Supplementary Information

Locus for severity implicates CNS resilience in progression of multiple sclerosis

International Multiple Sclerosis Genetics Consortium, MultipleMS Consortium

Supplementary Information

Table of Contents

Supplementary Note

Discovery and replication populations quality control

While the discovery population consisted exclusively of patients with MS, the replication population included MS cases and controls assembled as part of the MultipleMS consortium. To maximize subsequent phasing and imputation accuracy, all individuals were included in the quality control process. MS cases without disability measures and controls were excluded following imputation and did not contribute to the analyses (**Supplementary Fig. 1**).

Cohort-level quality control. Each cohort was mapped to the GRCh37 reference genome and oriented to the forward strand, using strand information from the Illumina array manifest files. Mitochondrial variants and insertions and deletions were excluded. We removed individuals with sample missingness greater than 0.05. We also removed individuals with a mismatch between genetic and reported sex, after selecting a linkage disequilibrium (LD) pruned (PLINK[1] --indep-pairphase 20000 2000 0.5) set of high-quality chromosome X variants with missingness < 0.02 and minor allele frequency (MAF) > 0.05. Following inspection of chromosome X F statistic histograms, thresholds for genetic sex determination were set to < 0.55 for females and > 0.8 for males. In the replication population, 4 out of 17 cohorts did not include sex chromosome data.

To account for the effects of population structure[2], we performed variant quality control using the subset of participants drawn from the largest ancestral group in each cohort. To define this group, we first selected high-confidence autosomal markers based on the following criteria:

- MAF > 0.05,
- Genotype missingness < 0.01,
- Hardy Weinberg equilibrium P > 10^{-10},
- Non-palindromic variants (excluding AT or CG variants),
- Retained following LD-pruned (PLINK2[3] --indep-pairwise 1000 kb 1 0.01),
- Outside regions with high principal component (PC) loadings[4],
- Common with 1000 Genomes phase 3[5].

The resulting set of variants was used to compute PC loadings from 2,534 unrelated (PLINK2 --king-cutoff 0.1) individuals from 1000 Genomes phase 3, onto which samples from each cohort were projected. Clustering was then performed on the PCs using the 'aberrant' package[6] in R, setting the parameter lambda to 30. The largest group comprised 85-100% of cohort participants closely matching the 1000 Genomes European superpopulation. The following variant filter criteria were then applied:

- Genotype missingness < 0.02 (0.05 in the replication cohort; threshold based on inspection of the empirical cumulative distribution function),
- Deviation from HWE P > 10^{-10} (10^{-6} for controls in the replication cohorts),
- MAF > 0.01,
- In the discovery only, between-cohort absolute allele frequency difference < 0.1 and absolute $\log_2$ fold-change < 5,
- In the replication only, differential missingness between cases and controls P > 10^{-4}.

For the sample quality control, we excluded individuals with an absolute inbreeding coefficient > 0.05, as well as those needed to resolve pairs of relatedness at the third degree or closer (PLINK2 --king-cutoff 0.0442). Finally, cohorts were merged into strata based on genotyping platforms (n=1 for discovery, n=4 for replication).

Stratum-level quality control. Stringent sample quality control was applied to each stratum. Cross-cohort duplicates and related samples were removed following the same approach as above. Population structure was assessed using a two-step approach. We applied PC analysis (PCA) as implemented in EIGENSOFT[7] to exclude outliers (> 6 standard deviations from stratum mean on PC1-10 across 5 iterations). Second, we calculated PCs on all 1000 Genomes phase 3 samples, projected our samples onto that space, and excluded outliers that fell 6 standard deviations away from the mean of the European 1000 Genomes superpopulation (N=503) on PC1-10. PCs after removal of outliers are presented in **Extended Data Fig. 2** and **Supplementary Fig. 2**.

Additional variant quality control was performed by applying the following criteria:

- Genotype missingness < 0.05,
- Deviation from HWE P > 10^{-10},
- For palindromic variants, alternate allele frequency < 0.4 or > 0.6,
- Absolute alternate allele frequency difference < 0.2 compared to European samples in the Haplotype Reference Consortium (HRC; version 1.1) imputation reference panel[8].

The number of variants and individuals passing quality control prior to imputation is described in **Supplementary Tables 3 and 4**.

Phasing and imputation. Phasing was performed using Eagle2 (version v2.4.1) in chunks of 20 Mb with 5 Mb overlapping flanking regions. The parameter for the number of conditioning haplotypes (--Kpbwt, default 10,000) was set to 20,000, with other parameters kept as default. Stratum samples were merged with HRC (n = 27,166) to maximize accuracy. Imputation of the phased genotypes was then performed using Minimac4[9]. We applied a range of analyses to examine chromosome continuity, imputation quality by chromosome position, and allele frequency difference compared to the reference panel. Across strata, the median imputation quality (R^2) was 0.965 to 0.985 for MAF ≥ 0.01 and the median imputation accuracy for genotyped variants (EmpR) was 0.977 to 0.997. Imputed variants with MAF < 0.01 or R^2 < 0.3 were excluded, resulting in a total of 7,722,279 to 7,830,995 autosomal variants for further analysis (**Supplementary Fig. 3** and **Supplementary Table 4**).

Sensitivity analysis of longitudinal disability outcomes

Given that the longitudinal analyses presented in **Figure 3** were performed in individuals that partially overlapped with the ARMSS score-based GWAS (N = 5,565), and that rs10191329 was selected based on its association in that GWAS, we conducted sensitivity analyses to ensure that this overlap did not bias the time-to-event and linear mixed model interaction results. First, we excluded all 5,565 EDSS scores (out of 54,113 in total, or 10%) that had been included in the cross-sectional GWAS. The A allele

for rs10191329 remained associated with 24-week confirmed disability worsening (HR = 1.12, 95% CI 1.03 to 1.20, P = 0.005), time to EDSS 6.0 (HR = 1.27, 95% CI 1.11 to 1.45, P = 0.0006) and rate of EDSS worsening (P = 0.01). Second, we examined if including participants from the initial GWAS could have caused bias beyond those overlapping EDSS scores, potentially due to correlation across EDSS scores. Here we leveraged the fact that a third of the participants in the longitudinal analyses (N = 2,760) were sourced from the replication cohort to compare the association of rs10191329 with each longitudinal outcome between participants who overlapped with the initial GWAS and those who did not. We introduced a variable indicating the cohort of origin (discovery or replication) and fitted a Cox regression with an interaction term between this indicator and rs10191329. The interaction term was not associated with 24-week confirmed disability worsening (*P* = 0.86) or time to EDSS 6.0 (*P* = 0.79), indicating that the association between rs10191329 and these outcomes is not different based on the original cohort. Similarly for the rate of EDSS worsening, a three-way interaction analysis indicated that the faster disability worsening observed among risk allele carriers was consistent across both cohorts (P = 0.35). Next, we analyzed the discovery and replication cohorts separately and tested for equivalence of the coefficients between cohorts using the formula:

$$Z = \frac{\beta_1 - \beta_2}{\sqrt{(SE\beta_1)^2 + (SE\beta_2)^2}}$$

where β is the regression coefficient and $SE\beta$ is the corresponding standard error (ref[10]). The results indicated that the differences in the coefficients for rs10191329 between the discovery and replication cohorts were not different from zero for 24-week confirmed disability worsening (*P* = 0.73), time to EDSS 6 (*P* = 0.78), and rate of EDSS change (*P* = 0.36).

