Genome-wide association analyses highlight the neuronal contribution to multiple sclerosis susceptibility

Lu Zeng[1§], Atlas Khan[2§], Kathryn C. Fitzgerald[3], Tsering Lama[1], Jessy Chen[1], Ruixi Li[4], International Multiple Sclerosis Genetics Consortium[†], Ellen A. Tsai[5], Ketian Yu[5], Tanuja Chitnis[6], Howard L. Weiner[6], Quentin Le Grand[7], Stéphanie Debette[7,8], Gao Wang[4], Masashi Fujita[1], Peter A. Calabresi[3], Frauke Zipp[9], Mariko Taga[1], Krzysztof Kiryluk[2] & Philip L. De Jager[1]*

[1]Center for Translational and Computational Neuroimmunology & Columbia Multiple Sclerosis Center, Department of Neurology, Columbia University Irving Medical Center, New York, NY, USA.
[2]Division of Nephrology, Dept of Medicine, Vagelos College of Physicians & Surgeons, Columbia University, New York, NY, USA.
[3]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA.
[4]The Gertrude H. Sergievsky Center and the Department of Neurology, Columbia University, New York, NY, USA.
[5]Biogen, Biomarkers and Systems Biology, Cambridge, MA, USA.
[6]Ann Romney Center for Neurologic Diseases and Brigham Multiple Sclerosis Center, Department of Neurology, Brigham & Women's Hospital, Mass General Brigham, Boston, MA, USA.
[7]University of Bordeaux, INSERM, Bordeaux Population Health research center, UMR1219, Bordeaux, France.
[8]Bordeaux University Hospital, Department of Neurology, Institute for Neurodegenerative Diseases, Bordeaux, France.
[9]Department of Neurology, Research Center for Immunotherapy (FZI) and Focus Program Translational Neuroscience (FTN), Rhine-Main Neuroscience Network (rmn2), University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany.

[§]These authors contributed equally to this work.
[†]A list of consortium authors appears at the end of the paper.

*Correspondence:
Philip L. De Jager
pld2115@cumc.columbia.edu

Abstract

Multiple sclerosis (MS) is a chronic inflammatory and neurodegenerative disease. Prior genetic studies have identified susceptibility loci that primarily impact immune cells and microglia. Here we performed a multi-ancestry genome-wide association study of 20,831 MS cases and 729,220 controls and identified 236 susceptibility variants outside the major histocompatibility complex, including four novel genomic loci. We also derived a polygenic score for MS; while optimized for European ancestry, it is informative for African-American and Latino individuals. Integrating single-cell data from blood and brain tissue, we identified 76 candidate causal genes. Notably, inhibitory neurons emerged as a key target cell type for MS-associated variants, with seven loci, including *STAT3*, displaying altered expression only in these cells. The *STAT3* variant is also associated with cognition and white matter integrity in non-MS individuals and greater sNfL levels in individuals with MS, suggesting that MS susceptibility may reflect reduced central nervous system resilience to inflammatory challenges.

Introduction

The genetic architecture of multiple sclerosis (MS) has become clearer over the past decade. Studies of MS susceptibility have identified more than 233 independent risk variants[1], and a recent study also reported one genome-wide significant disease severity locus[2]. While the functional consequences of some susceptibility variants have been characterized – such as the protective effect of rs2300747^{G} within the *CD58* locus[3,4] and the risk allele rs1800693-G in *TNFRSF1A*[5] – most variants remain poorly understood, and systematic efforts to map their functional effects remain limited[1,6–10]. Functional consequences of MS variants have been found primarily in peripheral immune cells and in microglia, the resident mesoderm-derived immune cells of the central nervous system (CNS)[1]. While some effects have been noted in non-immune cells, such as astrocytes, in targeted analyses[11–13], these studies highlight an important challenge in functional genomics: risk variants may influence function in multiple cell types and subtypes, creating ambiguity about the causal cell type or combination of cell types involved in exerting a variant's effect.

Despite some suggestions[14–16], there remains limited evidence that neuroectodermal derivatives within the CNS contribute to MS onset. Rather, the predominance of an initial peripheral auto-inflammatory response is supported by the fact that approximately half of MS susceptibility variants are shared with other autoimmune diseases[17,18]. However, the functional consequences of the remaining MS-specific susceptibility variants are less well understood. These variants may contribute to mechanisms that direct autoimmune responses towards the CNS.

Here, we investigated whether MS susceptibility variants exert functional consequences in neuronal and glial cell types. To power this analysis genome-wide, we expanded prior MS susceptibility results[1] by incorporating three new genome-wide datasets: the UK Biobank (UKBB)[19], the Electronic Medical Records and Genomics 3 (eMERGE-III)[20] study, and the initial release of the All of US cohort (AoU)[21]. These datasets were previously harmonized[22] into a coherent dataset of 750,051 participants and substantially expanded our prior MS genome-wide association study (GWAS) efforts[1]. The inclusion of these cohorts also enabled a multi-ancestry analysis, allowing us to examine the relevance of susceptibility loci identified in European ancestry (EUR) populations to African American (AFR) and Admixed American (AMR) participants. To prioritize candidate causal genes, we integrated the extended EUR GWAS results with gene expression data from EUR participants using a colocalization analysis, leveraging single-nucleus RNA sequencing (snucRNA-seq) across six major cell types of the dorsolateral prefrontal cortex (DLPFC)[23] and 14 major cell types of peripheral blood mononuclear cells (PBMCs)[24]. We further examined the overlap of MS risk loci with genetic associations across inflammatory, neurodegenerative, psychiatric, and metabolic traits. Finally, we designed and tested a genome-wide polygenic score (GPS)[25] for MS that maximizes performance across diverse ancestries, conducted a phenome-wide association study (PheWAS) to identify diseases/traits associated with the GPS, and evaluated GPS associations with brain MRI phenotypes in individuals with MS (**Fig. 1**).

Results

New European ancestry GWAS meta-analysis for MS

We first harmonized the genetic and phenotypic data from the UKBB, eMERGE-III, and AoU datasets[22], defining MS cases as individuals with one of the following ICD 9 codes: 340, 323 and 341. We then conducted a GWAS meta-analysis using METAL[26] that included 5,063 MS cases and 596,340 controls of European ancestry (see **Supplementary Methods** and **Supplementary Table 1**). These results were combined with our prior meta-analysis[1] (**Supplementary Table 2**), increasing the sample size to 19,865 MS cases and 623,043 controls. Since this project focused on exploring non-immune functional consequences, we excluded the extended major histocompatibility complex (MHC) region (Chr6: 25,383,722-33,368,421bp in GRCh37) from our primary analysis.

A total of 5,041 non-MHC SNPs exceeded a threshold of $P < 5\times10^{-8}$ in the new meta-analysis (**Fig. 2**). All shared SNPs showed concordant directions of effect between the prior study[1] and the three new cohorts (**Supplementary Fig. 1**). Using linkage disequilibrium (LD)-based clumping methods, we identified 236 SNPs independently associated with MS susceptibility among the 5,041 significant SNPs. We then removed SNPs with $r^2 > 0.1$ within a ±500-kb window of the 200 previously reported non-MHC susceptibility variants[1]. A total of 38 SNPs were not in LD with the previously reported SNPs and were considered to tag novel susceptibility variants. Most of these lie in regions previously implicated in MS. To identify genomic regions newly associated with MS susceptibility, we assessed ±250 kb around each of these 38 SNPs and removed those located near known variants. This approach identified four loci not previously reported to harbor MS susceptibility variants (**Table 1**, **Fig. 2** and **Supplementary Table 3**). Therefore, most newly associated variants (n = 34) fall within loci already harboring MS susceptibility variants. We also removed two susceptibility SNPs from our previous study (rs6498163 and rs11256593) because of $r^2 > 0.1$ with other MS SNPs; in these cases, the SNP with the smaller P-value was kept. This resulted in a total of 198 known and 38 novel autosomal variants, or 236 independent MS susceptibility effects labeled by lead SNPs (**Supplementary Table 4**). Results from this updated meta-analysis were used for all subsequent analyses.

Previous studies reported that multiple MS loci harbored more than one statistically independent genome-wide significant association[1]. This pattern was also observed in our updated list of MS risk variants, where multiple independent associations were detected at several loci, including the *DDX6-CXCR5* locus (**Supplementary Fig. 2a**), which has also been implicated in other autoimmune diseases. For example, the allele rs12365699^G, located in the *DDX6-CXCR5* locus, also increases the risk of rheumatoid arthritis and lupus[27–29]. Similar patterns of multiple independent associations were observed at the *CD6*, *AHI1* and *PTGER4* loci (**Supplementary Fig. 2b-d**).

Multi-ancestry GWAS meta-analysis for MS

Although the number of AFR (614 cases and 62,044 controls) and AMR (352 cases and 44,133 controls) ancestry individuals identified across the three biobanks was large for an MS study, it remains modest for GWAS. Our first study of European ancestry participants required 1,000 MS cases to yield two loci reaching genome-wide significance[30]. Nonetheless, we completed separate GWAS in these populations to identify ancestry-specific loci. Surprisingly, one locus reached genome-wide significance among AFR participants (rs76911648), whose minor allele frequency (MAF) is lower in EUR (MAF = 0.013) than in AFR (MAF = 0.035). The nearest

gene, *SMARCA2*, has been shown to promote metastasis in pancreatic cancer by epigenetically activating *STAT3* signaling, and *STAT3* is a target gene in EUR MS participants (**Table 1**)[31,32]. We also identified two significant SNPs among AMR participants (rs59061674, rs113284638) (**Table 1** and **Supplementary Fig. 3**), where rs59061674 (**Supplementary Fig. 3e**) has a lower MAF in EUR (MAF = 0.010) than in AMR (MAF = 0.038), while rs113284638 showed a slightly higher MAF in EUR (MAF = 0.072) than in AMR (MAF = 0.063). Given the modest AMR and AFR sample sizes, these results should be viewed cautiously; none of these three SNPs have $P < 0.05$ in EUR participants, and none are in LD with the significant SNPs described above. Additional non-European datasets for replication are currently unavailable, and these findings will require validation in larger and more diverse cohorts.

To be thorough, we next performed a multi-ancestry meta-analysis including 20,831 MS cases (>40% increase in the number of MS cases used in the previous GWAS[1]) and 729,220 controls using two methods: a multi-ancestry meta-regression implemented in MR-MEGA[33] and a random effects model implemented in PLINK (v1.9)[34]. No additional loci became significant in this slightly larger meta-analysis. Among the 236 significant SNPs from the EUR meta-analysis, 184 were available in AFR, of which 18 showed nominal evidence of association ($P < 0.05$, 14 with concordant effect direction). Similarly, 189 of the 236 SNPs were available in AMR participants, with 11 showing nominal association ($P < 0.05$, 9 with concordant effect direction) (**Supplementary Table 5**). We also observed significant cross-ancestry concordance in effect sizes between EUR and AFR participants (Pearson's $r = 0.33$, $P = 3.36\times10^{-6}$), and between EUR and AMR participants (Pearson's $r = 0.42$, $P = 2.39\times10^{-9}$) (**Supplementary Fig. 3f,g** and **Supplementary Table 6**). Although larger studies in non-European populations remain essential, our results suggest that certain findings from European-ancestry meta-GWAS are also relevant to AFR and AMR populations, consistent with earlier studies[35,36].

Susceptibility alleles overlap between MS and other autoimmune diseases

Next, we assessed the extent to which MS susceptibility was shared with other diseases. We assembled a list of SNPs that reached genome-wide significance ($P < 5\times10^{-8}$) in at least one of 12 autoimmune diseases using publicly available GWAS summary statistics (see **Methods**)[37–44]. Adding these SNPs to those reaching genome-wide significance in our MS analysis yielded 32,901 SNPs for cross-disease analysis (**Fig. 3a**). A single risk allele (rs3184504^{T}), a nonsynonymous SNP in *SH2B3*, exhibited the highest pleiotropy with concordant risk associations across seven autoimmune diseases (**Supplementary Table 7**), including MS, psoriasis, lupus, type 1 diabetes (T1D), celiac disease, inflammatory bowel disease (IBD) and thyroiditis. We also identified 7,849 SNPs shared between at least two autoimmune diseases, with progressively fewer SNPs associated with three or more diseases, including five SNPs implicated in six diseases.

To further explore pleiotropy among MS variants, we evaluated our updated list of 236 MS susceptibility SNPs (**Supplementary Table 4**) across other GWAS, including MS severity[2], 12 autoimmune diseases[37–44], four neurodegenerative diseases[45–48], four psychiatric disorders/traits[49–52], and three metabolic traits[53,54] (**Supplementary Table 8**). The analysis was performed using two thresholds: nominal significance ($P < 0.05$) and a more conservative threshold ($P < 0.001$). Results were partitioned into three groups: (1) SNPs with the same direction of effect as MS, (2) SNPs with the opposite direction, and (3) SNPs not meeting

significance threshold (**Fig. 3b** and **Supplementary Fig. 4**). T1D and IBD showed the most overlap, with 87 and 73 of the MS SNPs reaching nominal significance (**Fig. 3b**). In both cases, there was a significant skew toward the same direction of effect, although approximately one-quarter of variants showed the opposite direction (T1D: binomial $P = 1.27\times10^{-5}$; IBD: binomial $P = 2.12\times10^{-3}$; **Supplementary Table 9**). This pattern held true for the other autoimmune diseases and for the higher threshold of significance (T1D: binomial $P = 0.01$; IBD: binomial $P = 0.01$; **Supplementary Table 9**), consistent with previous reports[18]. The extent of sharing also depends on the size and power of the non-MS GWAS, as some disease studies remain relatively underpowered, such as the MS severity GWAS, which returned only one locus[2].

Interestingly, we observed notable genetic sharing between MS and neuropsychiatric traits, more than with neurodegenerative diseases (**Fig. 3b**). Alzheimer's disease is intriguing, with over 65% of shared SNPs showing inverse effects relative to MS risk. However, this trend did not reach statistical significance (binomial $P = 0.07$). Metabolic traits also showed overlap with MS, but effect directions were not significantly skewed, with ~50% of shared variants showing inverse associations. Under the more stringent statistical significance cutoffs ($P < 0.001$), few MS SNPs were associated with other traits, but the pattern among the autoimmune diseases was the same (**Supplementary Fig. 9**).

Figure 3c presents these results in greater detail by focusing on MS risk variants that reached genome-wide significance in at least one of the 24 phenotypes examined (**Supplementary Table 8**). While many variants appear to affect pathways predisposing to autoimmune responses, the substantial number of inverse effects suggests disease-specific roles for some immune pathways. For example, at the *STAT3* locus, rs1026916 ($P < 10^{-28}$ in MS) (**Fig. 3c**) has a concordant direction with psoriasis but an opposite direction in IBD, ulcerative colitis (UC), and Crohn's disease (CD). Despite many MS SNPs displaying nominal sharing with other autoimmune diseases, 50/236 were MS-specific at the most comprehensive threshold ($P < 0.05$) across the 12 autoimmune diseases, and 27 were specific to MS across all phenotypes we tested.

Using the GWAS summary statistics above, we estimated genetic correlations between MS and the 24 other phenotypes using LD score regression (LDSC) with an imputed reference panel of 1,217,312 quality-controlled HapMap3 SNPs[55] (**Fig. 3d** and **Supplementary Table 10**). MS showed the strongest genetic correlations with UC ($r_g = 0.250$, $P = 5.47\times10^{-9}$), systemic lupus erythematosus (SLE) ($r_g = 0.221$, $P = 1.00\times10^{-4}$, IBD ($r_g = 0.194$, $P = 3.60\times10^{-6}$) and rheumatoid arthritis (RA) ($r_g = 0.150$, $P = 1.50\times10^{-3}$), consistent with earlier studies[56,57]. We also observed a significant positive correlation between MS and neuroticism ($r_g = 0.090$, $P = 2.00\times10^{-4}$). MS severity showed no significant correlations with the traits examined. Larger studies will be needed to assess shared genetic architecture between MS susceptibility and severity. Notably, IgA nephropathy, chronic obstructive pulmonary disease (COPD), thyroiditis, IBD, and CD had significant genetic correlations with the psychiatric disorders/traits tested (**Fig. 3d**).

