **Table S15.** MS GWAS enrichment across OCRs in joint model in CD4 T cell types isolated from untreated MS patients. **Table S16.** MS GWAS enrichment across OCRs in joint model in B cell types isolated from untreated MS patients. **Table S17.** MS GWAS enrichment across OCRs in 6 cell types isolated from treated MS patients. **Table S18.** MS GWAS enrichment across OCRs in joint model in T4em isolated from untreated and treated MS patients. **Table S19.** MS GWAS enrichment across OCRs in joint model in cMBC isolated from untreated and treated MS patients. **Table S20.** MS GWAS enrichments in histone ChIP-seq from Th17 cells. **Table S21.** MS GWAS enrichments in histone ChIP-seq from B cells. **Table S20.** MS GWAS enrichments in chromatin states from Th17 cells. **Table S21.** MS GWAS enrichments in chromatin states from B cells. **Table S24.** MS fine-mapping enrichments in ATAC-seq OCRs. **Table S25.** Pairwise enrichments of MS fine-mapped SNPs in OCRs Index cell types are shown on the left. **Table S26.** MS fine-mapping enrichments in PCHIC + ATAC-seq OCRs. **Table S27.** Summary of gene prioritizations. **Table S28.** Canonical pathway enrichment for prioritized MS genes. **Table S29.** Protein-protein interaction connectivity summaries for prioritized gene lists. **Table S30.** CD4 T prioritized genes ranked in the opposite extreme 10% of gene expression changes in KD and OE cell lines. **Table S31.** B prioritized genes ranked in the opposite extreme 10% of gene expression changes in KD and OE cell lines. **Table S32.** Markers for FACS sorting strategy of Verily ATAC-seq data.

Additional file 2: Supplementary Figures. **Figure S1.** Correlation in ATAC-seq profiles across hematopoietic cell types. **Figure S2.** LDSC enrichments for MS GWAS in cell-type specific ATAC-seq peaks mature hematopoietic cell type. **Figure S3.** Enrichments of GWAS results from 10 neuropsychiatric or autoimmune conditions in OCRs across various hematopoietic cell types. **Figure S4.** LDSC results for MS GWAS in ATAC-seq peaks from treated MS patients. **Figure S5.** LDSC results for MS GWAS in histone ChIP-seq. **Figure S6.** LDSC results for MS GWAS in chromHMM partitions. **Figure S7.** Distribution of credible set variants. **Figure S8.** Number of GWAS loci with ATAC-seq peak overlapping fine-mapped SNP. **Figure S9.** Colocalization of MS GWAS loci with DICE CD4 T and B cell eQTLs. **Figure S10.** Colocalization enrichment of MS GWAS loci with DICE CD4 T and B cell eQTLs. **Figure S11.** Protein-protein interaction communities of putative causal genes in CD4 T cells. **Figure S12.** Protein-protein interaction communities of putative causal genes in B cells. **Figure S13.** Protein-protein interaction communities of putative causal genes shared in CD4 T and B cells. **Figure S14.** Protein-protein interaction communities of putative causal genes unique in CD4 T cells. **Figure S15.** Protein-protein interaction communities of putative causal genes unique in B cells. **Figure S16.** Enrichment of CD4 T cell putative causal genes in GTRD database. **Figure S17.** Enrichment of B cell putative causal genes in GTRD database. **Figure S18.** Enrichment of shared between CD4 T and B cells putative causal genes in GTRD database. **Figure S19.** Change of gene expression of CD4 T cell putative causal genes in knock-down (KD) and over-expression (OE) models in cancer cell lines. **Figure S20.** Change of gene expression of B cell putative causal genes in knock-down (KD) and over-expression (OE) models in cancer cell lines.

Additional file 3. Review history.

Acknowledgements

We thank Jacob Ulirsch for technical assistance with the use of LD score regression.

Peer review information

Anahita Bishop and Kevin Pang were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Review history

The review history is available as Additional file 3.

Authors' contributions

MHG and NAP initiated 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 gave final approval of the paper.

Funding

NAP was supported in part by National Multiple Sclerosis Society (grants JF-1808-32223 and RG-1707-28657). MHG was supported in part by NIH R25 NS065745. This work was supported in part by the Water Cove Charitable Foundation.