Lastly, we restricted the analysis to the replication cohort assuming a directionally concordant effect. Despite the considerably reduced power, rs10191329 remained associated with each outcome: 24-week confirmed disability worsening HR = 1.13, P = 0.033; time to EDSS 6.0 HR = 1.24, P = 0.026; rate of EDSS worsening β = 0.02, P = 0.009.

In summary, while two-thirds of the participants in the analyses of longitudinal disability outcomes overlapped with the initial GWAS, only 10% of the observations were shared. The exclusion of these EDSS scores did not impact the results. Moreover, associations derived from overlapping and non-overlapping individuals did not differ in direction or magnitude, and the findings were retained when considering only individuals from the replication cohorts. Consequently, we conclude that these analyses were not biased by this overlap.

Differential expression of prioritized genes in MS lesions and control tissues

We analyzed the expression of the prioritized genes identified from our MS severity GWAS in two single-nucleus RNA sequencing datasets from MS patients and controls. The first examined 17,799 nuclei from post-mortem white matter tissue samples collected from four individuals with progressive MS and five controls. In the MS samples, nuclei were isolated from four different white matter areas: normal appearing white matter (N = 3), active (N = 3), chronic active (N = 4), chronic inactive edge (N = 3), and remyelinated

(N = 2) lesions[11]. The second study examined 66,432 nuclei from three controls and five individuals with progressive MS (sampling two lesion cores, and four each of periplaque white matter tissue, chronic active and chronic inactive lesion edges)[12]. Processed unique molecular identifier counts were obtained from the original publications. Using a Bonferroni-corrected Wilcoxon Rank Sum two-tailed test, we compared the expression of the four prioritized severity genes in MS lesion types and white matter (normal-appearing and periplaque) to their respective control tissue from each study. This analysis revealed significant differential expression of all four genes (*DYSF*, *ZNF638*, *DNM3*, *PIGC*; **Supplementary Fig. 6**), with normal-appearing and periplaque white matter exhibiting the least difference (range of $\log_2$ fold-change between -1 and 1).

Exploratory analysis of genomics-driven drug discovery

We adopted a dual strategy to identify drug compounds and classes that could potentially reduce the severity of MS and slow its progression. First, we leveraged the scalable precision medicine open knowledge engine (SPOKE)[13] to search for approved or experimental compounds that regulate the expression of genes prioritized from our GWAS, or that bind to the proteins encoded by these genes (this knowledge graph integrates information from 41 biomedical databases). The regulatory effects of compounds on genes were obtained from perturbation experiments conducted using the L1000 assay platform as part of the NIH Library of Integrated Network-Based Cellular Signatures (LINCS) initiative[14]. Protein binding data was sourced from ChEMBL[15] and BindingDB[16]. We further screened for additional compounds by querying DrugBank[17], Therapeutic Target Database[18], PharmGKB[19], and the Open Targets Platform[20]. After controlling for the number of SPOKE edges between compounds and regulated genes (N = 968,460), two of our prioritized genes located in different loci, *DYSF* and *DNM3*, were found to be regulated by the same compound, entinostat, a small molecule histone deacetylase (HDAC) inhibitor. No additional compounds were identified in the other data sources.

Second, considering the limitations of single GWAS locus-based drug discovery[21], we applied Trans-Phar[22] to screen for compounds with gene expression regulatory profiles that negatively correlate with genetically determined expression associated with MS severity, with the notion that reversing the effects of the genetic variation could have therapeutic potential[21]. In brief, genetically regulated gene expression for MS severity was inferred in GTEx tissues (version 7) using FOCUS[23]. Genes with the top 10% absolute z-score were selected and compared against compound-perturbed gene expression in tissue-matched cell lines from the LINCS L1000 library[14] using Spearman's rank correlation tests. Given the prior evidence that significant associations are enriched in disease-relevant tissue[22], we initially focused our analysis on CNS tissues in GTEx (N = 10) and CNS cell lines in LINCS (N = 5; MNEU.E, NEU, NPC, NPC.CAS9, NPC.TAK), resulting in 82,680 correlation tests for pairs of genetically determined and compound-perturbed gene expression profiles. At an FDR of 0.1 (as recommended by the method developers[24]), we identified a significant negative correlation for the compound pracinostat, another small molecule HDAC inhibitor, in neural progenitor cells. Moreover, we observed apparent inflation in test statistics in CNS cell-type-compound pairs (**Supplementary Figure 7**). To verify that this was not the case in other tissues, we extended the analysis to additional GTEx tissues (N = 19) and LINCS cell types (N = 72) outside the CNS, and found no evidence of inflation (**Supplementary Figure 7**). A CNS

enrichment was further supported by significantly higher rankings for cell-type-compound pairs in the CNS compared to non-CNS tissues (Wilcoxon rank-sum test $P = 3.1\times10^{-4}$).

While currently approved MS therapies primarily target the peripheral immune system and offer limited or no benefit for progression, our analysis revealed significant signal enrichment for CNS-acting compounds, supporting the rationale for pursuing neuroprotection as a therapeutic strategy to slow MS progression. Additionally, using separate approaches we identified two small-molecule compounds, entinostat and pracinostat, belonging to the HDAC inhibitors class. HDACs regulate chromatin remodeling and are dysregulated in neurodegenerative diseases[25]. Although mainly indicated for cancer treatment and despite their lack of target specificity, HDAC inhibitors have demonstrated neuroprotective effects in cell and animal models[25,26], and have shown benefits in preclinical models of MS[27–30]. Lastly, a low frequency missense variant in *HDAC7* has also been associated with MS risk[31].