Genome-wide polygenic score for MS

Polygenic risk scores summarize an individual's genetic susceptibility to MS. They may help stratify individuals at high risk of developing MS[58]. Given the importance of the MHC in MS risk, we constructed a GPS using genome-wide SNPs both with and without variants in the MHC region (**Table 2**, **Supplementary Fig. 5** and **Supplementary Tables 11** and **12**). The model was

developed using the combined IMSGC, UKBB, and AoU datasets, excluding eMERGE-III. GPS models were then independently trained and tested in the eMERGE-III dataset. A summary of the analytical workflow is shown in **Figure 4a** (see **Methods**). The GPS for MS was tested with adjustment for age, sex, genotyping batch, and genetic ancestry. In the independent European-ancestry testing cohort, the GPS including the MHC region was strongly associated with MS risk (OR per standard deviation = 1.79, 95% CI: 1.60-1.99, $P = 3.27\times10^{-25}$; **Table 2**). Individuals in the top 1% vs. the remaining 99% of MS-GPS had more than a seven-fold increased risk (95% CI: 4.22-11.90, $P = 1.31\times10^{-13}$). The GPS without MHC behaved in a similar fashion (**Supplementary Table 12**). We also evaluated the GPS in the smaller AMR and AFR cohorts from eMERGE-III. Although effect sizes were reduced, the GPS remained significantly associated with MS in both cohorts (AMR: OR = 1.28, 95% CI: 1.08-1.52, $P = 4.61\times10^{-3}$; AFR: OR = 1.74, 95% CI: 1.32-2.31, $P = 1.14\times10^{-4}$; **Table 2**). These results suggest that, although further optimization in non-European populations is needed, the current GPS is already informative across the three tested populations.

Next, we performed a PheWAS in eMERGE-III participants not included in the GPS design. This approach offers a complementary strategy to assess shared genetic architecture across clinical traits. In the well-powered European ancestry analysis, the GPS association with MS was strongly replicated (OR = 2.05, 95% CI: 1.79-2.34, $P = 1.78\times10^{-26}$; **Fig. 4b** and **Supplementary Table 13**). We also found a GPS association with "other inflammatory demyelinating diseases of central nervous system" (OR = 1.79, 95% CI: 1.48-2.16, $P = 2.64\times10^{-9}$; **Fig. 4b** and **Supplementary Table 13**). This poorly defined diagnostic group may harbor certain individuals with MS and contains conditions that share symptomatology with MS but have different immune mechanisms. Thus, there may be some overlap in genetic architecture with these less common entities. An association with "functional disorders of the bladder" (OR = 1.21, 95% CI: 1.11-1.32, $P = 1.53\times10^{-5}$; **Fig. 4b** and **Supplementary Table 13**) likely reflected bladder dysfunction as a common symptom of MS. No other diagnostic categories were significantly associated with the MS GPS, suggesting that our GPS is relatively specific to MS.

In the smaller AFR dataset, the association of the GPS with MS was also phenome-wide significant (OR = 1.65, 95% CI: 1.33-2.05, $P = 6.29\times10^{-6}$; **Fig. 4c** and **Supplementary Table 14**), consistent with the earlier analyses. In the AMR cohort with fewer cases, the association with MS was nominally significant (OR = 2.09, 95% CI: 1.39-3.13, $P = 3.59\times10^{-4}$; **Fig. 4d** and **Supplementary Table 15**), likely due to limited power. We conclude that the GPS is associated with MS across different ancestral populations, but its predictive performance remains lower in non-European populations.

Finally, we evaluated the GPS in a subset of MS patients with magnetic resonance imaging (MRI) measurements (gray matter, white matter, and cerebrospinal fluid), Expanded Disability Status Scale (EDSS), and genotype data from the Comprehensive Longitudinal Investigation of Multiple Sclerosis at the Brigham and Women's Hospital (CLIMB) study[59] (see **Methods**). We used linear regression models to test the associations between the GPS and brain tissue compartments. We observed an association between the GPS calculated without the MHC region and lower white matter volume in MS patients ($P = 3.88\times10^{-4}$) (**Fig. 4e**). No significant associations were found between the MS GPS and other MRI measurements.

Functional characterization of MS variants using cell-type specific brain and blood expression quantitative trait loci (eQTLs)

We used our atlas of CNS cell type-specific eQTLs[23] and a publicly available PBMC scRNA-seq resource[24] to map the effects of MS variants in peripheral immune and brain cells. We identified MS susceptibility variants near eQTLs in blood and brain cell types and assessed colocalization of effects using Coloc (v5.1.0) (see **Methods**). As expected, several colocalized effects (PP.H4 > 0.8) were observed in blood-derived cells, particularly naïve T cells (**Fig. 5a** and **Supplementary Table 16**)[60,61]. Most effects were shared among several cell types (**Supplementary Fig. 6**), but some – such as *NR1D1* and *MMEL1* – appeared specific to naïve T cells. Further, we detected colocalization with microglial eQTLs, consistent with earlier observations[1,6].

However, the most interesting findings involve neurectoderm-derived glial and neuronal cells. In our reference, excitatory neurons are the most abundant cortical cell type and have the largest transcriptome, resulting in the greatest number of detectable eQTLs[23]. Despite this, inhibitory neurons harbored the most eQTLs that colocalize with MS susceptibility variants among CNS cell types (n = 15) (**Fig. 5a,b**), exceeding microglia (n = 6). Seven of these effects were unique to inhibitory neurons. We also identified five variants with functional consequences only in excitatory neurons, suggesting that neuronal cells may contribute to early events in MS pathogenesis. **Figure 6a** and **Supplementary Figure 7** illustrate this: *STAT3* encodes a cytokine-related factor involved in amplifying immune responses[62]. Our comparison of blood and brain cells indicates that this MS variant may influence immune responses only in inhibitory neurons, potentially linking peripheral immune dysregulation to CNS targeting. Further work is needed to understand how this effect interacts with other neuronal-specific cytokine-related signals such as *IL7*. Since our colocalization analyses suggested pleiotropic effects from some variants, we also performed summary-based Mendelian randomization (SMR) to assess potential causal relationships between gene expression and MS susceptibility. These analyses supported a causal relationship for *STAT3* expression by inhibitory neurons in MS risk (**Supplementary Fig. 8a** and **Supplementary Tables 17** and **18**).

While neurons harbor the most functional consequences of MS variants among CNS cells, glial cell types also showed several colocalizations, including astrocyte-specific effects (*KCTD13* and *RRAS2*) and oligodendrocyte-specific effects (*PHGDH* and *SYNGR1*). A previous report implicated an MS variant near *NFKB1* in altered astrocyte immune responses; however, this SNP did not alter gene expression in our brain dataset[63].

To investigate biological relevance of the colocalized genes, we performed Gene Ontology (GO) enrichment analysis using the Database for Annotation, Visualization, and Integrated Discovery (DAVID)[64], focusing on KEGG pathways and GO categories (Biological Process, Molecular Function, and Cellular Component). Due to relatively small number of colocalized genes identified per cell type, no pathways reached statistical significance after multiple testing correction (FDR).

Replication of colocalized eQTLs and epigenomic assessment

We accessed additional single-nucleus DLPFC datasets and found that the *STAT3* eQTL effect in inhibitory neurons replicated in two independent datasets[65] (**Fig. 6c-e** and **Supplementary Table**

19). Additional examples of replication include rs4896153 influencing *AHI1* expression in microglia and rs6032662 associated with *SLC12A5* expression in both excitatory and inhibitory neurons. Overall, 33 out of the 46 colocalized brain cis-eQTLs were replicated (**Supplementary Table 19**).

Review of reference epigenomic profiles[66] showed that many prioritized variants (**Fig. 5** and **Supplementary Table 16**) were not located in known open chromatin regions of the cell types implicated by the eQTL analyses. This may reflect limited statistical power, as only 92 samples were included in the snucATAC-seq dataset used here, and snucATAC-seq data are inherently sparse. One notable exception was rs3923387 near *PLEC*, which influences MS susceptibility, chromatin accessibility in microglia (GRCh37 chr8:145034681-145035181; PP.H4 = 0.84), and *PLEC* expression in the same cell type (colocalization of the ATAC QTL and eQTL PP.H4 = 0.93) (**Supplementary Fig. 8c,d**). This illustrates the next phase of consortium's work, i.e., generating improved, cell-resolved, multi-omic datasets to map how MS susceptibility variants propagate their molecular effects.

Finally, we confirmed expression of STAT3 protein in inhibitory neurons using immunofluorescence in DLPFC tissue from an MS donor obtained from the National MS Brain Bank. We found that 2.8% of $GAD1^{+}GAD2^{+}$ inhibitory neurons demonstrated elevated STAT3 protein expression (>2SD) (**Fig. 6f** and **Supplementary Fig. 9**). Morphological comparisons revealed no significant differences between $STAT3^{high}$ and other inhibitory neurons. Literature review showed that our MS *STAT3* risk allele, $rs102691^{A}$, is in LD ($r^2 = 0.70$) with $rs1053004^{A}$ associated with reduced cognitive performance and lower neurite density index in the white matter of 33,224 UKBB participants with MRI data[67]. These results suggest that the *STAT3* variant may reduce white matter integrity and cognitive performance in the general population, potentially increasing CNS vulnerability to inflammatory damage. To further explore this question, we examined rs1026916 in 3,480 individuals with MS of European ancestry with longitudinal measures of neuronal integrity (serum neurofilament light chain, sNfL). The risk allele was associated with faster sNfL increases over an average follow-up 1.4 years (SD: 0.68 years) and an average of 2.7 (SD: 0.8) samples per person, suggesting that individuals carrying the risk allele have a greater vulnerability to damage from inflammatory demyelination. In a linear mixed model, increasing A-allele dosage was associated with higher sNfL levels ($\beta = 0.05$; 95% CI: 0.004-0.10; $P = 0.03$). Comparing AA versus GG genotypes, the annual change in sNfL differed by 0.11 (95% CI: 0.01-0.21; $P = 0.03$; **Fig. 6b**).

Discussion

In this updated MS GWAS of 19,865 MS cases with genome-wide genotype data, we identified 38 novel MS risk variants and four genomic regions not previously implicated in MS susceptibility. Combined with SNPs from previous studies[1], our consortium has now reported 236 independent non-MHC autosomal MS risk variants in individuals of European ancestry. We estimated overall heritability of MS susceptibility at 23.2%, similar to our previous estimate of 19.2%[1].

We also conducted a multi-ancestry MS GWAS, including AFR and AMR ancestry participants. The three AFR- or AMR-specific significant variants require further replication, but we noticed

substantial concordance in effect sizes across ancestries for the 236 lead MS risk variants (**Table 2**, **Supplementary Fig. 3f,g** and **Supplementary Table 12**). Our GPS provides a new tool to evaluate genetic predisposition to MS in other datasets. While further optimization is required for non-European populations, the current GPS shows predictive capacity in AFR and AMR individuals, consistent with earlier reports[68,69]. Larger studies in non-European populations are needed to ensure the clinical utility of GPS across ancestries. Nonetheless, this cross-ancestry predictive ability likely reflects a shared genetic architecture underlying MS susceptibility. Previous studies have shown that effect sizes for many complex traits are largely conserved across ancestries[70], which may explain why a GPS derived primarily from European-ancestry data retains predictive utility in diverse cohorts. In addition, European admixture in AFR and AMR participants may contribute to GPS performance.

From the 236 genome-wide independent non-MHC MS risk variants, we prioritized 76 candidate causal genes by integrating scRNA-seq and snucRNA-seq data from blood and brain tissues. The functional consequences of the remaining variants remain unresolved, underscoring the complexity of regulatory mechanisms underlying MS susceptibility. Limited overlap between MS risk variants and eQTLs is consistent with patterns observed in other autoimmune diseases. For example, only 21.0% of RA susceptibility loci colocalize with an eQTL[71], and across autoimmune traits, an average of 38.1% of loci show regulatory QTL colocalization[72]. This gap highlights the need for larger and more diverse eQTL datasets with improved cellular resolution. Emerging technologies such as long-read sequencing and single-cell multi-omics may further clarify the regulatory consequences of MS variants on splicing, proteomic and epigenomic features[73].

Using our updated MS GWAS, we functionally classified susceptibility variants. One approach examined overlap with other GWAS to identify variants influencing broader autoimmune susceptibility. This is consistent with epidemiological studies reporting increased prevalence of autoimmune diseases such as T1D, thyroid disease, and IBD among individuals with MS[74,75] and within affected families[76]. The rs3184504 variant in the *SH2B3* locus illustrates this pattern: its risk allele is associated with seven autoimmune diseases. *SH2B3* encodes an adaptor protein that regulates immune signaling, and dysfunctional variants may promote excessive immune responses, which can lead to autoimmunity[77]. This locus also shows extensive pleiotropy, being associated with 79 different GWAS traits[78]. Overall, these findings support shared autoimmune mechanisms[56] contributing to the onset of MS.

At a broad threshold, 186 of the 236 variants showed evidence of association with another autoimmune disease. The remaining variants may contribute to mechanisms directing autoimmune responses toward the CNS rather than other tissues. Prior work demonstrated that many MS susceptibility variants act in peripheral immune cells, particularly $CD4^+$ T cells, as well as other bone marrow-derived cells and microglia[1,2]. Evidence for effects in non-microglial CNS cells has been limited.

Our colocalization analysis confirmed that naïve $CD4^+$ and $CD8^+$ T cells harbor the greatest number of shared MS susceptibility and gene expression signals. Surprisingly, the next most enriched cell type was inhibitory neurons, with several signals unique to inhibitory neurons (compared to other cortical and bone marrow-derived immune cells). For example, *STAT3* and

IL7 showed colocalization of susceptibility and expression effects only in inhibitory neurons (**Fig. 5**). These cytokine pathway genes amplify immune responses, with *IL7*-driven signaling partly mediated through *STAT3*[79]. The *STAT3* variant is pleiotropic, being associated with cognitive performance and neurite density index measures[67] and with faster elevation in sNfL levels in individuals with MS, suggesting accelerated neuronal injury after disease onset. Together, these findings suggest that functional consequences of the *STAT3* variant (and potentially other MS variants) occur in the CNS of the general population, potentially influencing CNS development and resilience to inflammatory injury. This may represent a second factor – after a propensity for autoimmunity – contributing to MS onset, particularly during adolescence when brain development is ongoing. STAT3 has established roles in neuronal differentiation, mitochondrial metabolism, axonal structure, plasticity, and regeneration[80–83]. It also regulates neural progenitor differentiation toward astrogliogenesis[84,85] and contributes to axonal regeneration after injury[81,86]. Prior studies further showed that astrocyte-neuron interactions mediated by the JAK-STAT pathway in individuals with benign MS can mitigate inflammatory neuroaxonal damage[11]. Experimental inhibition of STAT3 phosphorylation can prevent neuronal loss and persistent glial activation[87]. Despite these insights, the specific role of STAT3 in inhibitory neurons in MS remains unclear. STAT3 may act both as a vulnerability factor and an early modulator of neuroinflammatory responses, representing a third factor contributing to MS onset: altered neuroglial responses to inflammatory signals from infiltrating autoreactive immune cells.

We note that, similar to *STAT3*, *ZHX3* is the target of an MS variant in excitatory neurons (**Fig. 5b**). ZHX3 belongs to a family of transcriptional repressors involved in neural progenitor maintenance, hematopoietic cell development, and mesenchymal stem cell differentiation[88,89]. Dysfunction of *ZHX* family members has been linked to neurological disease[88]. Thus, MS onset may partly reflect altered CNS development, creating a parenchyma more susceptible to autoimmune processes arising in the peripheral immune system.