Availability of data and materials

MS GWAS summary statistics are available via request to the IMSCG (<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 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>. The Verily Life Sciences ATAC-seq data from the SystemeMS are available from NCBI GEO (accession GSE202087) [58]. Coloc results have been deposited in Zenodo: 10.5281/zenodo.5850919 [59]. Code used in this paper has been deposited in Zenodo: 10.5281/zenodo.6523097 [60].

Declarations

Ethics approval and consent to participate

This study was approved by the Mass Gen Brigham (MGB) Human Research Committee (IRB #2012p123456). All study participants provided written informed consent.

Consent for publication

Not applicable.

Competing interests

JB, CC, XD, PK, CCK, TO, TS, and DZS are employed by Verily Life Sciences at time of study. NAP is currently an employee of Novartis Institutes for BioMedical Research.

Author details

¹Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. ²Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA 02142, USA. ³Systems Biology and Computer Science Program, Ann Romney Center for Neurological Diseases, Department of Neurology, Brigham & Women's Hospital, Boston, MA 02115, USA. ⁴Division of Genetics, Department of Medicine, Brigham & Women's Hospital, Harvard Medical School, Boston, MA 02115, USA. ⁵Harvard Medical School, Boston, MA 02115, USA. ⁶Brigham Multiple Sclerosis Center, Brigham and Women's Hospital, Boston, MA 02115, USA. ⁷Ann Romney Center for Neurologic Diseases, Brigham and Women's Hospital, Boston, MA 02115, USA. ⁸Verily Life Sciences, South San Francisco 94080, USA.

Received: 23 August 2021 Accepted: 16 May 2022

Published online: 07 June 2022

References

- Filippi M, Bar-Or A, Piehl F, Preziosa P, Solari A, Vukusic S, et al. Multiple sclerosis. *Nat Rev Dis Prim*. 2018;4(1):43. <https://doi.org/10.1038/s41572-018-0041-4>.
- Baecher-Allan C, Kaskow BJ, Weiner HL. Multiple sclerosis: mechanisms and immunotherapy. *Neuron*. 2018;97:742–68. <https://doi.org/10.1016/j.neuron.2018.01.021>.
- van Langelaar J, Rijvers L, Smolders J, van Luijn MM. B and T cells driving multiple sclerosis: identity, mechanisms and potential triggers. *Front Immunol*. 2020;11:760. <https://doi.org/10.3389/fimmu.2020.00760>.
- Chihara N. Dysregulated T cells in multiple sclerosis. *Clin Exp Neuroimmunol*. 2018;9:20–9. <https://doi.org/10.1111/cen3.12438>.
- Negron A, Robinson RR, Stüve O, Forsthuber TG. The role of B cells in multiple sclerosis: Current and future therapies. *Cell Immunol*. 2019;339:10–23. <https://doi.org/10.1016/j.cellimm.2018.10.006>.
- Greenfield AL, Hauser SL. B-cell therapy for multiple sclerosis: entering an era. *Ann Neurol*. 2018;83:13–26. <https://doi.org/10.1002/ana.25119>.
- Hauser SL, Bar-Or A, Comi G, Giovannoni G, Hartung H-P, Hemmer B, et al. Ocrelizumab versus interferon beta-1a in relapsing multiple sclerosis. *N Engl J Med*. 2017;376(3):221–34. <https://doi.org/10.1056/NEJMoa1601277>.
- International Multiple Sclerosis Genetics Consortium, Patsopoulos NA, Baranzini SE, Santaniello A, Shoostari P, Cotsapas C, et al. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science*. 2019;365:eaav7188. <https://doi.org/10.1126/science.aav7188>.
- International Multiple Sclerosis Genetics Consortium. Low-frequency and rare-coding variation contributes to multiple sclerosis risk. *Cell*. 2018;175:1679–87. <https://doi.org/10.1016/j.cell.2018.09.049>.
- Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet*. 2017;101:5–22. <https://doi.org/10.1016/j.ajhg.2017.06.005>.
- Gusev A, Lee SH, Trynka G, Finucane H, Vilhjálmsón BJ, Xu H, et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am J Hum Genet*. 2014;95(5):535–52. <https://doi.org/10.1016/j.ajhg.2014.10.004>.
- Cano-Gamez E, Trynka G. From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. *Front Genet*. 2020;11:424. <https://doi.org/10.3389/fgene.2020.00424>.
- Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWAS: illuminating the dark road from association to function. *Am J Hum Genet*. 2013;93(5):779–97. <https://doi.org/10.1016/j.ajhg.2013.10.012>.
- Gallagher MD, Chen-Plotkin AS. The post-GWAS Era: from association to function. *Am J Hum Genet*. 2018;102(5):717–30. <https://doi.org/10.1016/j.ajhg.2018.04.002>.
- Farh KK-H, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*. 2015;518(7539):337–43. <https://doi.org/10.1038/nature13835>.

16. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet.* 2015;47(11):1228–35. <https://doi.org/10.1038/ng.3404>.
17. International Multiple Sclerosis Genetics Consortium. A systems biology approach uncovers cell-specific gene regulatory effects of genetic associations in multiple sclerosis. *Nat Commun.* 2019;10:2236. <https://doi.org/10.1038/s41467-019-09773-y>.
18. Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet.* 2016;48(10):1193–203. <https://doi.org/10.1038/ng.3646>.
19. Buenrostro JD, Corces MR, Lareau CA, Wu B, Schep AN, Aryee MJ, et al. Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation. *Cell.* 2018;173:1535–48. <https://doi.org/10.1016/j.cell.2018.03.074>.
20. Ulirsch JC, Lareau CA, Bao EL, Ludwig LS, Guo MH, Benner C, et al. Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nat Genet.* 2019;51(4):683–93. <https://doi.org/10.1038/s41588-019-0362-6>.
21. Calderon D, Nguyen MLT, Mezger A, Kathiria A, Müller F, Nguyen V, et al. Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat Genet.* 2019;51(10):1494–505. <https://doi.org/10.1038/s41588-019-0505-9>.
22. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015;47(3):291–5. <https://doi.org/10.1038/ng.3211>.
23. Dendrou CA, Fugger L, Friese MA. Immunopathology of multiple sclerosis. *Nat Rev Immunol.* 2015;15(9):545–58. <https://doi.org/10.1038/nri3871>.
24. 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;13:355. <https://doi.org/10.3389/fncel.2019.00355>.
25. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet.* 2013;45(12):1452–8. <https://doi.org/10.1038/ng.2802>.
26. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature.* 2014;511(7510):421–7. <https://doi.org/10.1038/nature13595>.
27. 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):977–83. <https://doi.org/10.1038/ng.943>.
28. Bradfield JP, Qu H-Q, Wang K, Zhang H, Sleiman PM, Kim CE, et al. A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. *PLoS Genet.* 2011;7(9):e1002293. <https://doi.org/10.1371/journal.pgen.1002293>.
29. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature.* 2012;491(7422):119–24. <https://doi.org/10.1038/nature11582>.
30. Bentham J, Morris DL, Graham DSC, Pinder CL, Tomblason P, Behrens TW, 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):1457–64. <https://doi.org/10.1038/ng.3434>.
31. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature.* 2014;506(7488):376–81. <https://doi.org/10.1038/nature12873>.
32. Cordell HJ, Han Y, Mells GF, Li Y, Hirschfeld GM, Greene CS, et al. International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. *Nat Commun.* 2015;6:8019. <https://doi.org/10.1038/ncomms9019>.
33. 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(7414):57–74. <https://doi.org/10.1038/nature11247>.
34. Kundaje A, Meuleman W, Ernst J, Bilienky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015;518(7539):317–30. <https://doi.org/10.1038/nature14248>.
35. Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc.* 2017;12(12):2478–92. <https://doi.org/10.1038/nprot.2017.124>.
36. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell.* 2016;167:1369–84. <https://doi.org/10.1016/j.cell.2016.09.037>.
37. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 2014;10(5):e1004383. <https://doi.org/10.1371/journal.pgen.1004383>.
38. Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J, et al. Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell.* 2018;175:1701–15. <https://doi.org/10.1016/j.cell.2018.10.022>.
39. Li T, Kim A, Rosenbluh J, Horn H, Greenfield L, An D, et al. GeNets: a unified web platform for network-based genomic analyses. *Nat Methods.* 2018;15(7):543–6. <https://doi.org/10.1038/s41592-018-0039-6>.
40. Yevshin I, Sharipov R, Kolmykov S, Kondrakhin Y, Kolpakov F. GTRD: a database on gene transcription regulation - 2019 update. *Nucleic Acids Res.* 2019;47(D1):D100–5. <https://doi.org/10.1093/nar/gky1128>.
41. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A next generation connectivity map: L1000 Platform and the First 1,000,000 Profiles. *Cell.* 2017;171(6):1437–52. <https://doi.org/10.1016/j.cell.2017.10.049>.
42. Jelcic I, Al Nimer F, Wang J, Lentsch V, Planas R, Jelcic I, et al. Memory B cells activate brain-homing, autoreactive CD4+ T cells in multiple sclerosis. *Cell.* 2018;175:85–100. <https://doi.org/10.1016/j.cell.2018.08.011>.
43. Li R, Rezk A, Miyazaki Y, Hilgenberg E, Touil H, Shen P, et al. Proinflammatory GM-CSF-producing B cells in multiple sclerosis and B cell depletion therapy. *Sci Transl Med.* 2015;7:310ra166. <https://doi.org/10.1126/scitranslmed.aab4176>.
44. Kebir H, Ifergan I, Alvarez JI, Bernard M, Poirier J, Arbour N, et al. Preferential recruitment of interferon- γ -expressing T H 17 cells in multiple sclerosis. *Ann Neurol.* 2009;66(3):390–402. <https://doi.org/10.1002/ana.21748>.