MS severity polygenic score and association with neurological diseases in UK Biobank

To estimate the aggregate impact of genetic factors that determine MS severity on other brain-related phenotypes, we constructed a polygenic score (PGS) for MS severity and evaluated its association with nervous system disorders in the UK Biobank (application number 61342). We first estimated the expected predictive power $E(R^2)$ of this PGS in relation to our modest GWAS sample size and calculated heritability, using the formula[32,33]:

$$E(R^2) = \frac{(h^2_{SNP})^2}{h^2_{SNP} + \frac{M}{N}}$$

where h^2_{SNP} is the SNP heritability, M is the effective number of independent markers, and N is the GWAS sample size. Based on previous estimates we set M to be 60,000[34]. This formula has been shown to provide a strong correlation between estimated and observed R^2 in a set of 26 phenotypes tested in the UK biobank[32]. Using a 95% confidence interval around our GRM heritability estimate, the $E(R^2)$ for an MS severity PGS based on our GWAS was 0.00035 to 0.0051. We then constructed a PGS using PRSice-2 (with default settings)[35]. The best model fit incorporated 143,368 independent variants with $P < 0.36$ and demonstrated an R^2 of 0.001 ($P = 4\times10^{-3}$) in the replication cohort, consistent with the above theoretical predictions. In comparison, the variance explained by the top GWAS variant (rs10191329) in the replication population was 5.0×10^{-4}.

Next, we selected 408,817 individuals of white British ancestry as defined in the UK Biobank[2] and extracted case status for 67 conditions falling under Chapter VI Diseases of the nervous system of the ICD-10 (UK Biobank Category 2406). Cases were sourced from primary care and hospital inpatient data, death registry records, and self-reports. Data fields with fewer than 275 cases (first quartile) were excluded. The median number of cases for the 50 remaining conditions was 1,462 with a range of 276 to 26,417. Mean age was 56.9 years (standard deviation 8.0) and 45.9% of participants were male. Individuals were scored for the MS severity PGS using PLINK2 and 142,985 (99.7%) overlapping variants. Logistic regression adjusted for age, sex, and the first 10 PCs (as provided by the UK Biobank) showed no association between the MS severity PGS and neurological phenotypes (FDR ≤ 0.43). To

summarize, the MS severity PGS demonstrated limited predictive capacity in accordance with the GWAS sample size and heritability, and was specific to MS outcomes when evaluated against a range of case-control neurological disorders in a populational biobank.

Observational cohorts and variables in educational attainment analysis

The Epidemiological Investigation of Multiple Sclerosis (EIMS) study recruited incident MS cases aged 16 to 70 years from 42 neurological clinics in Sweden from April 2005 to June 2011. The Genes and Environment in MS (GEMS) study comprised prevalent cases identified from the Swedish MS registry between November 2009 and November 2011[36,37]. Participants in EIMS and GEMS provided written informed consent and ethical approval was obtained from the Regional Ethical Review Board in Stockholm. **Supplementary Table 20** reports the characteristics of MS participants included in the present study from each cohort.

Clinical characteristics, including age at onset, disease course, and disability level measured by the EDSS were extracted from the Swedish MS registry[37]. EDSS scores were converted to rank inverse normal transformed age-related MS severity score (ARMSS) as in the GWAS. Data on education and income was extracted from Statistics Sweden. Using the Swedish Educational Terminology (SUN 2000), the highest educational level commenced by individuals was coded as years of study (**Supplementary Table 24**). Data on education outside Sweden was not available. Individualized disposable income was characterized by five ordinal categories based on data from 27,644 MS patients and 285,676 controls matched by age, sex, and residential area. Smoking was assessed through questionnaires and individuals were categorized as ever (past and current) or never smokers. Additional covariates including year of birth and sex were obtained using government-issued personal identification numbers. Individuals with missing data were excluded from the analysis.

CNS cell-type heritability enrichment

In light of our finding of enrichment for MS severity associations in genes with high relative expression in CNS tissue, we extended our analysis to individual cell types following the approach by Bryois et al.[38] We collected publicly available single-nucleus RNA sequencing data from two different studies. The first sampled 19,550 nuclei from human prefrontal cortex and the hippocampus[39]; the second examined 36,166 nuclei from human frontal and visual cortex, and the cerebellum[40]. We replicated the processing steps outlined in Bryois et al.[38] by adapting code provided with the paper. Briefly, non-protein-coding genes and those not expressed in any cell type were excluded. After scaling, cell type specificity was quantified by dividing the expression level of each gene by its total expression across cell types. The top 10% most specific genes for each cell type were then selected, followed by annotation of 100-kb genomic regions upstream and downstream of these genes. Analysis was performed using stratified LDSC adjusted for the baseline model[41]. We found no significant cell type enrichment after multiple testing correction for the number of cell types (FDR > 0.3, **Supplementary Fig. 10**). As demonstrated in a recent comparative analysis of previous and current GWAS of schizophrenia[42], statistically significant cell type-specific enrichment may emerge for MS severity as the sample size increases in future iterations.

Relative contribution of relapses and progression to MS severity

In principle, relapses and progression could both influence the MS severity outcomes used in this study. However, although relapses typically lead to transient increase in disability, it is recognized that their contribution to long-term disability and confirmed disability progression is limited, especially after the first few years post diagnosis[43]. In addition, relapse frequency spontaneously diminishes over time[44]. In this context, a recent study of relapse activity in MS reported distinct genetic association signals and alternate pathways[45]. Given the average age and disease duration of our population (respectively 51.7 and 18.2 years), as well as the associations with time to EDSS 6.0 and cortical pathology, our findings are likely to reflect independent mechanisms underlying MS progression. Nevertheless, additional longitudinal analyses will be required to further refine the effects of these severity variants on MS phenotypes, including molecular and imaging.

Consortia Members

All authors are listed at the end of the main manuscript. In addition to those authors, these lists recognize the involvement of consortium members who did not directly contribute to the paper. Individuals may be part of one or both consortia.