Overall, pathways predisposing to autoimmunity may interact with vulnerability of neuroglial structures to inflammation and altered immune responses in neurons and glial cells. These responses may lead to recurrent inflammatory demyelination and a progressive neurodegenerative process that remains poorly understood. The predilection of inhibitory neurons as a target of MS risk variants is intriguing, given that post-mortem studies showing reduced inhibitory neurons, and a shift in excitation-inhibition balance toward hyperexcitability and excitotoxicity across MS stages[90–92]. Our observations generate hypotheses to explore downstream molecular and functional effects of these variants in the implicated cell types. The role of the adaptive immune system in MS is well established, with $CD8^+$ T cells abundant in MS white matter[93]. $CD4^+$ T cells likely play an equally important role given the convergence of genetic susceptibility effects in this cell type[94]. Our results suggest that interactions between T cells and neurons or glia may contribute to MS onset through both lymphocytes populations, consistent with extensive evidence of immune responses in neuroglial cells as well as shifts of neuronal excitation-inhibition balance, possibly linked to a cognitive phenotype[95,96]. More broadly, tissue-resident cells may play a similar role in other inflammatory diseases.

A notable methodological point relates to our large excess of control participants in our MS GWAS meta-analysis: it improves effect size estimation, enables detection of novel risk loci, and

strengthens downstream analyses such as polygenic score development, fine-mapping, and genetic correlation analysis. While important for our study, its utility may vary by disease context. For more prevalent autoimmune disorders, increasing control numbers beyond a certain threshold may offer diminishing returns. However, for less common autoimmune diseases with fewer cases, maximizing well-matched controls can substantially improve statistical power and the reliability of association signals. We therefore recommend that future studies carefully consider disease-specific factors in their study design.

In summary, our results advance understanding of the biological etiology of MS, refocusing efforts on disease onset to include molecular pathways in the developing brain as well as neuroglial interactions with autoreactive immune cells as a propensity for autoimmunity is realized in the periphery. While many loci act in immune cells, our findings emphasize an underappreciated neuronal contribution to MS susceptibility. These insights may inform future strategies for MS prevention and treatment targeting CNS pathways.

International Multiple Sclerosis Genetics Consortium

Adil Harroud[10], Alastair Compston[11], Stephen J. Sawcer[11], Allan G. Kermode[12], An Goris[13], Bernhard Hemmer[14], Bénédicte Dubois[15], Bruce Taylor[16], Dana Horakova[17], David A. Hafler[18], Efthimios Dardiotis[19], Farren B. S. Briggs[20], Federica Esposito[21], Filippo Martinelli-Boneschi[22], Georgios Hadjigeorgiou[23], Grant P. Parnell[24], Hanne F. Harbo[25], Helle B. Søndergaard[26], Ingrid Kockum[27], Tomas Olsson[27], Pernilla Stridh[27], Lars Alfredsson[27], Jacob L. McCauley[28], Margaret A. Pericak-Vance[28], Janna Saarela[29], Jeannette Lechner-Scott[30], Joost Smolders[31], Jorge R. Oksenberg[32], Roland G. Henry[32], Sergio E. Baranzini[32], Stepher L. Hauser[32], Lisa F. Barcellos[33], Massimo Filippi[34], Manuel Comabella[35], Michael Khalil[36], Neil P. Robertson[37], Noriko Isobe[38], Pierre-Antoine Gourraud[39], Nicolas Vince[39], Roland Martin[40], Sandra D'Alfonso[41], Seema Kalra[42], Teresa Fazia[43] & Voon Wee Yong[44]

[10]The Neuro (Montreal Neurological Institute-Hospital), Montréal, QC, Canada & Department of Neurology and Neurosurgery, McGill University, Montréal, QC, Canada & Department of Human Genetics, McGill University, Montréal, Québec, Canada.
[11]Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK.
[12]Perron Institute for Neurological and Translational Science, University of Western Australia, Perth, Western Australia, Australia & Personalised Medicine Centre, Health Futures Institute, Murdoch University, Perth, Western Australia, Australia.
[13]Department of Neurosciences, Leuven Brain Institute, KU Leuven, Leuven, Belgium.
[14]Department of Neurology, School of Medicine, Technical University of Munich, Munich, Germany & Munich Cluster for Systems Neurology (SyNergy), Munich, Germany.
[15]Department of Neurology, University Hospitals Leuven, Leuven, Belgium & Department of Neurosciences, Leuven Brain Institute, KU Leuven, Leuven, Belgium.
[16]Menzies Institute for Medical Research, University of Tasmania, Hobart, Tasmania, Australia.
[17]Department of Neurology and Center of Clinical Neuroscience, First Faculty of Medicine, Charles University and General University, Prague, Czech Republic.
[18]Departments of Neurology and Immunobiology, Yale School of Medicine, New Haven, CT, USA.
[19]Department of Neurology, University General Hospital of Larissa, Faculty of Medicine, School of Health Sciences, University of Thessaly, Larissa, Greece.

[20]Department of Public Health Sciences, University of Miami Miller School of Medicine, Miami, FL, USA.
[21]Neurology Unit and Laboratory of Human Genetics of Neurological Disorders, IRCCS Ospedale San Raffaele Scientific Institute, Milan, Italy.
[22]Clinical Neurology, Department of Health Science, CRC "Aldo Ravelli" for Experimental Brain Therapeutics, University of Milan, Italy & Hospital San Paolo, ASST Santi Paolo e Carlo, Milan, Italy.
[23]Medical School, University of Cyprus, Nicosia, Cyprus.
[24]Centre for Immunology and Allergy Research, The Westmead Institute for Medical Research, Westmead, New South Wales, Australia & School of Medical Sciences, Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia.
[25]Department of Neurology, Oslo University Hospital, Oslo, Norway & Institute of Clinical Medicine, University of Oslo, Oslo, Norway.
[26]Danish Multiple Sclerosis Center, Department of Neurology, Copenhagen University Hospital–Rigshospitalet, Glostrup, Denmark.
[27]Department of Clinical Neuroscience, Karolinska Institutet, Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden.
[28]John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA & The Dr John T. Macdonald Foundation Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, FL, USA.
[29]Centre for Molecular Medicine Norway, University of Oslo, Oslo, Norway & Institute for Molecular Medicine Finland, Helsinki Institute for Life Sciences, University of Helsinki, Helsinki, Finland.
[30]Department of Neurology, John Hunter Hospital, Hunter New England Health District, Newcastle, New South Wales, Australia & Hunter Medical Research Institute, University of Newcastle, Newcastle, New South Wales, Australia.
[31]Departments of Neurology and Immunology, MS Center ErasMS, Erasmus University Medical Center, Rotterdam, The Netherlands & Neuroimmunology Research Group, Netherlands Institute for Neuroscience, Amsterdam, The Netherlands.
[32]UCSF Weill Institute for Neurosciences, Department of Neurology, University of California, San Francisco, CA, USA.
[33]Genetic Epidemiology and Genomics Laboratory, Division of Epidemiology, School of Public Health, University of California, Berkeley, CA, USA.
[34]Neuroimaging Research Unit, Division of Neuroscience, IRCCS San Raffaele Scientific Institute, Neurology Unit, IRCCS San Raffaele Scientific Institute, Vita-Salute San Raffaele University, Neurorehabilitation Unit, IRCCS San Raffaele Scientific Institute, Neurophysiology Service, IRCCS San Raffaele Scientific Institute, Milan, Italy.
[35]Servei de Neurologia, Centre d'Esclerosi Múltiple de Catalunya (Cemcat), Vall d'Hebron Institut de Recerca, Vall d'Hebron Hospital Universitari, Barcelona, Spain.
[36]Department of Neurology, Medical University of Graz, Graz, Austria.
[37]Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, UK.
[38]Department of Neurology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan.
[39]Nantes Université, CHU Nantes, Centrale Nantes, Inserm, Center for Research in Transplantation and Translational Immunology, Nantes, France.
[40]Institute of Experimental Immunology, University of Zurich, Zurich, Switzerland & Cellerys AG, Wagistrasse 21, Schlieren, Switzerland & Neuroimmunology and MS Research, Neurology Clinic, University Hospital Zurich, University of Zurich, Zurich, Switzerland & Therapeutic Immune Design Unit, Department of Clinical Neuroscience, Karolinska Institute, Center for Molecular Medicine, Stockholm, Sweden.
[41]Department of Health Sciences and Center on Auto-immune and Allergic Diseases (CAAD), University of Eastern Piedmont, Novara, Italy.
[42]Neurology Department, University Hospital North Midlands NHS Trust, Stoke-on-Trent, UK.

[43]Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy.
[44]Department of Clinical Neurosciences and the Hotchkiss Brain Institute, University of Calgary, Calgary, Alberta, Canada.

Acknowledgements

We thank the participants who donated their time, life experiences and DNA to this research and the clinical and scientific teams that worked with them.

Funding

This work was funded by NIH grants to P.L.D.J. (U01 AG061356 and the Foundation for the National Institutes of Health's Accelerating Medicines Program), and to A.K. (K25DK128563), as well as by a National Multiple Sclerosis Society (NMSS) grant to L.Z. (BI-2408-43968) and the German Research Foundation (DFG) grant SFB/CRC 1292 (project number 318346496) (F.Z.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author Contributions

P.L.D.J. conceived the study and supervised the research. L.Z. and A.K. performed computational analyses. K.C.F., R.L., E.A.T., K.Y., T.C., H.L.W., Q.L.G., S.D., P.A.C., M.F., F.Z., G.W., K.K., and P.L.D.J. provided and analyzed the data. M.T. and T.L. carried out the immunofluorescence staining analysis. P.L.D.J., L.Z. and A.K. acquired the fundings. L.Z., A.K. and P.L.D.J. wrote the original manuscript draft. L.Z., A.K., K.C.F., J.C., M.F., M.T., F.Z., G.W., K.K., P.L.D.J., and all consortium authors (A.H., A.C., A.G.K., A.G., B.H., B.T., B.D., D.H., D.A.H., E.D., F.B.S.B., F.E., F.M.-B., G.H., G.P.P., H.F.H., H.B.S., I.K., J.L.M., J. Smolders, J.L.-S., J. Saarela, J.R.O., L.A., L.F.B., M.C., M.A.P.-V., M.F., M.K., N.P.R., N.V., N.I., P.S., P.-A.G., R.G.H., R.M., S.D'A., S.K., S.E.B., S.J.S., S.L.H., T.F., T.O., V.W.Y.) reviewed and edited the manuscript draft.

Competing Interests

E.A.T. and K.Y. are employees and stockholders of Biogen Inc. K.F. has served as a consultant for SetPoint Medical. P.-A.G. is the founder of Methodomics and its spin-off Big Data Santé ("Octopize"). He consults for pharmaceutical companies and start-ups, including AstraZeneca, Amgen, Biogen, Boston Scientific, Cemka, Cepton, Cook, Docaposte/Heva, Edimark, Ellipses, Elsevier, Grunenthal, Hemovis, Janssen, IAGE, Lek, Merck, Mérieux, Novartis, Octopize, Sanofi-Genzyme, Lifen, TuneInsight and Aspire UAE. He serves as a volunteer board member at AXA (not-for-profit mutual insurance company, 2021). He has no prescription activity with either drugs or devices. He receives no wages from these activities. T.O. has received lecture and advisory board honoraria from Merck, Novartis and Sanofi, and unrestricted research grants from Merck, Novartis, Sanofi and Sandoz, and academic grants from the Swedish Research Council, the Swedish Brain Council, Knut and Alice Wallenberg foundation and Margaretha af Ugglas Foundation. H.L.W. has served on advisory boards for Sanofi and Tiziana Life Sciences and holds stock in vTv Therapeutics and Tiziana Life Sciences, and research support from Cure Alzheimer's Fund, National Institutes of Health and National Multiple Sclerosis Society. I.K. has received lecture honoraria from Merck and research funding from Pfizer. J. Smolders has

received consultancy and/or lecture fees from Biogen, Merck, Novartis, Roche and Sanofi; his institution has received research funding from Biogen, Merck, Roche and Siemens Healthineers, all outside the submitted work. A.H. has received consultancy fees from Amgen, Biogen, Novartis, Pfizer, and Roche; compensation for serving on a scientific advisory board for Amgen, Alexion, and Roche; and research funding from Amgen and Roche, all unrelated to the present work. R.M. has received unrestricted grants from Biogen, Novartis, Roche, Third Rock and honoraries for advisory roles and lectures from Roche, Novartis, Biogen, Genzyme, Neuway, CellProtect, Third Rock, Teva, Swiss Rockets. He is a patent holder and co-holder on patents of: daclizumab in MS (held by NIH), JCV VP1 for vaccination against PML; JCV-specific neutralizing antibodies to treat PML; antigen-specific tolerization with peptide-coupled cells, novel autoantigens in MS, and designer neoantigens for tumor vaccination (all held by University of Zurich). He is a co-founder of Abata Therapeutics, Watertown, MA, USA, and co-founder and employee of Cellerys AG, Schlieren, Switzerland. F.Z. has received research grants and/or consultation funds from Max Planck Society (MPG), Ministry of Education and Research (BMBF), Bristol-Meyers-Squibb, Celgene, German Research Foundation (DFG), Hexal, Horizon, Janssen, Juvisé, Merck Serono, Novartis, Progressive MS Alliance (PMSA), Roche, Sanofi Genzyme, Sandoz. N.P.R. has received grants for research held by his institution from Novartis, Genzyme, and Sanofi, and support for attendance and or honoraria from Merck, Novartis, Genzyme, Sanofi, Bayer. F.E. received compensation for consulting services from Alexion, Almirall, Biogen, Merck, Novartis, Roche, Sanofi, speaking activities from Bayer, Biogen, Celgene, Chiesi Italia SpA, Eli Lilly, Genzyme, Janssen, Merck-Serono, Neopharmed Gentili, Novartis, Novo Nordisk, Roche, Sanofi, Takeda, and TEVA, participation in Advisory Boards for Alexion, Biogen, Bristol-Myers Squibb, Merck, Novartis, Roche, Sanofi, Sanofi-Aventis, Sanofi-Genzyme, Takeda, scientific direction of educational events for Biogen, Merck, Roche, Celgene, Bristol-Myers Squibb, Lilly, Novartis, Sanofi-Genzyme, he receives research support from Biogen Idec, Merck-Serono, Novartis, Roche, the Italian Ministry of Health, the Italian Ministry of University and Research, and Fondazione Italiana Sclerosi Multipla. M.F. received compensation for consulting services from Alexion, Almirall, Biogen, Merck, Novartis, Roche, Sanofi, speaking activities from Bayer, Biogen, Celgene, Chiesi Italia SpA, Eli Lilly, Genzyme, Janssen, Merck-Serono, Neopharmed Gentili, Novartis, Novo Nordisk, Roche, Sanofi, Takeda, and TEVA, participation in Advisory Boards for Alexion, Biogen, Bristol-Myers Squibb, Merck, Novartis, Roche, Sanofi, Sanofi-Aventis, Sanofi-Genzyme, Takeda, scientific direction of educational events for Biogen, Merck, Roche, Celgene, Bristol-Myers Squibb, Lilly, Novartis, Sanofi-Genzyme, he receives research support from Biogen Idec, Merck-Serono, Novartis, Roche, the Italian Ministry of Health, the Italian Ministry of University and Research, and Fondazione Italiana Sclerosi Multipla. J.L.-S. received travel compensation from Biogen, Merck, Novartis; has been involved in clinical trials with Biogen, Novartis, Roche; her institution has received honoraria for talks and advisory board service from Biogen, Merck, Novartis, Roche. She is on the Board of Directors for MS Plus and MSA. S.L.H. currently serves on the scientific advisory boards of Accure, Alector, Hinge Bio; and as an advisor for Neurona, he has also received nonfinancial support (travel reimbursement and writing support for anti-CD20-therapy-related meetings and presentations) from F. Hoffmann-La Roche and Novartis AG; and has received no personal compensation from any entity that sells or tests therapeutics for MS. D.A.H. has received research funding from Bristol-Myers Squibb, Novartis, Sanofi, and Genentech. He has been a consultant for Bayer Pharmaceuticals, Repertoire Inc, Bristol Myers Squibb, Compass Therapeutics, EMD Serono, Genentech, Juno therapeutics, Novartis

Pharmaceuticals, Proclara Biosciences, Sage Therapeutics, and Sanofi Genzyme. F.M.-B. received compensation for consulting services and/or speaking activities from Teva Pharmaceutical Industries, Sanofi Genzyme, Merck-Serono, Biogen Idec, Roche, Medday, Excemed, and received research support from Merck, Teva Pharmaceutical Industries, Italian Ministry of Health, Fondazione Italiana Sclerosi Multipla and Fondazione Cariplo. L.A. has received lecture honoraria from Merck, Novartis and Teva. I.K. has received academic research grants from Horizon 2020 (MultipleMS 733161), Horizon Europe (WISDOM 101137154), Swedish Research Council and Swedish Brain Foundation during the conduct of the study. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. H.S. has received grants from NIH/NINDS (R35NS111644), Valhalla Foundation, and C.D. Spangler Foundation. L.A. has received grants from Swedish Research Council, grants from Swedish Research Council for Health Working Life and Welfare and grants from Swedish Brain Foundation. P.S. has received academic grants from Deutsche Forschungsgemeinschaft (DFG, no. 523862973), the Swedish Research Council, the NEURO, and Margaretha af Ugglas Foundation. All other authors declare no financial interests or potential conflicts of interest.