45. Kebir H, Kreymborg K, Ifergan I, Dodelet-Devillers A, Cayrol R, Bernard M, et al. Human TH17 lymphocytes promote blood-brain barrier disruption and central nervous system inflammation. *Nat Med.* 2007;13(10):1173–5. <https://doi.org/10.1038/nm1651>.
46. Benner C, Havulinna AS, Järvelin M-R, Salomaa V, Ripatti S, Pirinen M. Prospects of fine-mapping trait-associated genomic regions by using summary statistics from genome-wide association studies. *Am J Hum Genet.* 2017;101(4):539–51. <https://doi.org/10.1016/j.ajhg.2017.08.012>.
47. Bai X, Huang L, Niu L, Zhang Y, Wang J, Sun X, et al. Mst1 positively regulates B-cell receptor signaling via CD19 transcriptional levels. *Blood Adv.* 2016;1(3):219–30. <https://doi.org/10.1182/bloodadvances.2016000588>.
48. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357–9. <https://doi.org/10.1038/nmeth.1923>.
49. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
50. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008;9(9):R137. <https://doi.org/10.1186/gb-2008-9-9-r137>.
51. Amariuta T, Luo Y, Gazal S, Davenport EE, van de Geijn B, Ishigaki K, 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):879–95. <https://doi.org/10.1016/j.ajhg.2019.03.012>.
52. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, MA DP, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012;491:56–65. <https://doi.org/10.1038/nature11632>.
53. Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis ET, Schaffner SF, Yu F, et al. Integrating common and rare genetic variation in diverse human populations. *Nature.* 2010;467(7311):52–8. <https://doi.org/10.1038/nature09298>.
54. Finucane HK, Reshef YA, Anttila V, Slowikowski K, Gusev A, Byrnes A, et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet.* 2018;50(4):621–9. <https://doi.org/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. <https://doi.org/10.1186/s13742-015-0047-8>.
56. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559–75. <https://doi.org/10.1086/519795>.
57. Cairns J, Freire-Pritchett P, Wingett SW, Várnai C, Dimond A, Plagnol V, et al. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.* 2016;17(1):127. <https://doi.org/10.1186/s13059-016-0992-2>.
58. Guo MH, Sama P, LaBarre BA, Lokhande H, Balibalos J, Chu C, et al. Dissection of multiple sclerosis genetics identifies B and CD4+ T cells as driver cell subsets. *Datasets. Gene Expression Omnibus.* <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE202087>. 2022.
59. Guo MH, Sama P, LaBarre BA, Lokhande H, Balibalos J, Chu C, et al. Dissection of multiple sclerosis genetics identifies B and CD4+ T cells as driver cell subsets. *Zenodo.* 2022. <https://doi.org/10.5281/zenodo.5850919>.
60. Guo MH, Sama P, LaBarre BA, Lokhande H, Balibalos J, Chu C, et al. Dissection of multiple sclerosis genetics identifies B and CD4+ T cells as driver cell subsets. *Zenodo.* 2022. <https://doi.org/10.5281/zenodo.6523097>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