Members of the International Multiple Sclerosis Genetics Consortium

Lars Alfredsson[1], Katayoun Alikhani[2], Till F. M. Andlauer[3], Maria Ban[4], Sergio E. Baranzini[5], Lisa F. Barcellos[6], Nadia Barizzone[7], Ashley H. Beecham[8], Tone Berge[9], Luisa Bernardinelli[10], Achim Berthele[3], Stefan Bittner[11], Steffan D. Bos[12], Farren Briggs[13], Stacy J. Caillier[5], Domenico Caputo[14], Paola Cavalla[15], Elisabeth G. Celius[12], Gabriel Cerono[5], Tanuja Chitnis[16], Ferdinando Clarelli[17], Manuel Comabella[18], Giancarlo Comi[19], Alastair Compston[4], Chris Cotsapas[20], Bruce C. A. Cree[5], Sandra D'Alfonso[7], Efthimios Dardiotis[21], Philip L. De Jager[22], Silvia R. Delgado[23], Bénédicte Dubois[24], Sinah Engel[11], Hendrik J. Engelenburg[25], Federica Esposito[26], Marzena J. Fabis-Pedrini[27], Massimo Filippi[28], Christiane Gasperi[3], Lissette Gomez[8], Refujia Gomez[5], An Goris[24], Pierre-Antoine Gourraud[29], Georgios Hadjigeorgiou[30], David A. Hafler[31], Jonathan L. Haines[32], Jörg Hamann[25], Hanne F. Harbo[12], Adil Harroud[5], Stephen L. Hauser[5], Friederike Held[3], Bernhard Hemmer[3], Roland G. Henry[5], Jan Hillert[1], Dana Horakova[33], Jesse Huang[1], Inge Huitinga[25], Noriko Isobe[34], Maja Jagodic[1], Seema Kalra[35], Allan G. Kermode[27], Michael Khalil[36], Trevor J. Kilpatrick[37], Ingrid Kockum[1], Ioanna Konidari[8], Karim L. Kreft[38], Jeannette Lechner-Scott[39], Maurizio Leone[40], Felix Luessi[11], Sunny Malhotra[18], Ali Manouchehrinia[1], Clara P. Manrique[8], Roland Martin[41], Filippo Martinelli-Boneschi[42], Elisabetta Mascia[17], Jacob L. McCauley[8], Luanne M. Metz[2], Luciana Midaglia[18], Xavier Montalban[18], Jorge R. Oksenberg[5], Tomas Olsson[1], Annette Oturai[43], Kimmo Pääkkönen[44], Grant P. Parnell[45], Nikolaos A. Patsopoulos[46], Margaret A. Pericak-Vance[8], Fredrik Piehl[1], Justin P. Rubio[37], Janna Saarela[44], Adam Santaniello[5], Silvia Santoro[17], Stephen J. Sawcer[4], Catherine Schaefer[47], Finn Sellebjerg[43], Hengameh Shams[5], Klementy Shchetynsky[48], Claudia Silva[2], Vasileios Siokas[21], Joost Smolders[49,49], Helle B. Søndergaard[43], Melissa Sorosina[17], Graeme J. Stewart[50], Pernilla Stridh[1], Bruce Taylor[51], Aletta M. R. van den Bosch[25], Marijne Vandebergh[24], Domizia Vecchio[52], Margarete M. Voortman[36], Howard L. Weiner[16], V. Wee Yong[2], Frauke Zipp[11].

[1]Department of Clinical Neuroscience, Karolinska Institutet, Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden. [2]Department of Clinical Neurosciences and the Hotchkiss Brain Institute, University of Calgary, Calgary, Canada. [3]Department of Neurology, School of Medicine, Technical University of Munich, Munich, Germany. [4]Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK. [5]UCSF Weill Institute for Neurosciences, Department of Neurology, University of California, San Francisco, San Francisco, CA, USA. [6]Genetic Epidemiology and Genomics Laboratory, Division of Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, CA, USA. [7]Department of Health Sciences and Center on Auto-immune and Allergic Diseases (CAAD), University of Eastern Piedmont, Novara, Italy. [8]John P Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA. [9]Department of Research, Innovation and Education, Oslo University Hospital, Oslo, Norway. [10]Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy. [11]Department of Neurology, Focus Program Translational Neuroscience (FTN) and Immunotherapy (FZI), University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany. [12]Department of Neurology, Oslo University Hospital, Oslo, Norway. [13]Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University,

Cleveland, OH, USA. [14]IRCCS Fondazione Don Gnocchi ONLUS, Milano, Italy. [15]Department Neuroscience and Mental Health, City of Health and Science University Hospital of Turín, Turín, Italy. [16]Brigham Multiple Sclerosis Center, Brigham and Women's Hospital, Boston, MA, USA. [17]Laboratory of Human Genetics of Neurological Disorders, IRCCS San Raffaele Scientific Institute, Milan, Italy. [18]Servei de Neurologia-Neuroimmunologia, Centre d'Esclerosi Múltiple de Catalunya (Cemcat), Vall d'Hebron Institut de Recerca, Vall d'Hebron Hospital Universitari, Barcelona, Spain. [19]Vita-Salute San Raffaele University, Milan, Italy. [20]Departments of Neurology and Genetics, Yale School of Medicine, New Haven, CT, USA. [21]Department of Neurology, University General Hospital of Larissa, Faculty of Medicine, School of Health Sciences, University of Thessaly, Larissa, Greece. [22]Center For Translational & Computational Neuroimmunology and the Multiple Sclerosis Center, Department of Neurology, Columbia University Irving Medical Center, New York, NY, USA. [23]Multiple Sclerosis Division, Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL, USA. [24]KU Leuven, Leuven Brain Institute, Department of Neurosciences, Leuven, Belgium. [25]Neuroimmunology Research Group, Netherlands Institute for Neuroscience, Amsterdam, Netherlands. [26]Neurology Unit and Laboratory of Human Genetics of Neurological Disorders, IRCCS San Raffaele Scientific Institute, Milan, Italy. [27]Perron Institute for Neurological and Translational Science, University of Western Australia, Perth, Australia. [28]Neurology Unit, Neurorehabilitation Unit, Neurophysiology Service and Neuroimaging Research Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy. [29]Université de Nantes, CHU Nantes, Inserm, Centre de Recherche en Transplantation et Immunologie, Nantes, France. [30]Medical School, University of Cyprus, Nicosia, Cyprus. [31]Departments of Neurology and Immunobiology, Yale School of Medicine, New Haven, CT, USA. [32]Case Western Reserve University School of Medicine, Cleveland, Ohio, USA. [33]Department of Neurology and Center of Clinical Neuroscience, First Faculty of Medicine and General University Hospital, Charles University in Prague, Prague, Czech Republic. [34]Department of Neurology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. [35]Royal Stoke MS Centre, University Hospitals of North Midlands NHS Trust, Royal Stoke University Hospital, Newcastle Road, Stoke-on-Trent, UK. [36]Department of Neurology, Medical University of Graz, Graz, Austria. [37]Florey institute of Neuroscience and Mental Health, Melbourne, Australia. [38]Department of Neurology, MS center ErasMS, Erasmus University Medical Center, Rotterdam, Netherlands. [39]Department of Neurology, John Hunter Hospital, Hunter New England Health District, Newcastle, Australia. [40]SC Neurologia, Dipartimento di Scienze Mediche, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Foggia, Italy. [41]Neuroimmunology and Multiple Sclerosis Research Section, Neurology Clinic, University Hospital Zurich & University of Zurich, Zurich, Switzerland. [42]IRCCS Fondazione Ca' Granda Ospedale Maggiore Policlinico, Neurology Unit, Milan, Italy. [43]Danish Multiple Sclerosis Center, Department of Neurology, Copenhagen University Hospital - Rigshospitalet, Glostrup, Denmark. [44]Institute for Molecular Medicine Finland, Helsinki Institute for Life Sciences, University of Helsinki, Helsinki, Finland. [45]School of Medical Sciences, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia. [46]Systems Biology and Computer Science Program, Ann Romney Center for Neurological Diseases, Department of Neurology, Brigham & Women's Hospital, Boston, MA, USA. [47]Kaiser Permanente Division of Research, Oakland, CA, USA. [48]Department of Neurology, Yale School of Medicine, New Haven, CT, USA. [49]Departments of Neurology and Immunology, MS center ErasMS, Erasmus University Medical Center, Rotterdam, Netherlands. [50]Westmead Institute for Medical Research, Sydney, Australia. [51]Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia. [52]Department of Translational Medicine and Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara, Italy.