Tables

Table 1 | GWAS meta-analysis uncovers 38 additional MS susceptibility variants in EUR, one in AFR and two in AMR.

Locus	Anc.	CHR	BP	Lead variant	*P*-value	OR	s.e.	RA/OA	RAF	Genes
1	EUR	1	2701575	rs375915427	3.15×10^{-8}	1.079	0.014	T/C	0.076	*MMEL1*[a], *TTC34*[b]
2	EUR	1	85748811	rs529392609	2.63×10^{-10}	2.063	0.115	G/A	0.999	*BCL10*[c]
3	EUR	1	85764886	rs11161589	6.12×10^{-12}	1.042	0.006	G/A	0.403	*CYR61*[b], *BCL10*[b]
4	EUR	1	92956978	rs113561235	3.60×10^{-8}	1.101	0.018	T/C	0.029	*GFI1*[c]
5	EUR	1	92973242	rs79285232	5.85×10^{-11}	1.202	0.028	C/T	0.017	*EVI5*[c]
6	EUR	1	93088923	rs72724541	1.26×10^{-11}	1.136	0.019	A/G	0.025	*GFI1*[b]
7	EUR	1	93291944	rs12042488	1.09×10^{-11}	1.063	0.009	A/T	0.800	*RPL5*[c]
8	EUR	1	101289496	rs142860878	1.77×10^{-8}	1.198	0.032	G/C	0.013	*AC93157.1*[a]
9	EUR	1	101307053	rs12047318	4.93×10^{-10}	1.070	0.011	G/A	0.922	*AC93157.1*[a], *EXTL2*[b]
10	EUR	1	101544143	rs147885102	3.69×10^{-10}	1.052	0.008	T/A	0.718	*AC93157.1*[a]
11	EUR	1	157660829	rs77191363	7.86×10^{-9}	1.078	0.013	C/G	0.946	*FCRL3*[a]
12	EUR	2	30478386	rs4952115	4.31×10^{-9}	1.048	0.008	G/T	0.836	*LBH*[a,b]
13	EUR	2	61066666	rs1432295	2.74×10^{-8}	1.033	0.006	G/A	0.432	*REL*[b]
14	EUR	3	101661456	rs74482986	2.46×10^{-8}	1.065	0.011	C/A	0.928	*NXPE3*[b]
15	EUR	3	121770539	rs2255214	6.19×10^{-16}	1.049	0.006	G/T	0.495	*CD86*[c]
16	EUR	3	159702290	rs9858816	2.64×10^{-8}	1.034	0.006	C/T	0.379	*IL12A*[c]
17	EUR	5	40393852	rs1992662	1.67×10^{-17}	1.054	0.006	A/G	0.651	*PTGER4*[c]
18	EUR	5	118815815	rs28762138	1.46×10^{-8}	1.244	0.039	G/T	0.009	*HSD17B4*[c]
19	EUR	5	158944266	rs7727104	8.17×10^{-9}	1.039	0.007	A/G	0.737	*C1QTNF2*[b]
20	EUR	6	135749682	rs13218824	4.01×10^{-8}	1.079	0.014	C/T	0.047	*AHI1*[a], *LINC00271*[a]
21	EUR	6	135904197	rs76892387	1.44×10^{-8}	1.085	0.014	G/A	0.044	*AHI1*[a], *LINC00271*[a]
22	EUR	7	56091706	rs6975311*	5.30×10^{-9}	1.039	0.007	G/A	0.727	*PSPH*[c]
23	EUR	9	4981602	rs10758669*	2.20×10^{-8}	1.035	0.006	C/A	0.353	*JAK2*[c]
24	EUR	10	64384640	rs77051803	3.30×10^{-8}	1.053	0.009	A/G	0.109	*ADO*[c], *ZNF365*[c]
25	EUR	11	321235	rs56232455	1.78×10^{-8}	1.042	0.007	A/G	0.443	*RP11-326C3.15*[a], *IFITM3*[a,b]
26	EUR	11	60783062	rs75064517	6.85×10^{-9}	1.121	0.020	G/A	0.035	*CD6*[a]
27	EUR	11	60827933	rs11230581	5.55×10^{-15}	1.048	0.006	T/C	0.582	*CD6*[a]
28	EUR	11	72450091	rs77267834*	2.71×10^{-12}	1.103	0.014	A/T	0.046	*ARAP1*[b], *ATG16L2*[b]
29	EUR	14	88407917	rs12432149	4.08×10^{-12}	1.041	0.006	A/G	0.512	*GALC*[b]
30	EUR	16	11053656	rs117283010	3.07×10^{-16}	1.131	0.015	A/G	0.062	*CLEC16A*[c]
31	EUR	16	11185464	rs55898143	1.38×10^{-13}	1.081	0.011	T/C	0.085	*CLEC16A*[c]
32	EUR	16	11242497	rs794423	1.62×10^{-10}	1.104	0.015	A/C	0.059	*CLEC16A*[c]
33	EUR	16	11247847	rs80207443	1.60×10^{-13}	1.107	0.014	T/C	0.048	*CLEC16A*[c]
34	EUR	16	11335999	rs814260	9.30×10^{-9}	1.036	0.006	G/A	0.360	*SOCS1*[c]
35	EUR	16	11398467	rs10852332	4.02×10^{-9}	1.043	0.007	C/G	0.213	*PRM1*[c]
36	EUR	17	40508559	rs58905292	1.96×10^{-9}	1.106	0.017	A/T	0.049	*STAT3*[a]
37	EUR	17	57963873	rs1292052	1.49×10^{-9}	1.072	0.012	C/T	0.890	*TUBD1*[b]
38	EUR	20	47253487	rs3935549*	1.62×10^{-8}	1.034	0.006	C/T	0.506	*PREX1*[c]
1	AFR	9	1827489	rs76911648	3.28×10^{-9}	3.169	0.20	G/C	0.035	*SMARCA2*[c]
2	AMR	5	18904547	rs59061674	4.00×10^{-8}	3.461	0.23	G/A	0.038	*CDH18*[c]
3	AMR	15	77706452	rs113284638	3.82×10^{-8}	2.689	0.18	C/A	0.063	*PEAK1*[c]

Genes were assigned on the basis of [a]colocalization results, [b]SNP-to-Gene linking strategies and [c]nearest gene to the index. Anc., ancestry; AMR, Hispanic/Latinx; AFR, African-American; CHR, chromosome; BP, GRCh37 base-pair position; OR, odds ratio; RA/OA, risk/other allele; RAF, risk allele frequency using 1000 Genomes Project (1KG phase 3) EUR/AFR/AMR populations. *P*-values are two-sided and based on an inverse-variance-weighted fixed-effects meta-analysis. *Asterisks highlights susceptibility loci not previously associated with MS.

Table 2 | Performance metrics for the genome-wide polygenic score (GPS) in MS. A multi-ancestry GPS was constructed using PRS-CSx. GPS performance was evaluated in the remaining 30% of eMERGE-III participants of European ancestry (EUR) and in all participants of Admixed American (AMR) and African (AFR) ancestry. Odds ratios (ORs) with 95% confidence intervals (CIs) were estimated per standard deviation (SD) increase in GPS and across GPS percentile groups; unadjusted two-sided *P* values are reported. Predictive performance was assessed using the area under the receiver operating characteristic curve (AUC).

eMERGE-III (Ancestries)	Case/control	OR per SD (95% CI), *P*	AUC (Crude)	GPS threshold	OR (95% CI), *P* value
EUR (30%) genetic ancestry	287/22,796	1.79 (1.60-1.99), $P = 3.27\times10^{-25}$	0.7324 (0.6559)	Top 20% vs. other 80%	2.89 (2.27-3.67), $P = 3.44\times10^{-18}$
				Top 10% vs. other 90%	3.10 (2.36-4.07), $P = 4.84\times10^{-16}$
				Top 5% vs. other 95%	4.01 (2.92-5.53), $P = 1.90\times10^{-17}$
				Top 2% vs. other 98%	4.85 (3.14-7.50), $P = 1.05\times10^{-12}$
				Top 1% vs. other 99%	7.10 (4.22-11.90), $P = 1.31\times10^{-13}$
AFR genetic ancestry	142/15,600	1.28 (1.08-1.52), $P = 4.61\times10^{-3}$	0.7442 (0.5379)	Top 20% vs. other 80%	1.63 (0.83-3.20), $P = 0.160$
				Top 10% vs. other 90%	2.72 (1.27-5.80), $P = 9.58\times10^{-3}$
				Top 5% vs. other 95%	3.42 (1.28-9.13), $P = 0.014$
				Top 2% vs. other 98%	7.54 (2.50-22.80), $P = 3.47\times10^{-4}$
				Top 1% vs. other 99%	7.24 (1.59-33.00), $P = 0.011$
AMR genetic ancestry	55/5,148	1.74 (1.32-2.31), $P = 1.14\times10^{-4}$	0.7636 (0.5837)	Top 20% vs. other 80%	1.52 (1.04-2.23), $P = 0.032$
				Top 10% vs. other 90%	1.29 (0.77-2.14), $P = 0.331$
				Top 5% vs. other 95%	2.29 (1.32-3.97), $P = 0.003$
				Top 2% vs. other 98%	3.19 (1.52-6.70), $P = 0.002$
				Top 1% vs. other 99%	4.26 (1.80-10.10), $P = 0.001$

Figure legends

Figure 1 | MS GWAS study design. Top panel: four cohorts used in the meta-analysis. Middle panel: meta-analysis and the three analytical methods used. METAL provides a computationally efficient tool for meta-analysis of genome-wide association scans in European ancestry, MR-MEGA (middle) can identify risk variants with heterogeneous effects due to population stratification introduced by ancestry differences, whereas random-effect (bottom) is better suited for risk variants with homogeneous effect direction across different ancestries. Red dashed lines indicate *P*-value threshold of $P < 5\times10^{-8}$. Bottom panel: overview of downstream analyses and representative examples. Illustrations created in BioRender. Zeng, L. (2026) https://BioRender.com/cgzse7r.

Figure 2 | Circular presentation of loci associated with MS identified in European ancestry. The $-\log_{10}(P)$ (unadjusted two-sided *P* values from an inverse-variance-weighted fixed-effects meta-analysis) for genetic association with MS are arranged by chromosomal position, indicated by alternating blue and green points. Association *P* values are truncated at $P < 1\times10^{-30}$. Genome-wide significance ($P < 5\times10^{-8}$) is indicated by the red line. Genes showing Coloc effects with DLPFC cell types are highlighted in red, genes showing Coloc effects in PBMC cell types are highlighted in blue, and shared Coloc genes are annotated with black. The inner circle indicates MS loci that colocalize with DLPFC eQTLs, colored by cell type. Color keys representing cell types are indicated in the plot center. Chromosomes are indicated by numbered panels 1-22. Red diamonds highlight the four novel genomic risk loci for MS, defined as regions located more than 250 kb away from any previously reported MS-associated locus. Ast, astrocytes; Exc, excitatory neurons; Inh, inhibitory neurons; Mic, microglia; Oli, oligodendrocytes; OPC, oligodendrocyte precursor cells; PBMC, peripheral blood mononuclear cells; DLPFC, dorsolateral prefrontal cortex.

Figure 3 | Overlap of the genetic architecture of MS with other diseases and traits. a, Number of SNPs that reached genome-wide significance ($P < 5\times10^{-8}$) and were shared across 13 autoimmune diseases ($n = 32{,}901$). rs3184504 is highlighted as the only SNP associated with seven autoimmune diseases, including celiac disease, inflammatory bowel disease, multiple sclerosis, psoriasis, systemic lupus erythematosus, type 1 diabetes and thyroiditis. **b**, Percentage of non-MHC SNPs GWAS from MS severity, 12 inflammatory, 4 neurodegenerative, 4 psychiatric and 3 BMI-associated diseases or traits that are not statistically significant (NS), or significant in the same direction (SD) or the opposite direction (OD) in the current 236 MS susceptibility risk variants using a unadjusted *P*-value cut-off ($P \leq 0.05$). A two-sided binomial test was used to assess whether the observed number of SNPs with the same effect direction significantly differs from those with opposite directions (see **Supplementary Table 9**). **c**, Comparison of 45 MS risk variants across 24 diseases/traits. Colors represent effect directions and *P* values. White denotes SNPs that were not detected in the corresponding phenotype. **d**, Genetic correlation (r_g) estimated across MS and other 24 diseases/traits using LDSC. The areas of the squares represent the absolute value of corresponding genetic correlations. Statistical significance was evaluated using two-sided *P* values, and after Benjamini-Hochberg FDR correction for 325 tests at a 5% significance level, correlations significantly different from 0 are indicated by asterisks (*0.01 < FDR ≤ 0.05; **0.001 < FDR ≤ 0.01; ***FDR ≤ 0.001). Red color denotes positive genetic correlation, and blue color represents negative genetic correlation. Larger squares correspond to more significant *P* values. Tissue illustrations in **c** and **d** were created with BioRender. Zeng, L. (2026) https://BioRender.com/sp83o2t. CD, Crohn's disease; CeD, celiac disease; COPD, chronic obstructive pulmonary disease; IBD, inflammatory bowel disease; IGA, IgA nephropathy, OB, obesity; PS, psoriasis, RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; T1D, type 1 diabetes; TRD, thyroiditis; UC, ulcerative colitis; MS.sev, multiple sclerosis severity; AD, Alzheimer's disease; ALS, amyotrophic lateral sclerosis; FTD, frontotemporal dementia; PD, Parkinson's disease; BIP, bipolar disorder; MDD, major depressive disorder; Neuro, neuroticism;

SCZ, schizophrenia; BMI, body mass index; WHRadjBMI, waist-to-hip ratio adjusted BMI; T2D, type 2 diabetes.

Figure 4 | Workflow for the analysis of MS GPS. a, The MS GPS was developed using GWAS summary statistics from the IMSGC, All of Us (AoU), and UK Biobank (UKBB). Optimization was performed using 70% of European ancestry participants from eMERGE-III. GPS performance was validated in the remaining 30% of eMERGE-III participants of EUR and all AMR and AFR participants. **b-d**, PheWAS results are shown for EUR (n = 23,120) (**b**), AFR (n = 15,862) (**c**), and AMR (n = 5,223) (**d**) participants. The analysis includes eMERGE-III participants with both genotype and phenotype information. Logistic regression was performed adjusting for age, sex, batch, and genetic ancestry. Red horizontal lines indicate the phenome-wide significance threshold after Bonferroni correction for multiple testing ($P = 2.8\times10^{-5}$). The y-axis represents $-\log_{10}(P$ values) (two-sided, no adjustment for multiple testing), and the x-axis displays system-based phecode groupings. Upward-pointing triangles indicate increased odds for a given phecode, while downward-pointing triangles indicate reduced risk. **e**, Scatter plot showing the relationship between the MS GPS (with and without MHC variants) and white matter volume. The blue line represents the linear regression fit, and the gray shaded area indicates the 95% confidence interval of the regression model. P values were computed using two-sided t-tests.