Members of the MultipleMS Consortium

Ana-Katharina Klein[1], Imad Abugessaisa[2], Amber Alexander[3], Jeppe Sonne Andersen[4], Till F. M. Andlauer[1], Jennifer Arjona[3], Amie Baker[5], Maria Ban[5], Sergio E. Baranzini[3], Stephan Beck[6], Pål Berg-Hansen[7], Tone Berge[7], Anna Berntson[8], Mona Beyer[7], Viola Biberacher[1], Yolanda Blanco[9], Anna Blazier[10], Steffan D. Bos[7], Rune Brautaset[11], Sophie Buhelt [4], Cecilia Byström[11], Lynn Castelot[12], Elisabeth G. Celius[7], Ambra Cerri[13], Wojciech Chacholski[8], Clement Chatelain[10], Ferdinando Clarelli[14], Noah Connally[12], Andrea Corona[15], Chris Cotsapas[12], Sandra D'Alfonso[16], Amy Dao[12], Kati Donner[17], Bénédicte Dubois[18], Simone Ecker[6], Anita Eliasson[19], Federica Esposito[14], Begoña Fernandez[9], Mika Frederiksen[17], Lars Fugger[20], Guro Gafvelin[11], Christiane Gasperi[1], David Gomez-Cabrero[2], Javier Gonzalo[9], An Goris[18], Tobias Granberg[11], Kim Grasboeck[1], Courtney Grimm[12], Hans Grönlund[11], Verena Grummel[1], Clara Guaschino[14], Maria Gustafsson[8], Stefan Gustavsen[4], Alexandra Gyllenberg[11], Hanne F. Harbo[7], Peter Harrison[21], Adil Harroud[3], Bernhard Hemmer[1], Roland Henry[3], Sari Honkanen[17], Jesse Huang[11], Maja Jagodic[11], Sophie Janacek[21], Jean Jones[3], Ingileif Jónsdóttir[22], Seema Kalra[23], Timo Kanninen[19], Rajinder Kaul[12], Tobias Keiner[1], Anisha Keshavan[3], Narsis Kiani[2], Kicheol Kim[3], Gina Kirkish[3], Benedikt Knier[1], Ingrid Kockum[11], Myrto Kostadima[21], Adrian Lang[1], Annika Langkilde[4], Ferre Laura[14], Maurizio Leone[24], Gildas Lepentier[1], Tony Lewis[5], Cecilia Lindgren[25], Eva Lindström[11], Sara Llufriu[9], Lohith Madireddy[3], Ali Manouchehrinia[11], Filippo Martinelli-Boneschi[15], Elena H Martinez-Lapiscina[9], Meoni Massimiliano[14], Diana Maxwell[8], Timo Miettinen[17], Thomas Moridi[11], Alisha Mota[12], Sinéad Moylett[18], Mark Mühlau[1], Shaun Muzic[3], Maria Needhamsen[11], Maria Nilsson[11], Mathieu Niquille[26], Dimitry Ofengeim[10], Anna Gabriella Olsson[4], Tomas Olsson[11], Ana Maria Osiceanu[14], Annette Oturai[4], Kimmo Pääkkönen[17], Staffan Paulie[27], Fredrik Piehl[11], Laura Planas[9], Albert Pukaj[1], Renate Raedlein[1], Ryan Ramanujam[8], Marianne Randen[7], Philippe Rocolle[10], Philip Rosensteil[28], Janna Saarela[17], Albert Saiz[9], Silvia Santoro[14,14], Stephen Sawcer[5], Finn Sellebjerg[4], Klementy Shchetynsky[12], Parisa Shooshtari[12], Gilad Silberberg[2], Emma Sinha[21], Helle B. Søndergaard[4], Melissa Sorosina[14], Ingvild Sorum[7], Kári Stefánsson[22], Pernilla Stridh[11], Miljana Tanic[6], Jesper Tegnér[2], Päivi Tikka-Kleemola[19], Lori Kim Tonnes-Priddy[12], Anusha Uppinkudru[1], Gemma Vila[9], Pablo Villoslada[9], Amy Webster[6], Martina Wenzel[1], Mathew Woodwark[29], Andrea Zauli[14], Daniel Zerbino[21], Mindy Zhang[10], Galina Zheleznyakova[11], Nikolay Zhukovsky[26], Iaroslav Zinevych[1], Frauke Zipp[30], Irati Zubizarreta[9].

[1]Department of Neurology, School of Medicine, Technical University of Munich, Munich, Germany. [2]Karolinska Institutet, Stockholm, Sweden. [3]UCSF Weill Institute for Neurosciences, Department of Neurology, University of California, San Francisco, San Francisco, CA, USA. [4]Danish Multiple Sclerosis Center, Department of Neurology, Copenhagen University Hospital, Rigshospitalet, Glostrup, Denmark. [5]Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK. [6]University College London, London, UK. [7]Department of Neurology, Oslo University Hospital, Oslo, Norway. [8]KTH-The Royal Institute of Technology, Stockholm, Sweden. [9]Department of Neurology, Hospital Clinic Barcelona and Institut d'Investigacions Biomediques August Pi Sunyer (IDIBAPS), Barcelona, Spain. [10]SANOFI, Chilly-Mazarin, France. [11]Department of Clinical Neuroscience, Karolinska Institutet, Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden. [12]Altius Institute for Biomedical Sciences, Seattle, WA, USA. [13]IRCCS Policlinico San Donato. [14]Neurology Unit and Laboratory of Human Genetics of Neurological Disorders, IRCCS San Raffaele Scientific Institute, Milan, Italy. [15]IRCCS Fondazione Ca' Granda Ospedale Maggiore Policlinico, Neurology Unit, Milan, Italy. [16]University of Eastern Piedmont, Novara, Italy. [17]Institute for Molecular Medicine Finland, Helsinki Institute for Life Sciences, University of Helsinki, Helsinki, Finland. [18]KU Leuven, Leuven Brain Institute, Department of Neurosciences, Leuven,

Belgium. [19]BC Platforms, Zurich, Switzerland. [20]Oxford Centre for Neuroinflammation, Nuffield Department of Clinical Neurosciences, Oxford University Hospitals, University of Oxford, Oxford, UK. [21]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Cambridge, UK. [22]deCODE Genetics/Amgen, Inc., Reykjavik, Iceland. [23]Royal Stoke MS Centre, University Hospitals of North Midlands NHS Trust, Royal Stoke University Hospital, Newcastle Road, Stoke-on-Trent, UK. [24]Fondazione IRCCS Casa Sollievo Della Sofferenza, San Giovanni Rotondo, Italy. [25]Big Data Institute, Li Ka Shing Center for Health Information and Discovery, Oxford University, Oxford, UK. [26]NEURIX, Geneva, Switzerland. [27]Mabtech AB, Nacka Strand, Sweden. [28]Institute of Clinical Molecular Biology, Kiel University and University Medical Center Schleswig-Holstein, Campus Kiel, Kiel, Germany. [29]MedImmune, Gaithersburg, MD, USA. [30]Department of Neurology, Focus Program Translational Neuroscience (FTN) and Immunotherapy (FZI), University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany.

Data Access Restrictions

Individual-level genetic and phenotype data deposited in the European Genome-phenome Archive (EGAS00001007162) is organized by contributing center to comply with varying consent and privacy laws. The following datasets may be shared with bone fide third parties for research approved by the International Multiple Sclerosis Consortium (IMSGC) data access committee: Austria (EGAD00010002487), Greece (EGAD00010002493), Italy-Milan (EGAD00010002486), Italy-Piedmont (EGAD00010002490), Netherlands (EGAD00010002488), Norway (EGAD00010002485), United Kingdom (EGAD00010002491), and Spain (EGAD00010002492). The datasets from Germany-Mainz (EGAD00010002484) and Germany-Munich (EGAD00010002489) may be shared with not-for-profit third parties for research approved by the IMSGC. For non-European centers, Individual-level genetic and phenotype data is available in dbGaP as a single dataset (phs002929.v1.p1). Participants are grouped into one of four consent categories: general research use (N = 3,503), health/medical/biomedical research excluding the study of population origins or ancestry (N = 648), disease-specific research on MS (N = 291); disease-specific research on MS requiring collaboration with the primary study investigator (N = 181).

Ethical approvals for the GWAS study

The ethical committees or institutional review boards of the following organizations approved the study: University of Graz Ethics Committee (Austria), Royal Melbourne Hospital Human Research Ethics Committee (Australia), Hunter New England Local Health District Research Ethics & Governance Office (Australia), Sir Charles Gairdner Group Human Research Ethics Committee (Australia), University of Tasmania Office of Research Services (Australia), University of Leuven Ethics Committee (Belgium), University of Calgary Conjoint Health Research Ethics Board (Canada), Scientific Ethics Committees for the Capital Region and the Municipalities of Copenhagen and Frederiksberg (Denmark), University of Mainz Ethics Committee (Germany), Technical University of Munich Ethics Committee (Germany), University Hospital of Larissa Local Ethics Committee (Greece), University of Piemonte Novara Ethics Committee (Italy), San Raffaele Hospital (IRCCS) Ethics Committee (Italy), Erasmus University Rotterdam Ethics Committee (Netherlands), Regional Committee for Medical Research Ethics (Norway), Clinical Research Ethics Committee of the Hospital Clinic of Barcelona (Spain), Vall D'Hebron University Hospital Research Ethics Committee (Spain), Karolinska Institutet Ethical Assurances (Sweden), Thames Valley Multi Centre Research Ethics Committee (UK), Brigham and Women’s Hospital Partners Human Research Committee (USA), University of California San Francisco Committee on Human Research (USA), University of California at Berkeley Committee for protection of Human Subjects (USA), and University of Miami Human Subject Research Office (USA).