Figure 5 | Overlap of the results from PBMC/DLPFC eQTLs and GWAS of MS. a, Heatmap reports the PP.H4 of the Coloc method, which assumes that GWAS and eQTLs share a single causal SNP. Rows represent the overlap for individual gene and SNP pairs, and columns represent the PP.H4 score in each of cell types. Color of each square is based on the code found to the right; the darker color denotes higher confidence that the same variant influences susceptibility and gene expression in that cell type. The top bar chart shows the number of colocalized eGenes with high confidence (PP.H4 > 0.8) in each cell type. **b**, Cartoon illustration summarizes the colocalization effects of neurons compared to the 18 cell types included in our analysis, colored by cell type. Illustrations in **b** were created in BioRender. Zeng, L. (2026) https://BioRender.com/k6pgcki. BIN, naïve B cells; Bmem, memory B cells; CD4ET, $CD4^+$ effector memory T cells; CD4NC, $CD4^+$ naïve and central memory T cells; CD4SOX4, $CD4^+$ T cells expressing SOX4; CD8ET, $CD8^+$ effector memory T cells; CD8NC, $CD8^+$ naïve and central memory T cells; CD8S100B, $CD8^+$ T cells expressing S100B; DC, dendritic cells; MonoC, classical monocytes; MonoNC, nonclassical monocytes; NKR, NK recruiting cells; NK, natural killer cells; Ast, astrocytes; Exc, excitatory neurons; Inh, inhibitory neurons; Mic, microglia; Oli, oligodendrocytes; OPC, oligodendrocyte progenitor cells.

Figure 6 | Examples of colocalization results. a, LocusCompare plots showing the colocalization between *STAT3* cis-eQTLs in inhibitory neurons and MS GWAS. The y-axis shows $-\log_{10}(P$ values) (two-sided, no adjustment for multiple testing) for the MS GWAS, and x-axis shows $-\log_{10}(P$ values) (two-sided, no adjustment for multiple testing) for the snucRNA-seq eQTL in corresponding cell type. **b**, Association between rs1026916 and serum neurofilament light chain (sNfL) levels in 3,480 individuals with MS of European ancestry. Red dots indicate the difference in mean change in sNfL value for each genotype relative to the reference genotype (GG), and error bars represent the 95% confidence intervals. A linear regression was applied to patient-specific slopes of sNfL changes and allele dosage. Effect size β is shown in the figure. P values were computed using two-sided t-tests. **c-e**, Expression quantitative trait loci (eQTL) box plots of associations between genotype rs1026916 and *STAT3* expression in inhibitory neurons using snucRNA-seq data from Fujita et al. (CUIMC study 1) (**c**), Mathys et al. (MIT cohort) (**d**), and our in-house multiome datasets (CUIMC study 2) (**e**). A linear regression was applied to *STAT3* expression and allele dosage. Effect size β is shown in the figure. P values were computed using two-sided t-tests. Centerline represents the median expression value, lower and upper hinges represent the lower and upper quartiles, respectively, and whiskers represent 1.5× interquartile range (IQR). **f**, Immunohistochemistry of DLPFC in human MS brain tissue (n = 1), stained for STAT3 (green), GAD1/2

(red), and NeuN (yellow), with DAPI (blue) to visualize nuclei. Expression of STAT3 was observed in $NeuN^+GAD1/2^+$ neurons. White triangles highlight the colocalization of DAPI, STAT3, GAD1/2, and NeuN. Scale bar, 50 μm. This analysis was performed on tissue from a single MS brain sample; no independent biological replicates were available.

References

1. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* **365**, eaav7188 (2019).
2. International Multiple Sclerosis Genetics Consortium & MultipleMS Consortium. Locus for severity implicates CNS resilience in progression of multiple sclerosis. *Nature* **619**, 323–331 (2023).
3. De Jager, P. L. *et al.* The role of the CD58 locus in multiple sclerosis. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 5264–5269 (2009).
4. International Multiple Sclerosis Genetics Consortium *et al.* Risk alleles for multiple sclerosis identified by a genomewide study. *N. Engl. J. Med.* **357**, 851–862 (2007).
5. Ottoboni, L. *et al.* Clinical relevance and functional consequences of the TNFRSF1A multiple sclerosis locus. *Neurology* **81**, 1891–1899 (2013).
6. Ma, Q. *et al.* Integration of epigenetic and genetic profiles identifies multiple sclerosis disease-critical cell types and genes. *Commun. Biol.* **6**, 342 (2023).
7. Gregory, S. G. *et al.* Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis. *Nat. Genet.* **39**, 1083–1091 (2007).
8. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
9. Raj, T. *et al.* Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* **344**, 519–523 (2014).
10. Maier, L. M. *et al.* Soluble IL-2RA levels in multiple sclerosis subjects and the effect of soluble IL-2RA on immune responses. *J. Immunol.* **182**, 1541–1547 (2009).
11. Ponath, G., Park, C. & Pitt, D. The role of astrocytes in multiple sclerosis. *Front. Immunol.* **9**, 217 (2018).
12. Absinta, M. *et al.* A lymphocyte-microglia-astrocyte axis in chronic active multiple sclerosis. *Nature* **597**, 709–714 (2021).
13. Jäkel, S. *et al.* Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* **566**, 543–547 (2019).
14. Dutta, R. & Trapp, B. D. Mechanisms of neuronal dysfunction and degeneration in multiple sclerosis. *Prog. Neurobiol.* **93**, 1–12 (2011).
15. Schirmer, L. *et al.* Neuronal vulnerability and multilineage diversity in multiple sclerosis. *Nature* **573**, 75–82 (2019).
16. Wegner, C., Esiri, M. M., Chance, S. A., Palace, J. & Matthews, P. M. Neocortical neuronal, synaptic, and glial loss in multiple sclerosis. *Neurology* **67**, 960–967 (2006).
17. Parnell, G. P. & Booth, D. R. The multiple sclerosis (MS) genetic risk factors indicate both acquired and innate immune cell subsets contribute to MS pathogenesis and identify novel therapeutic opportunities. *Front. Immunol.* **8**, 425 (2017).

18. Patsopoulos, N. A. *et al.* Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Ann. Neurol.* **70**, 897–912 (2011).
19. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
20. eMERGE Consortium. Harmonizing clinical sequencing and interpretation for the eMERGE III Network. *Am. J. Hum. Genet.* **105**, 588–605 (2019).
21. Ramirez, A. H. *et al.* The All of Us Research Program: data quality, utility, and diversity. *Patterns* **3**, 100570 (2022).
22. Khan, A. *et al.* Polygenic risk alters the penetrance of monogenic kidney disease. *Nat. Commun.* **14**, 8318 (2023).
23. Fujita, M. *et al.* Cell subtype-specific effects of genetic variation in the Alzheimer's disease brain. *Nat. Genet.* **56**, 605–614 (2024).
24. Yazar, S. *et al.* Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science* **376**, eabf3041 (2022).
25. Ruan, Y. *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nat. Genet.* **54**, 573–580 (2022).
26. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
27. Zhernakova, A. *et al.* Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. *PLoS Genet.* **7**, e1002004 (2011).
28. Eyre, S. *et al.* High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat. Genet.* **44**, 1336–1340 (2012).
29. Wiley, M. M. *et al.* Variants in the DDX6-CXCR5 autoimmune disease risk locus influence the regulatory network in immune cells and salivary gland. *Ann. Rheum. Dis.* **84**, 1512–1527 (2025).
30. Nakatsuka, N. *et al.* Two genetic variants explain the association of European ancestry with multiple sclerosis risk in African-Americans. *Sci. Rep.* **10**, 16902 (2020).
31. Zhang, Z. *et al.* BRM/SMARCA2 promotes the proliferation and chemoresistance of pancreatic cancer cells by targeting JAK2/STAT3 signaling. *Cancer Lett.* **402**, 213–224 (2017).
32. Valori, M., Jansson, L. & Tienari, P. J. CD8+ cell somatic mutations in multiple sclerosis patients and controls-Enrichment of mutations in STAT3 and other genes implicated in hematological malignancies. *PloS One* **16**, e0261002 (2021).
33. Mägi, R. *et al.* Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum. Mol. Genet.* **26**, 3639–3650 (2017).
34. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
35. Isobe, N. *et al.* An ImmunoChip study of multiple sclerosis risk in African Americans. *Brain J. Neurol.* **138**, 1518–1530 (2015).
36. Beecham, A. H. *et al.* The genetic diversity of multiple sclerosis risk among Hispanic and African American populations living in the United States. *Mult. Scler.* **26**, 1329–1339 (2020).
37. Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53**, 1415–1424 (2021).

38. Jiang, L., Zheng, Z., Fang, H. & Yang, J. A generalized linear mixed model association tool for biobank-scale data. *Nat. Genet.* **53**, 1616–1621 (2021).
39. Stuart, P. E. *et al.* Transethnic analysis of psoriasis susceptibility in South Asians and Europeans enhances fine-mapping in the MHC and genomewide. *HGG Adv.* **3**, 100069 (2022).
40. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).
41. Julià, A. *et al.* Genome-wide association study meta-analysis identifies five new loci for systemic lupus erythematosus. *Arthritis Res. Ther.* **20**, 100 (2018).
42. Chiou, J. *et al.* Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. *Nature* **594**, 398–402 (2021).
43. Dubois, P. C. A. *et al.* Multiple common variants for celiac disease influencing immune gene expression. *Nat. Genet.* **42**, 295–302 (2010).
44. de Lange, K. M. *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261 (2017).
45. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436 (2022).
46. van Rheenen, W. *et al.* Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nat. Genet.* **53**, 1636–1648 (2021).
47. Ferrari, R. *et al.* Frontotemporal dementia and its subtypes: a genome-wide association study. *Lancet Neurol.* **13**, 686–699 (2014).
48. Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **18**, 1091–1102 (2019).
49. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).
50. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).
51. Nagel, M. *et al.* Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat. Genet.* **50**, 920–927 (2018).
52. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
53. Xue, A. *et al.* Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat. Commun.* **9**, 2941 (2018).
54. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174 (2019).
55. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
56. Lincoln, M. R. *et al.* Genetic mapping across autoimmune diseases reveals shared associations and mechanisms. *Nat. Genet.* **56**, 838–845 (2024).
57. Yang, Y. *et al.* Investigating the shared genetic architecture between multiple sclerosis and inflammatory bowel diseases. *Nat. Commun.* **12**, 5641 (2021).

58. Xia, Z. *et al.* Genes and Environment in Multiple Sclerosis project: a platform to investigate multiple sclerosis risk. *Ann. Neurol.* **79**, 178–189 (2016).
59. Gauthier, S. A., Glanz, B. I., Mandel, M. & Weiner, H. L. A model for the comprehensive investigation of a chronic autoimmune disease: the multiple sclerosis CLIMB study. *Autoimmun. Rev.* **5**, 532–536 (2006).
60. Kaskow, B. J. & Baecher-Allan, C. Effector T cells in multiple sclerosis. *Cold Spring Harb. Perspect. Med.* **8**, a029025 (2018).
61. Fletcher, J. M., Lalor, S. J., Sweeney, C. M., Tubridy, N. & Mills, K. H. G. T cells in multiple sclerosis and experimental autoimmune encephalomyelitis. *Clin. Exp. Immunol.* **162**, 1–11 (2010).
62. Rébé, C., Végran, F., Berger, H. & Ghiringhelli, F. STAT3 activation: a key factor in tumor immunoescape. *JAK-STAT* **2**, e23010 (2013).
63. Ponath, G. *et al.* Enhanced astrocyte responses are driven by a genetic risk allele associated with multiple sclerosis. *Nat. Commun.* **9**, 5337 (2018).
64. Sherman, B. T. *et al.* DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **50**, W216–W221 (2022).
65. Mathys, H. *et al.* Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer's disease pathology. *Cell* **186**, 4365–4385.e27 (2023).
66. Xiong, X. *et al.* Epigenomic dissection of Alzheimer's disease pinpoints causal variants and reveals epigenome erosion. *Cell* **186**, 4422–4437.e21 (2023).
67. Le Grand, Q. *et al.* Diffusion imaging genomics provides novel insight into early mechanisms of cerebral small vessel disease. *Mol. Psychiatry* **29**, 3567–3579 (2024).
68. Duncan, L. *et al.* Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* **10**, 3328 (2019).
69. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
70. Shi, H. *et al.* Population-specific causal disease effect sizes in functionally important regions impacted by selection. *Nat. Commun.* **12**, 1098 (2021).
71. Chun, S. *et al.* Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat. Genet.* **49**, 600–605 (2017).
72. Mu, Z. *et al.* The impact of cell type and context-dependent regulatory variants on human immune traits. *Genome Biol.* **22**, 122 (2021).
73. Arthur, T. D. *et al.* Multiomic QTL mapping reveals phenotypic complexity of GWAS loci and prioritizes putative causal variants. *Cell Genomics* **5**, 100775 (2025).
74. Magyari, M. & Sorensen, P. S. Comorbidity in multiple sclerosis. *Front. Neurol.* **11**, 851 (2020).
75. Hauer, L., Perneczky, J. & Sellner, J. A global view of comorbidity in multiple sclerosis: a systematic review with a focus on regional differences, methodology, and clinical implications. *J. Neurol.* **268**, 4066–4077 (2021).
76. Criswell, L. A. *et al.* Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. *Am. J. Hum. Genet.* **76**, 561–571 (2005).

77. Devallière, J. & Charreau, B. The adaptor Lnk (SH2B3): an emerging regulator in vascular cells and a link between immune and inflammatory signaling. *Biochem. Pharmacol.* **82**, 1391–1402 (2011).
78. Liu, L. *et al.* Genetic regulation of serum IgA levels and susceptibility to common immune, infectious, kidney, and cardio-metabolic traits. *Nat. Commun.* **13**, 6859 (2022).
79. Winer, H. *et al.* IL-7: Comprehensive review. *Cytokine* **160**, 156049 (2022).
80. Schwaiger, F. W. *et al.* Peripheral but not central axotomy induces changes in Janus kinases (JAK) and signal transducers and activators of transcription (STAT). *Eur. J. Neurosci.* **12**, 1165–1176 (2000).
81. Lang, C., Bradley, P. M., Jacobi, A., Kerschensteiner, M. & Bareyre, F. M. STAT3 promotes corticospinal remodelling and functional recovery after spinal cord injury. *EMBO Rep.* **14**, 931–937 (2013).
82. Luo, X. *et al.* Enhanced transcriptional activity and mitochondrial localization of STAT3 co-induce axon regrowth in the adult central nervous system. *Cell Rep.* **15**, 398–410 (2016).
83. Selvaraj, B. T., Frank, N., Bender, F. L. P., Asan, E. & Sendtner, M. Local axonal function of STAT3 rescues axon degeneration in the pmn model of motoneuron disease. *J. Cell Biol.* **199**, 437–451 (2012).
84. Chen, E. *et al.* A novel role of the STAT3 pathway in brain inflammation-induced human neural progenitor cell differentiation. *Curr. Mol. Med.* **13**, 1474–1484 (2013).
85. Su, Y. *et al.* STAT3 regulates mouse neural progenitor proliferation and differentiation by promoting mitochondrial metabolism. *Front. Cell Dev. Biol.* **8**, 362 (2020).
86. Bareyre, F. M. *et al.* In vivo imaging reveals a phase-specific role of STAT3 during central and peripheral nervous system axon regeneration. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 6282–6287 (2011).
87. Martín-Suárez, S., Cortes, J. M. & Bonifazi, P. Blockage of STAT3 during epileptogenesis prevents GABAergic loss and imprinting of the epileptic state. *Brain J. Neurol.* **146**, 3416–3430 (2023).
88. Liu, Y., Ma, D. & Ji, C. Zinc fingers and homeoboxes family in human diseases. *Cancer Gene Ther.* **22**, 223–226 (2015).
89. Igata, T. *et al.* Loss of the transcription repressor ZHX3 induces senescence-associated gene expression and mitochondrial-nucleolar activation. *PloS One* **17**, e0262488 (2022).
90. Woo, M. S., Engler, J. B. & Friese, M. A. The neuropathobiology of multiple sclerosis. *Nat. Rev. Neurosci.* **25**, 493–513 (2024).
91. Ellwardt, E. *et al.* Maladaptive cortical hyperactivity upon recovery from experimental autoimmune encephalomyelitis. *Nat. Neurosci.* **21**, 1392–1403 (2018).
92. Zoupi, L. *et al.* Selective vulnerability of inhibitory networks in multiple sclerosis. *Acta Neuropathol. (Berl.)* **141**, 415–429 (2021).
93. Smolders, J. *et al.* Tissue-resident memory T cells populate the human brain. *Nat. Commun.* **9**, 4593 (2018).
94. Baecher-Allan, C., Kaskow, B. J. & Weiner, H. L. Multiple sclerosis: mechanisms and immunotherapy. *Neuron* **97**, 742–768 (2018).
95. Salvador, A. F., de Lima, K. A. & Kipnis, J. Neuromodulation by the immune system: a focus on cytokines. *Nat. Rev. Immunol.* **21**, 526–541 (2021).
96. Zipp, F., Bittner, S. & Schafer, D. P. Cytokines as emerging regulators of central nervous system synapses. *Immunity* **56**, 914–925 (2023).