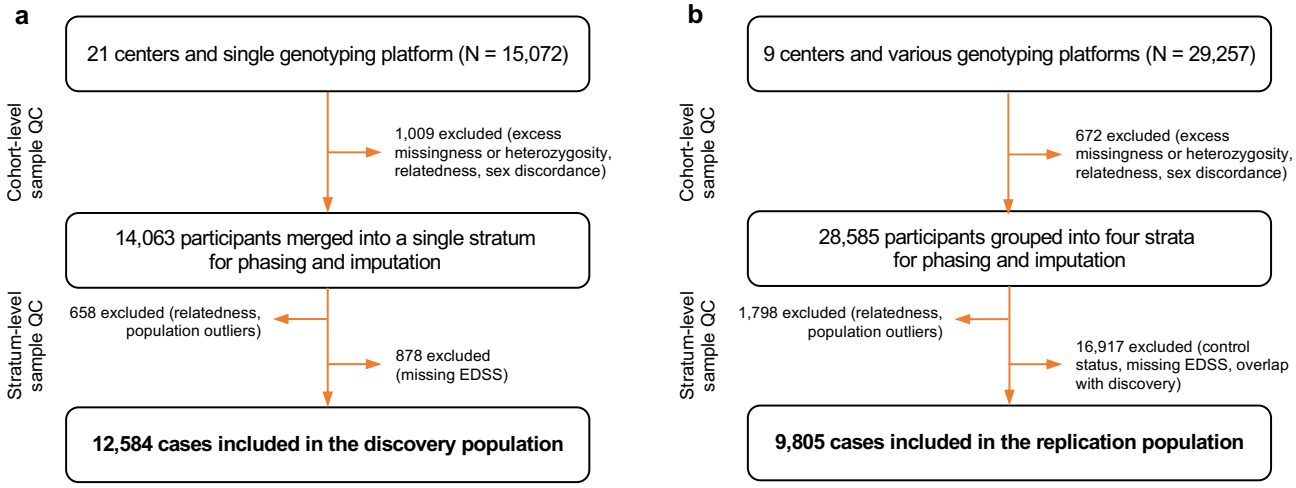
Supplementary References

1. 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).
2. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
3. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* vol. 4 Preprint at https://doi.org/10.1186/s13742-015-0047-8 (2015).
4. Anderson, C. A. *et al.* Data quality control in genetic case-control association studies. *Nature Protocols* vol. 5 1564–1573 Preprint at https://doi.org/10.1038/nprot.2010.116 (2010).
5. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
6. Bellenguez, C. *et al.* A robust clustering algorithm for identifying problematic samples in genome-wide association studies. *Bioinformatics* **28**, 134–135 (2012).
7. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
8. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
9. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
10. Clogg, C. C., Petkova, E. & Haritou, A. Statistical Methods for Comparing Regression Coefficients Between Models. *American Journal of Sociology* vol. 100 1261–1293
11. Jäkel, S. *et al.* Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* **566**, 543–547 (2019).
12. Absinta, M. *et al.* A lymphocyte-microglia-astrocyte axis in chronic active multiple sclerosis. *Nature* **597**, 709–714 (2021).
13. Morris, J. H. *et al.* The scalable precision medicine open knowledge engine (SPOKE): a massive knowledge graph of biomedical information. *Bioinformatics* **39**, btad080 (2023).
14. Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437–1452 (2017).
15. Mendez, D. *et al.* ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* **47**, D930–D940 (2019).
16. Chen, X., Liu, M. & Gilson, M. K. BindingDB: a web-accessible molecular recognition database. *Comb. Chem. High Throughput Screen.* **4**, 719–725 (2001).
17. Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).
18. Liu, X. *et al.* The Therapeutic Target Database: an Internet resource for the primary targets of approved, clinical trial and experimental drugs. *Expert Opin. Ther. Targets* **15**, 903–912 (2011).
19. Whirl-Carrillo, M. *et al.* An Evidence-Based Framework for Evaluating Pharmacogenomics Knowledge for Personalized Medicine. *Clin. Pharmacol. Ther.* **110**, 563–572 (2021).
20. Ochoa, D. *et al.* Open Targets Platform: supporting systematic drug-target identification and prioritisation. *Nucleic Acids Res.* **49**, D1302–D1310 (2021).
21. Reay, W. R. & Cairns, M. J. Advancing the use of genome-wide association studies for drug

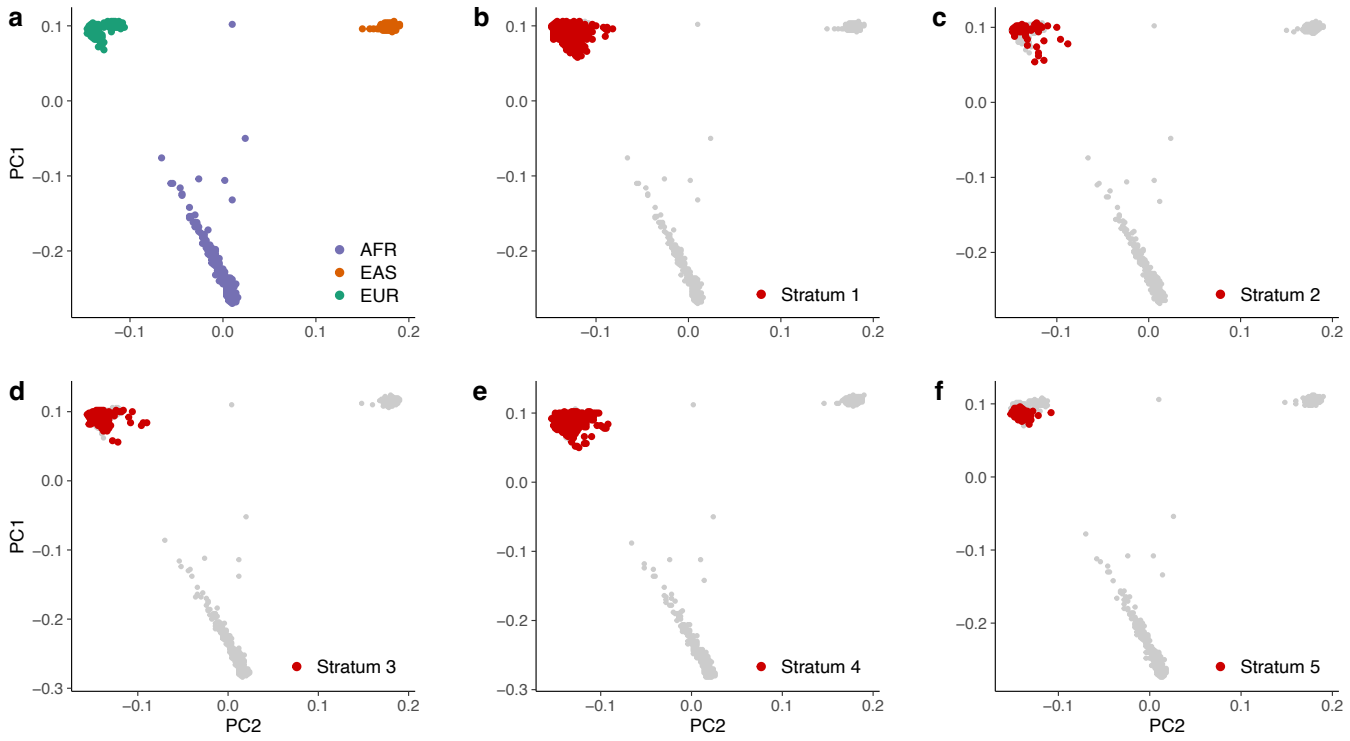
repurposing. *Nat. Rev. Genet.* **22**, 658–671 (2021).
22. Konuma, T., Ogawa, K. & Okada, Y. Integration of genetically regulated gene expression and pharmacological library provides therapeutic drug candidates. *Hum. Mol. Genet.* **30**, 294–304 (2021).
23. Mancuso, N. *et al.* Probabilistic fine-mapping of transcriptome-wide association studies. *Nat. Genet.* **51**, 675–682 (2019).
24. Namba, S. *et al.* A practical guideline of genomics-driven drug discovery in the era of global biobank meta-analysis. *Cell Genom* **2**, 100190 (2022).
25. Kazantsev, A. G. & Thompson, L. M. Therapeutic application of histone deacetylase inhibitors for central nervous system disorders. *Nat. Rev. Drug Discov.* **7**, 854–868 (2008).
26. Shukla, S. & Tekwani, B. L. Histone Deacetylases Inhibitors in Neurodegenerative Diseases, Neuroprotection and Neuronal Differentiation. *Front. Pharmacol.* **11**, 537 (2020).
27. Zhang, Z., Zhang, Z.-Y., Wu, Y. & Schluesener, H. J. Valproic acid ameliorates inflammation in experimental autoimmune encephalomyelitis rats. *Neuroscience* **221**, 140–150 (2012).
28. Camelo, S. *et al.* Transcriptional therapy with the histone deacetylase inhibitor trichostatin A ameliorates experimental autoimmune encephalomyelitis. *J. Neuroimmunol.* **164**, 10–21 (2005).
29. Ge, Z. *et al.* Vorinostat, a histone deacetylase inhibitor, suppresses dendritic cell function and ameliorates experimental autoimmune encephalomyelitis. *Exp. Neurol.* **241**, 56–66 (2013).
30. Castelo-Branco, G. *et al.* Acute treatment with valproic acid and l-thyroxine ameliorates clinical signs of experimental autoimmune encephalomyelitis and prevents brain pathology in DA rats. *Neurobiol. Dis.* **71**, 220–233 (2014).
31. International Multiple Sclerosis Genetics Consortium. Low-Frequency and Rare-Coding Variation Contributes to Multiple Sclerosis Risk. *Cell* **175**, 1679–1687 (2018).
32. Becker, J. *et al.* Resource profile and user guide of the Polygenic Index Repository. *Nat Hum Behav* **5**, 1744–1758 (2021).
33. Daetwyler, H. D., Villanueva, B. & Woolliams, J. A. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* **3**, e3395 (2008).
34. Wray, N. R. *et al.* Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* **14**, 507–515 (2013).
35. Choi, S. W. & O'Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* **8**, giz082 (2019).
36. Hedström, A. K., Hillert, J., Olsson, T. & Alfredsson, L. Alcohol as a modifiable lifestyle factor affecting multiple sclerosis risk. *JAMA Neurol.* **71**, 300–305 (2014).
37. Hillert, J. & Stawiarz, L. The Swedish MS registry – clinical support tool and scientific resource. *Acta Neurol. Scand.* **132**, 11–19 (2015).
38. Bryois, J. *et al.* Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease. *Nat. Genet.* **52**, 482–493 (2020).
39. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
40. Lake, B. B. *et al.* Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
41. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
42. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia.