Methods

Ethics

Our study complied with all relevant ethical regulations and was approved by the institutional review boards (IRB-AAAR4962) at the Columbia University Irving Medical Center.

Study design

This cross-sectional study involves a combined analysis of the IMSGC, UKBB, eMERGE-III, and AoU cohorts (**Supplementary Methods**). All participants provided informed consent to participate in genetic studies. Each cohort was first analyzed separately, and cohort-specific results were combined using fixed-effects meta-analysis.

MS phenotyping and case-control definitions

The MS phenotype was defined using ICD codes from the UKBB, eMERGE-III, and AoU datasets. Cases were identified by at least one occurrence of the following ICD codes: ICD-9: 340, 323, or 341. Participants without any of these codes were classified as controls.

Meta-GWAS

The MHC is the most gene-dense and most polymorphic part of the human genome. The region exhibits haplotype-specific linkage disequilibrium patterns, extreme structural variation and copy number variations, and an extremely high level of genetic diversity; the use of a single reference sequence to analyze GWAS data in this area is problematic[97]. Therefore, the Extended MHC region (xMHC) was set aside in our meta-analysis (defined as the regions between *HIST1H2AA* and *RPL12P1* genes: chr6: 25,383,722-33,368,421Mb; ~7.6Mb, GRCh37), resulting in ~68,000 SNPs located in xMHC being removed from the primary meta-analysis. However, association results within this region are provided separately in **Supplementary Figure 10b-e** and **Supplementary Table 20** for completeness.

We first performed a meta-analysis using an inverse-variance-weighted fixed-effects model in METAL (version 2011-03-25)[26] combining UKBB, AoU, and eMERGE-III cohorts for European ancestry (5,063 MS cases and 596,340 controls), African-American (614 MS cases and 62,044 controls) and Hispanic American (352 MS cases and 44,133 controls) populations, respectively. In addition, another meta-analysis using METAL was performed exclusively for the European ancestry cohort, which included the GWAS summary statistics from the IMSGC discovery cohort[1], along with UKBB, AoU, and eMERGE-III (19,865 MS cases and 623,043 controls). A genome-wide significant locus was defined as the region around a SNP with $P < 5 \times 10^{-8}$ and LD $r^2 > 0.1$, within a 500-kb window, using the reference panel from phase 3 of the 1000 Genomes Project as the reference population.

Two models were used to conduct multi-ancestry meta-analyses (20,831 MS cases and 729,220 controls). Random effects models were performed using PLINK (v1.9)[34], while a separate analysis was performed using MR-MEGA (v0.2)[33]. PLINK v1.9 was preferred over METAL due to its capacity to perform random effects analyses in parallel. A random effects model provides a more conservative framework that allows each study to have unique effects, as expected in different populations. MR-MEGA was also employed since it is well-powered to detect associations at loci with allelic heterogeneity. MR-MEGA models allelic effects as a function of

axes of genetic variation that are derived from the input GWAS summary statistics. This method can result in reduced variant sets since it requires that variants have sufficient overlap between the input datasets ($k > 3$), where k is the number of inputs GWAS, in contrast to random effects models implemented in PLINK v1.9, which were limited to $k > 2$ to quantify heterogeneity accurately.

To identify novel genomic risk loci, LD blocks of independent significant SNPs ($r^2 > 0.1$, ±500 kb, 1KG phase 3) were merged into a single genomic locus if the distance between LD blocks was less than 250 kb. These loci were compared to the previous GWAS[1] to assess whether these regions were known to be associated with MS. There was no evidence of stratification artifacts or uncontrolled inflation of test statistics in the results from any cohort ($\lambda_{GC} = 1.02–1.14$; **Supplementary Fig. 3a-d**).

Conditional analysis

To identify secondary association signals, we used the program GCTA-COJO[98] to perform conditional analysis on the summary meta-analysis. Specifically, we applied GCTA-COJO (--cojo-cond) to assess the loci with multiple nearby risk variants within ±100 kb (e.g., the *DDX6*, *CD6*, *AHI1* and *PTGER* loci, where multiple nearby variants suggested possible independent effects using LD clumping). This method enables approximate conditional association testing based on summary-level GWAS data, using LD information from a reference panel instead of individual-level genotypes.

Summary statistics for autoimmune diseases and other traits

We downloaded complete summary statistics for autoimmune and inflammatory disease GWAS available in the NHGRI-EBI GWAS Catalog (https://www.ebi.ac.uk/gwas/downloads/summary-statistics) and PubMed (https://pubmed.ncbi.nlm.nih.gov/) (**Supplementary Table 8**). We focused on European ancestry studies with at least 2,000 study participants for which signed summary statistics were available. We chose the study with the largest cohort size, among those available for a given trait. By applying these filters, we obtained GWAS statistics for the IgA nephropathy[37], chronic obstructive pulmonary disease[38], obesity[38], psoriasis[39], rheumatoid arthritis[40], systemic lupus erythematosus[41], type 1 diabetes[42], thyroiditis[37], celiac disease[43], and inflammatory bowel disease (IBD), for which IBD summary statistics also included results for Crohn's disease and ulcerative colitis[44]. We also downloaded GWAS summary statistics for four neurodegenerative diseases (Alzheimer's disease[45], amyotrophic lateral sclerosis[46], frontotemporal dementia[47], and Parkinson's disease[48]), four psychiatric disorders/traits (bipolar disorder[49], major depressive disorder[50], neuroticism[51], and schizophrenia[52]), and three metabolic traits (type 2 diabetes[53], body mass index (BMI), and waist-to-hip ratio adjusted BMI[54]). Given that most of the GWAS we collected were conducted in participants of European ancestry, we used the results of updated MS GWAS summary statistics in European ancestry for this analysis.

We removed the Extended MHC region (xMHC) from the summary statistics (defined as the regions between *HIST1H2AA* and *RPL12P1* genes: chr6: 25,383,722-33,368,421Mb; ~7.6Mb, GRCh37). We then removed indels and SNPs inconsistent with the 1000 Genomes Project (phase 3) reference panel and filtered for strand-unambiguous biallelic SNPs with minor allele frequency (MAF) > 0.01 in the 1000 Genomes European (EUR) reference individuals.

Cross-trait LD score regression

LDSC v1.0.1[99] bivariate genetic correlations attributed to genome-wide SNPs (r_g) were estimated across 25 human diseases/traits from published GWAS datasets, as mentioned above. We used LD scores from the 'eur_w_ld_chr' file available from https://alkesgroup.broadinstitute.org/LDSCORE, computed using 1000 Genomes Project[100] Europeans as a reference panel[101]. Bonferroni adjusted FDR < 0.05 was used to define significant genetic correlations by adjusting for the number of traits tested.

In addition, LDSC was used to estimate the phenotypic variance in MS explained by common SNPs (SNP-heritability, h^2_{SNP}) from GWAS summary statistics using 1000 Genomes Phase 3 reference panel.

Genome-wide polygenic score (GPS) design and optimization

We used PRS-CSx, a Bayesian polygenic modeling framework, to develop genomic prediction scores (GPS) across diverse ancestries[25]. PRS-CSx v.1.1.0 integrates GWAS summary statistics from multiple populations, accounting for population-specific LD patterns. Specifically, we utilized GWAS summary statistics from three ancestral groups: African (AFR), European (EUR), and Admixed American (AMR), and combined them using the 'meta' setting in PRS-CSx. In our study, 70% of the training data consisted of individuals of European ancestry from the eMERGE-III cohort (615 MS cases and 53,250 controls) to optimize model selection. To ensure no overlap between the GWAS discovery cohort and the GPS development dataset, the eMERGE-III dataset was excluded from the MS GWAS discovery cohort. We evaluated model robustness by running PRS-CSx with different values of the global shrinkage parameter: 1, 10^{-1}, 10^{-2}, 10^{-4}, 10^{-6}, and 10^{-8}. The final GPS was selected based on the best-performing model for the training dataset (**Supplementary Table 11**). The score was standardized to zero mean and unit variance based on ancestry-matched population controls. In the optimization dataset, the shrinkage parameter (10^{-4}) explained 2% of the variance (R^2), with 1 s.d. of the score increasing MS risk by 62% (odds ratio (OR) = 1.62, 95% confidence interval (CI) = 1.49–1.75, $P < 5.33 \times 10^{-32}$) after controlling for age, sex, batch effects, and four genetic PCs. The final PRS-CSx output included 1,161,784 HapMap3[102] variants and their weights.

PheWAS

The derived polygenic predictors for MS were used to score all 102,138 eMERGE-III participants with available genotypes and electronic health record (EHR) data. To test the association of these polygenic predictors with diseases in a phenome-wide manner, we first harmonized the diagnostic data by converting all available ICD-10-CM codes to the ICD-9-CM system. A total of 102,138 genotyped eMERGE-III participants had 20,783 unique ICD-9 codes, which were subsequently mapped to 1,817 distinct phecodes. Phenome-wide association analyses (PheWAS) were conducted using the PheWAS R package[103], which applies predefined control groups for each phecode. For case definition, at least two occurrences of ICD-9 codes within the case grouping of each phecode were required. Logistic regression was used to test associations between the MS polygenic score and each of the 1,817 phecodes, with case-control status as the outcome. The polygenic score for MS was adjusted for age, sex, study site, and ancestry's first three principal components (PCs). We applied a Bonferroni correction for multiple testing to determine statistically significant disease associations, setting the significance threshold at 2.75×10^{-5} (0.05 divided by 1,817).

MRI analysis

Multiple sclerosis (MS) participants were from the Comprehensive Longitudinal Investigation of Multiple Sclerosis at the Brigham and Women's Hospital (CLIMB) study[59]. CLIMB is a natural history observational study of MS in which participants undergo semi-annual neurological examinations and annual magnetic resonance imaging (MRI). MS lesions and brain tissue compartments (gray matter, white matter, and cerebrospinal fluid) were segmented using template-driven segmentation and partial volume artifact correction (TDS+) method[104]. Results underwent quality control and manual correction where necessary[105] (**Supplementary Fig. 10**). MRI and genome-wide genotyping data were available for 145 MS patients; 136 of them were European ancestry, 7 were AFR ancestry, and 2 were Hispanic. Among them, 130 are diagnosed with relapsing-remitting MS, and 15 are clinically isolated syndrome. GPS score for each participant was calculated using the PLINK command '--bfile --score sum –out'[34], and a regression model was used to test the association between GPS and MRI, adjusted for age at visit, sex, and top three genotype PCs.

Serum neurofilament light chain analysis

For this analysis, we included participants from the Multiple Sclerosis Partners Advancing Technology and Health Solutions (MS PATHS) which is a 10 site consortium of MS centers from the US and Europe in which standardized clinical and neuroperformance information is collected during routine clinic visits. A subset of MS PATHS participants provides blood samples which were used to characterize genetic information and serum neurofilament light chain levels (sNfL). Whole genome sequencing for MS PATHS was performed at 30× as a part of a larger effort. Joint genotyping was performed using GATK best practices including the variant quality score recalibration for final variant filtering and further filtered by the removal of variants with genotype quality scores <20, variants with missing rates >10% or those which were monomorphic. We also removed variants with allelic imbalances of 0.15 for SNPs and 0.20 for indels. For sNfL, levels were quantified by the Siemens Healthcare Laboratory, LLC (Berkeley, CA, USA) using a high-throughput scalable acridinium-ester immunoassay, which has a range of 3.9 to 500 pg/mL and excellent repeatability on the on the Atellica® Solution platform. sNfL measures were then translated to age-normative *Z* scores from a reference of healthy controls MS using Generalized Additive Models for Location, Scale, and Shape. To assess longitudinal change in sNfL levels at the individual level, we fit linear mixed effects models with patient-specific slopes and intercepts; samples obtained within 6 months of a recent relapse were excluded. We then extracted patient-specific slopes of sNfL changes and regressed these values on rs1026916 adjusting for baseline age, sex, and five ancestry principal components.

Colocalization analysis

The COLOC package (version 5.1.0)[106] was applied to test the approximate Bayes factor (ABF) colocalization hypothesis, which assumes a single causal variant. Under ABF analysis, the association of a trait with a SNP is assessed by calculating the posterior probability (value from 0 to 1), with the value of 1 indicating the causal SNP. In addition, the ABF analysis has five hypotheses, where PP.H0.abf indicates there is neither an eQTL nor a GWAS signal at the loci; PP.H1.abf indicates the locus is only associated with the GWAS; PP.H2.abf indicates the locus is only associated with the eQTL; PP.H3.abf indicates that both the GWAS and eQTL are associated but to a different genetic variant; PP.H4.abf indicates that the eQTL and the GWAS

are associated to the same genetic variant. With the posterior probability of each SNP and aiming to find the casual variants between the GWAS and eQTL, we focused on extracting the PP.H4 value for each SNP in our study.

For MS GWAS, we used the reported lead SNPs of 236 loci. For each locus, we searched for the expression-associated SNPs (eSNPs) that are within 500 kb of the lead SNP, and listed eGenes that were paired with the eSNP. We then obtained the cis-eQTL output for the eGenes within a 1-Mb window centered around the lead SNP. In addition, we extracted GWAS summary statistics around the reported 236 lead SNP. At last, we conducted COLOC for respective pairs of cis-eQTL and eSNP-GWAS (SNPs that serve as both eQTLs and show association with MS susceptibility in GWAS) for each cell type, using eQTL summary statistics from the OneK1K cohort (982 PBMC samples, 14 blood cell types, browsable results are available at www.onek1k.org)[24] and ROSMAP (424 DLPFC samples, 6 brain cell types, https://doi.org/10.7303/syn52335732) cohort[23].

Summary-based Mendelian randomization (SMR)

SMR (v1.03)[107] was used to evaluate whether genes colocalized with MS risk loci were associated with susceptibility via their *cis*-regulated gene expression in DLPFC/PBMC. We used the eQTL results and our MS GWAS EUR summary statistics to perform SMR. We applied two filtering strategies to assessed whether the MS GWAS and eQTL associations for colocalized genes reflect a shared causal variant consistent with mediation through gene expression. The first approach used the filtering criteria from the OneK1K[24] SMR pipeline: SMR FDR < 0.05, MS GWAS $P < 1\times10^{-8}$, and eQTL $P < 1\times10^{-4}$, while the second applied an SMR FDR < 0.05, and a HEIDI $P > 0.05$ to control for heterogeneity. Multiple testing correction was performed using the Benjamini-Hochberg false discovery rate (FDR) method.

Immunohistochemistry staining for STAT3 and GAD1+GAD2

For validation immunostaining, a 6-µm formalin-fixed paraffin-embedded (FFPE) tissue section from the dorsolateral prefrontal cortex (Brodmann Area 9) of an MS individual was obtained from the New York Brain Bank at Columbia University. The tissue was stained with STAT3 (clone 9D8, 1:300, Abcam, catalog no. ab119352, lot no. 1076952-1), GAD1+GAD2 (clone EPR19366, 1:100, Abcam, catalog no. ab183999, lot no. 1071027-7). The FFPE tissue section was deparaffinized using CitriSolv (d-limonene, Decon Laboratories, Inc. cat.# 1601H) as a clearing agent for 20 min. The section was rehydrated and prepared for staining through a series of graded ethanol washes. Heat-mediated antigen retrieval was performed with citrate buffer (pH 6, Sigma-Aldrich catalog no. C9999) using a microwave (800 W, 30% power setting) for 25 minutes. Following this, the section was blocked for 30 min at room temperature (RT) using a Bovine Serum Albumin-blocking medium (BSA, 3%, Sigma-Aldrich, catalog no. A7906) to minimize non-specific antibody binding. The section was incubated overnight with the primary antibodies (anti-STAT3 and anti-GAD1+GAD2) at 4 °C. After washing, the tissues were incubated for 1 h with fluorochrome-conjugated secondary antibodies (donkey anti-mouse Alexa Fluor 488: polyclonal, 1:500, Invitrogen, catalog no. A21202, lot no. 2563848 and donkey anti-rabbit Alexa Fluor 555: polyclonal, 1:500, Invitrogen, catalog no. A10042, lot no. 2540901) to bind to the primary antibody for protein detection and signal enhancement. After washing, the slides were again incubated in 3% BSA for 30 min and stained with the NeuN-conjugated-647 antibody (clone EPR12763, 1:100, Abcam, catalog no. ab190565, lot no. 1015266-1) for 1 h at

RT. After incubation, the section was washed and treated with True Black Lipofuscin Autofluorescence Quencher for 2 min at RT to minimize endogenous autofluorescence. An anti-fading DAPI mounting agent (347 channel, Invitrogen, catalog no. P36931) was used to coverslip.

Images were acquired using the Nikon Eclipse Ni-E immunofluorescence microscope at a magnification of ×20), and approximately 44 pictures were acquired in zigzag pattern to cover all cortical layers from the MS individual. The captured images were analyzed using CellProfiler[108] software. An extensive pipeline has been developed to automatically segment the neurons and detect STAT3 expressed by $GAD1^+$ and $GAD2^+$ cells[109]. DAPI and NeuN was defined as the primary object using the "IdentifyPrimaryObjects" module. The Robust Background method was used for thresholding. The typical diameter for DAPI objects was set to range between 15 and 80 pixels and between 30 and 80 pixels for NeuN objects. Then, the 'RelateObjects' module was applied to filter NEUN objects positive for DAPI objects (NeuN+DAPI+). The module "IdentifyPrimaryObjects" was used to segment GAD1/GAD2+ cells, using the Robust Background as the thresholding method, with a typical diameter ranging from 30 to 80 pixels. The segmented GAD1/GAD2+ objects were related to NeuN+DAPI+ filter GAD1/GAD2+NeuN+DAPI+ objects. The STAT3 intensity was measured within the GAD1/GAD2+NeuN+DAPI+ objects.

Data availability

GWAS summary statistics are publicly available on the Zenodo website (https://zenodo.org/records/17593969). The final formulation of the GPS for MS was deposited at https://zenodo.org/records/15537740. The DLPFC cell type-level eQTL summary statistics are available at Synapse (https://doi.org/10.7303/syn52335732). The PBMC cell type-level eQTLs are available at https://onek1k.org/. This study included some publicly available datasets accessed through the 1000 Genomes Phase 3 reference panel (https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase3/) and the HRC reference panel v1.0 (http://www.haplotype-reference-consortium.org/home). The summary statistics used in the autoimmune diseases analyses can be accessed from the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) website (https://www.ebi.ac.uk/gwas/) or are available in the referenced publications.

Code availability

All software used in the study is publicly available as described in the Methods. METAL v.2011-03-25 (https://genome.sph.umich.edu/wiki/METAL), PLINK v.1.9 (https://www.cog-genomics.org/plink/1.9/), MR-MEGA v0.2 (https://tools.gi.ut.ee/tools/MR-MEGA_v0.2.zip), GCTA-COJO (https://yanglab.westlake.edu.cn/software/gcta/#COJO), PRS-CSx v.1.1.0 (https://github.com/getian107/PRScsx), COLOC v5.1.0 (https://chr1swallace.github.io/coloc/), LDSC v1.0.1 (https://github.com/bulik/ldsc), SMR v1.0.3 (https://yanglab.westlake.edu.cn/software/smr/#Overview). The custom code for the analyses can be accessed at Zenodo (https://zenodo.org/records/18895815) and GitHub (https://github.com/luzengAdelaide/MS-GWAS).

Methods-only References

97. Kennedy, A. E., Ozbek, U. & Dorak, M. T. What has GWAS done for HLA and disease associations? *Int. J. Immunogenet.* **44**, 195–211 (2017).
98. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
99. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
100. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
101. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
102. International HapMap 3 Consortium *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
103. Denny, J. C. *et al.* PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* **26**, 1205–1210 (2010).
104. Wu, Y. *et al.* Automated segmentation of multiple sclerosis lesion subtypes with multichannel MRI. *NeuroImage* **32**, 1205–1215 (2006).
105. Roostaei, T. *et al.* Cell type- and state- resolved immune transcriptomic profiling identifies glucocorticoid-responsive molecular defects in multiple sclerosis T cells. 2022.06.29.498195 Preprint at https://doi.org/10.1101/2022.06.29.498195 (2022).
106. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
107. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
108. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
109. Duba-Kiss, R., Niibori, Y. & Hampson, D. R. GABAergic gene regulatory elements used in adeno-associated viral vectors. *Front. Neurol.* **12**, 745159 (2021).

Figure 1.

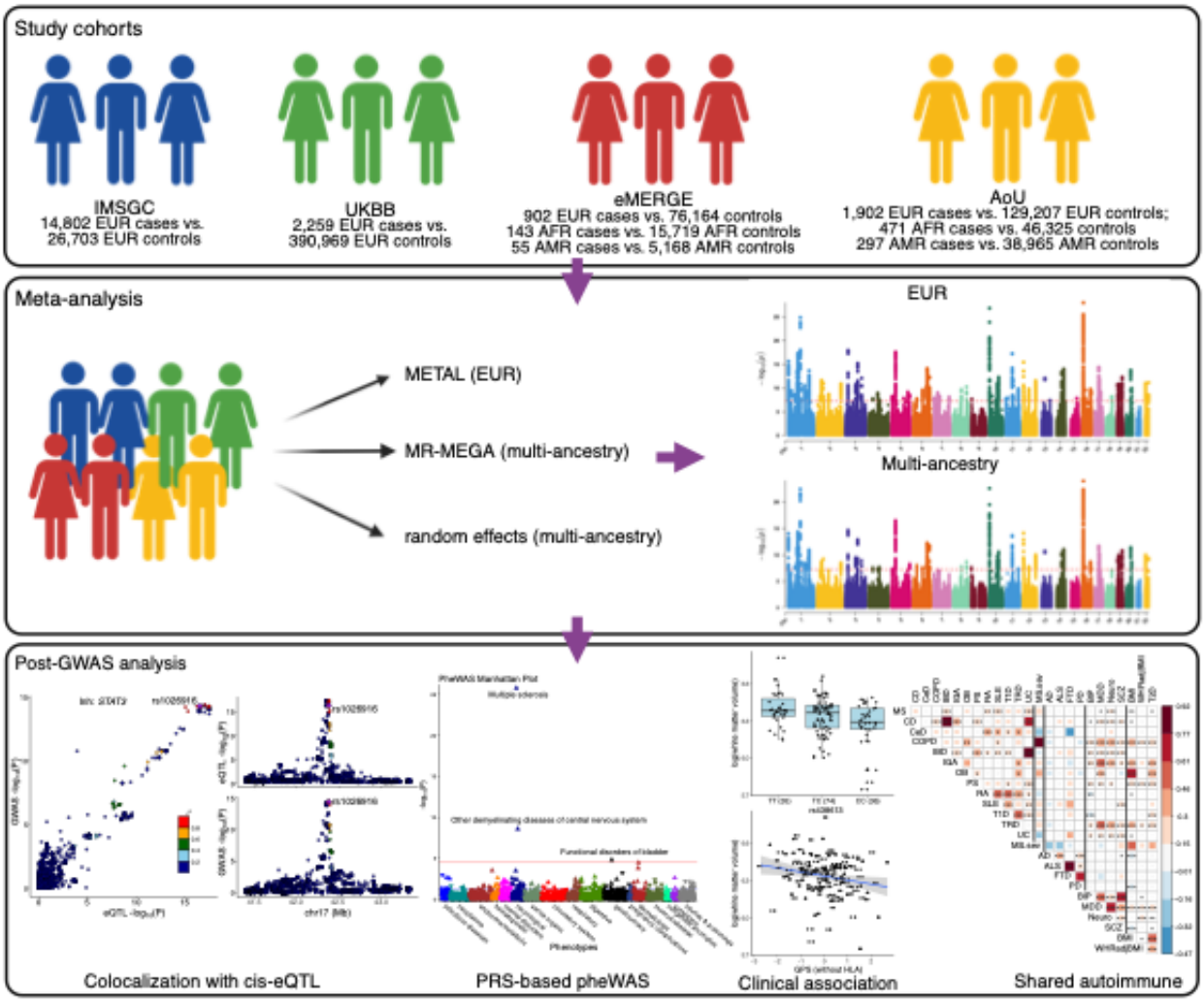

Study cohorts
IMSGC
14,802 EUR cases vs.
26,703 EUR controls
UKBB
2,259 EUR cases vs.
390,969 EUR controls
eMERGE
902 EUR cases vs. 76,164 controls
143 AFR cases vs. 15,719 AFR controls
55 AMR cases vs. 5,168 AMR controls
AoU
1,902 EUR cases vs. 129,207 EUR controls;
471 AFR cases vs. 46,325 controls
297 AMR cases vs. 38,965 AMR controls
Meta-analysis
METAL (EUR)
MR-MEGA (multi-ancestry)
random effects (multi-ancestry)
EUR
Multi-ancestry
Post-GWAS analysis
Colocalization with cis-eQTL
PRS-based pheWAS
Clinical association
Shared autoimmune

Figure 2.

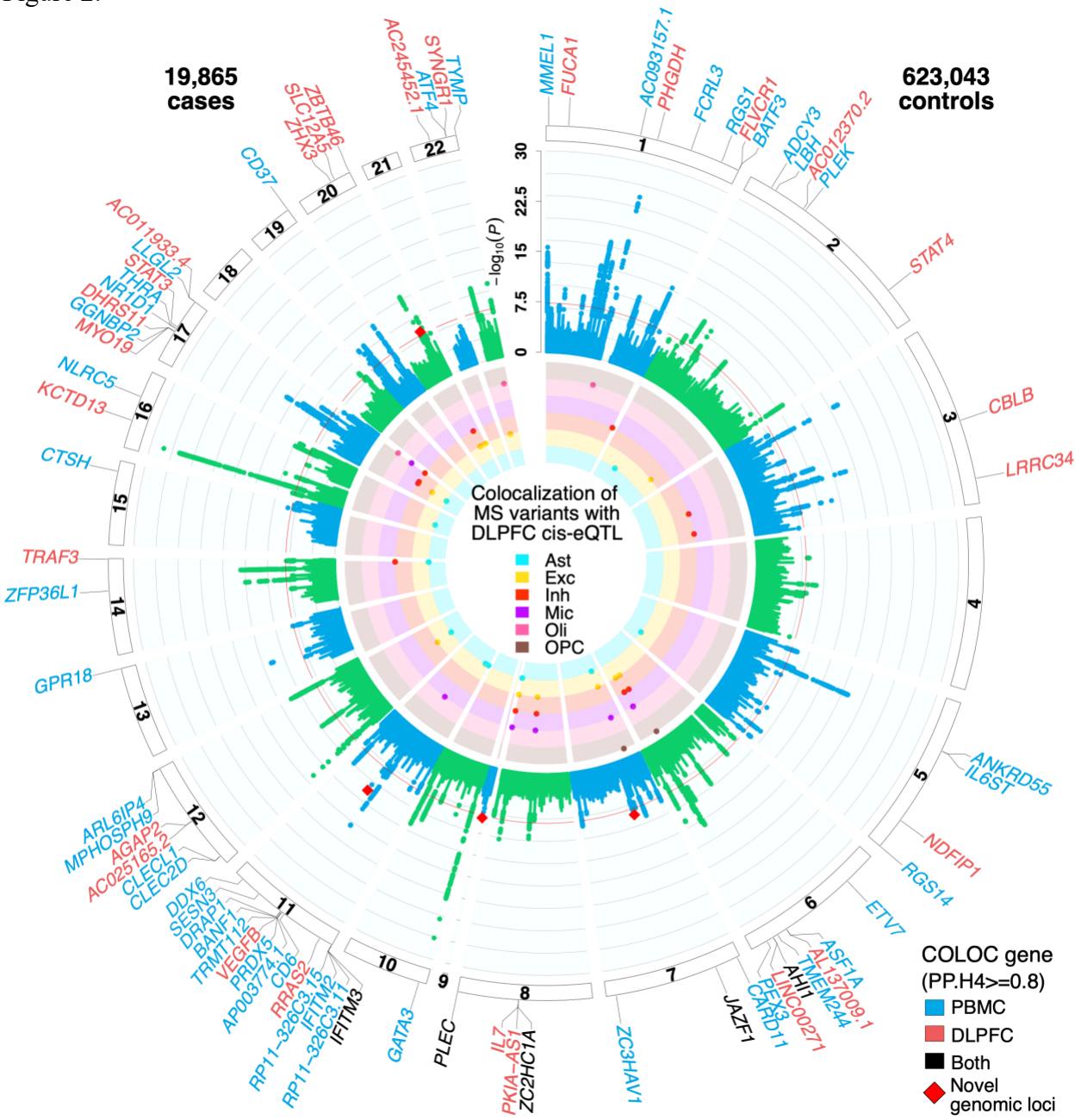

19,865
cases
623,043
controls
− log10(P)
0
7.5
15
22.5
30
Colocalization of
MS variants with
DLPFC cis-eQTL
Ast
Exc
Inh
Mic
Oli
OPC
COLOC gene
(PP.H4>=0.8)
PBMC
DLPFC
Both
Novel
genomic loci
MMEL1
FUCA1
AC093157.1
PHGDH
FCRL3
RGS1
FLVCR1
BATF3
ADCY3
LBH
AC012370.2
PLEK
STAT4
CBLB
LRRC34
ANKRD55
IL6ST
NDFIP1
RGS14
ETV7
ASF1A
AL137009.1
TMEM244
AHI1
LINC00271
PEX3
CARD11
JAZF1
ZC3HAV1
IL7
PKIA-AS1
ZC2HC1A
PLEC
GATA3
IFITM3
RP11-326C3.11
IFITM2
RP11-326C3.15
RRAS2
CD6
AP003774.1
PRDX5
VEGFB
TRMT112
BANF1
DRAP1
SESN3
DDX6
CLEC2D
CLEC1A
AC025165.1
AGAP2-AS1
MPHOSPH9
ARL6IP4
GPR18
ZFP36L1
TRAF3
CTSH
KCTD13
NLRC5
MYO19
GGNBP2
DHRS11
NR1D1
THRA
STAT3
LLGL2
AC011933.4
CD37
ZHX3
SLC12A5
ZBTB46
AC245452.1
ATF4
SYNGR1
TYMP

Figure 3.

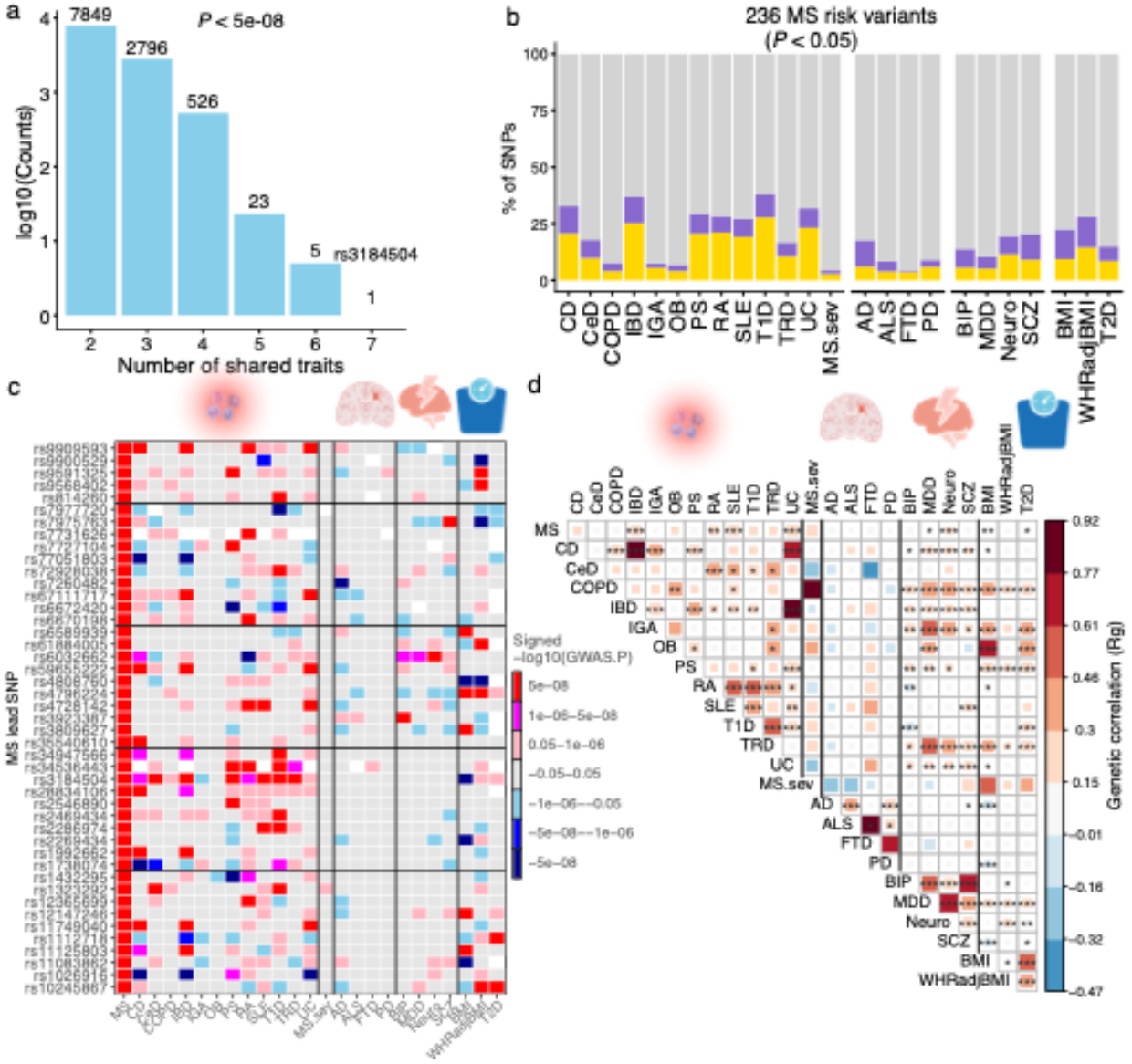

a
P < 5e-08
7849
2796
526
23
5
rs3184504
1
log10(Counts)
Number of shared traits
b
236 MS risk variants
(P < 0.05)
% of SNPs
CD
CeD
COPD
IBD
IGA
OB
PS
RA
SLE
T1D
TRD
UC
MS.sev
AD
ALS
FTD
PD
BIP
MDD
Neuro
SCZ
BMI
WHRadjBMI
T2D
c
MS lead SNP
Signed
−log10(GWAS.P)
5e−08
1e−06–5e−08
0.05–1e−06
−0.05–0.05
−1e−06–−0.05
−5e−08–−1e−06
−5e−08
d
MS
Genetic correlation (Rg)
0.92
0.77
0.61
0.46
0.3
0.15
−0.01
−0.16
−0.32
−0.47

Figure 4.

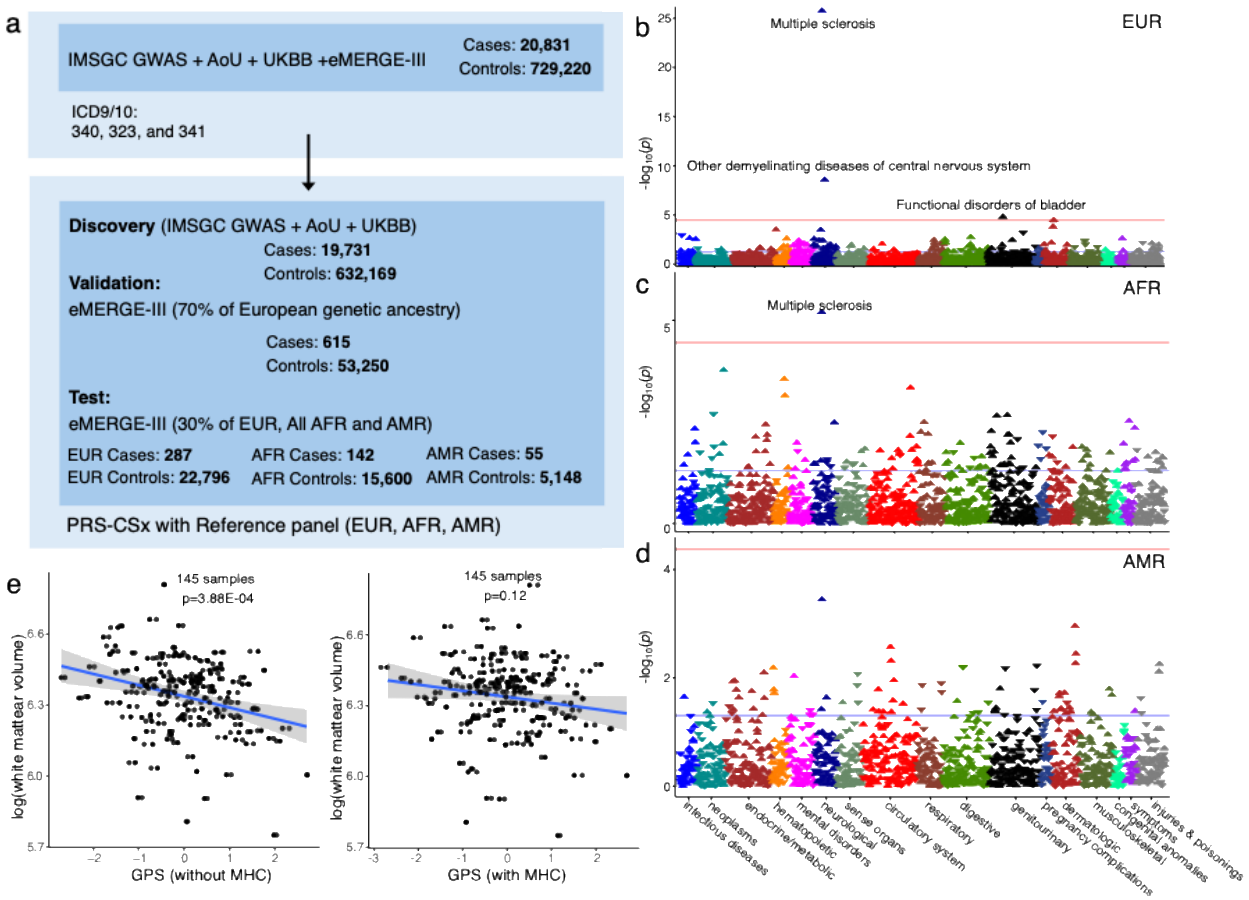
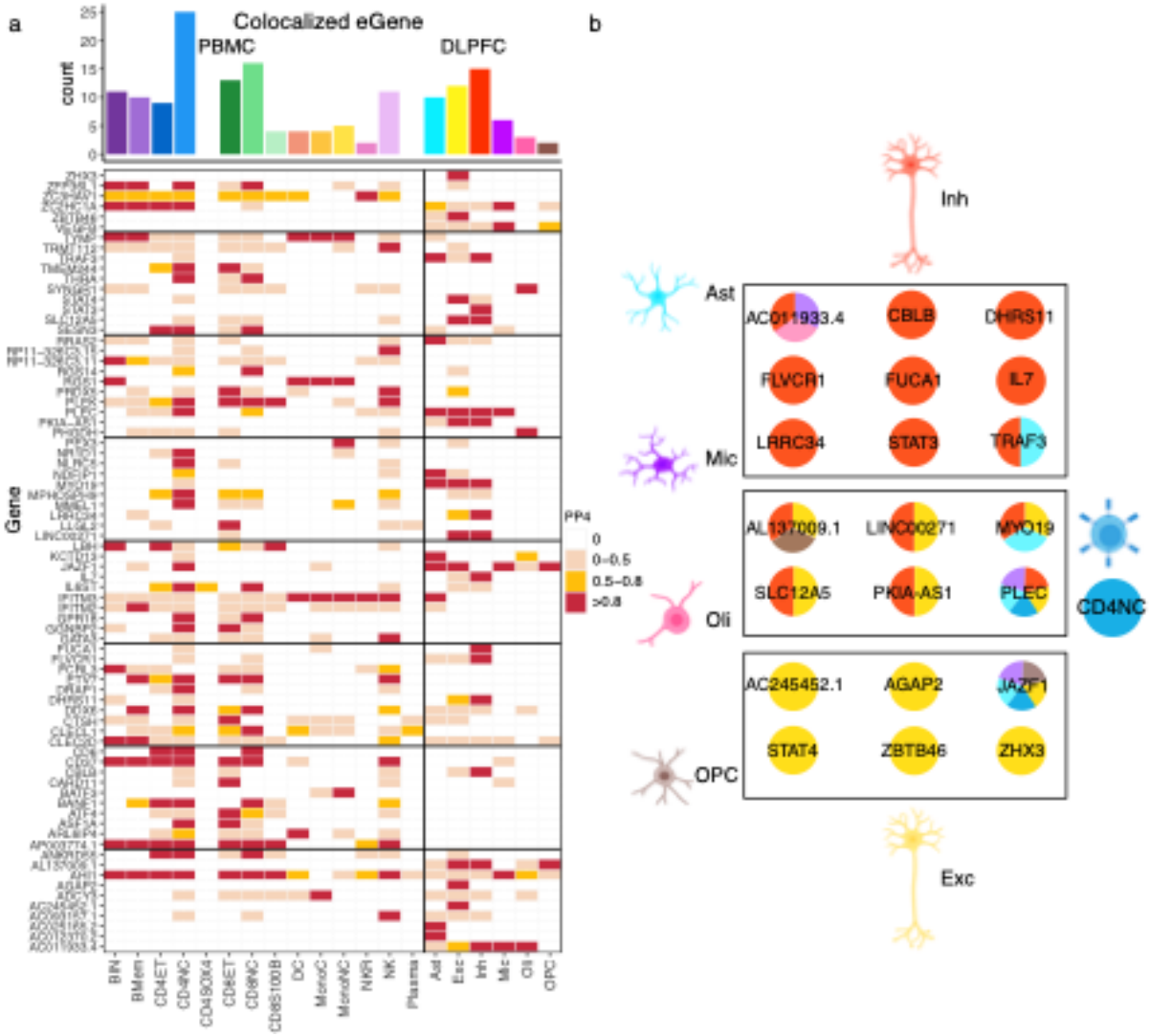
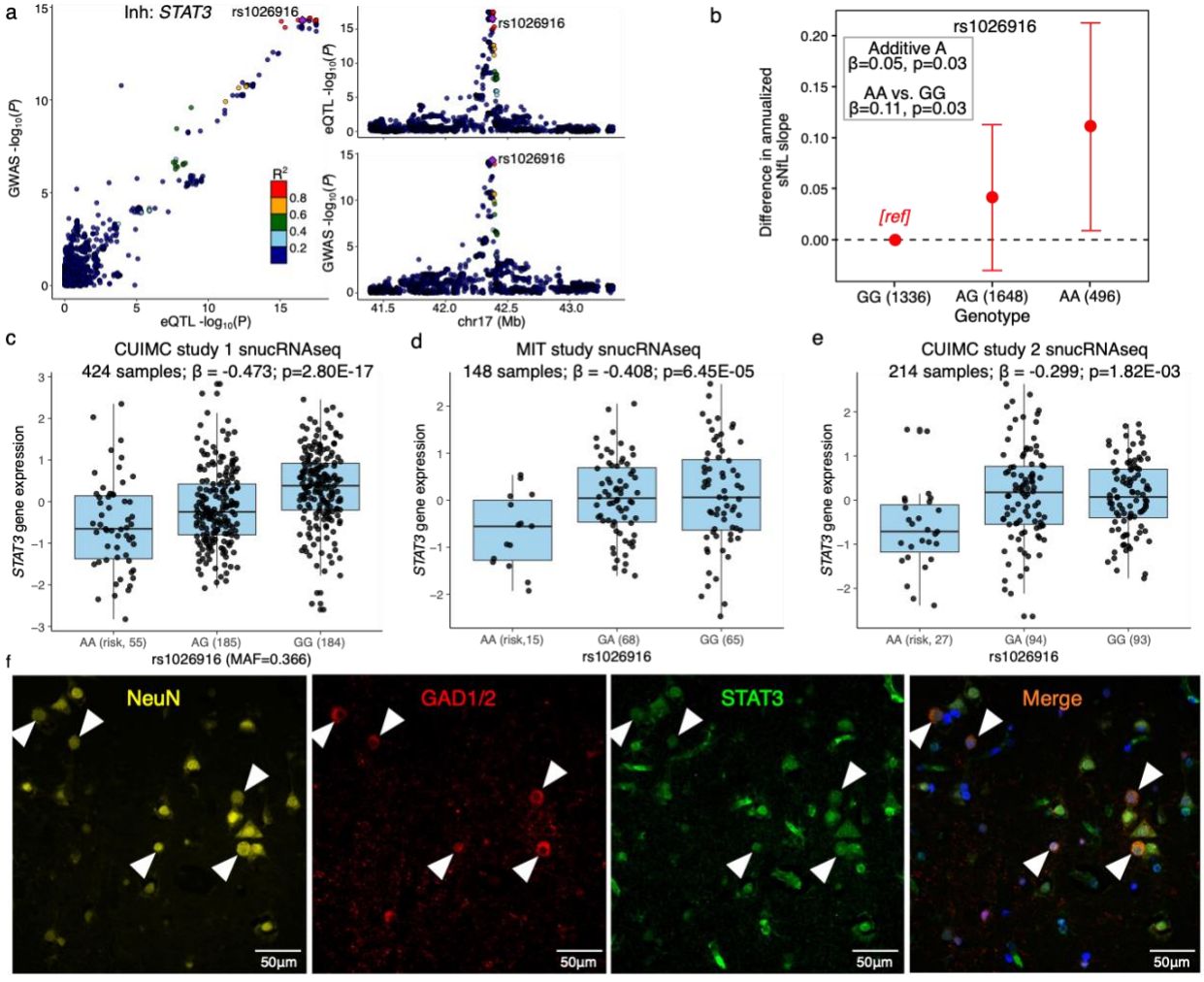

a
IMSGC GWAS + AoU + UKBB +eMERGE-III
Cases: 20,831
Controls: 729,220
ICD9/10:
340, 323, and 341
Discovery (IMSGC GWAS + AoU + UKBB)
Cases: 19,731
Controls: 632,169
Validation:
eMERGE-III (70% of European genetic ancestry)
Cases: 615
Controls: 53,250
Test:
eMERGE-III (30% of EUR, All AFR and AMR)
EUR Cases: 287
EUR Controls: 22,796
AFR Cases: 142
AFR Controls: 15,600
AMR Cases: 55
AMR Controls: 5,148
PRS-CSx with Reference panel (EUR, AFR, AMR)
b
Multiple sclerosis
EUR
Other demyelinating diseases of central nervous system
Functional disorders of bladder
c
Multiple sclerosis
AFR
d
AMR
e
145 samples
p=3.88E-04
145 samples
p=0.12
log(white mattear volume)
GPS (without MHC)
GPS (with MHC)

Figure 5.


a
Colocalized eGene
PBMC
DLPFC
count
Gene
PP4
0
0–0.5
0.5–0.8
>0.8
B IN
B Mem
CD4 ET
CD4 NC
CD4 SOX4
CD8 ET
CD8 NC
CD8 S100B
DC
Mono C
Mono NC
NK R
NK
Plasma
Ast
Exc
Inh
Mic
Oli
OPC
b
Inh
Ast
ACO11933.4
CBLB
DHRS11
FLVCR1
FUCA1
IL7
Mic
LRRC34
STAT3
TRAF3
AL137009.1
LINC00271
MYO19
Oli
SLC12A5
PKIA-AS1
PLEC
CD4NC
AC245452.1
AGAP2
JAZF1
OPC
STAT4
ZBTB46
ZHX3
Exc

1418 Figure 6.

1419



1420