Nature **604**, 502–508 (2022).
43. University of California, San Francisco MS-EPIC Team *et al.* Silent progression in disease activity-free relapsing multiple sclerosis. *Ann. Neurol.* **85**, 653–666 (2019).
44. Tremlett, H., Zhao, Y., Joseph, J., Devonshire, V. & UBCMS Clinic Neurologists. Relapses in multiple sclerosis are age- and time-dependent. *J. Neurol. Neurosurg. Psychiatry* **79**, 1368–1374 (2008).
45. Vandebergh, M. *et al.* Genetic Variation in WNT9B Increases Relapse Hazard in Multiple Sclerosis. *Ann. Neurol.* **89**, 884–894 (2021).

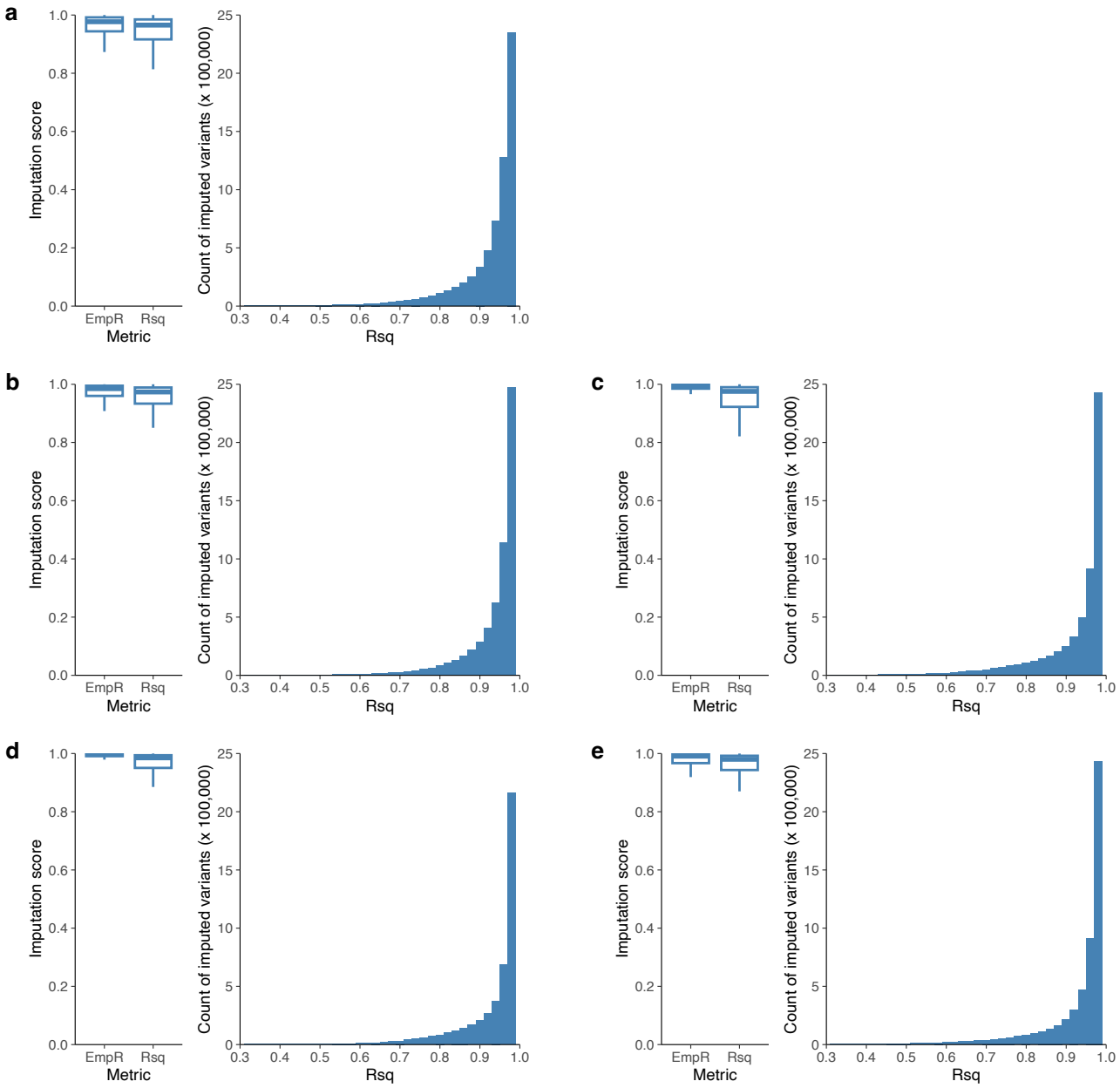
Supplementary Figures



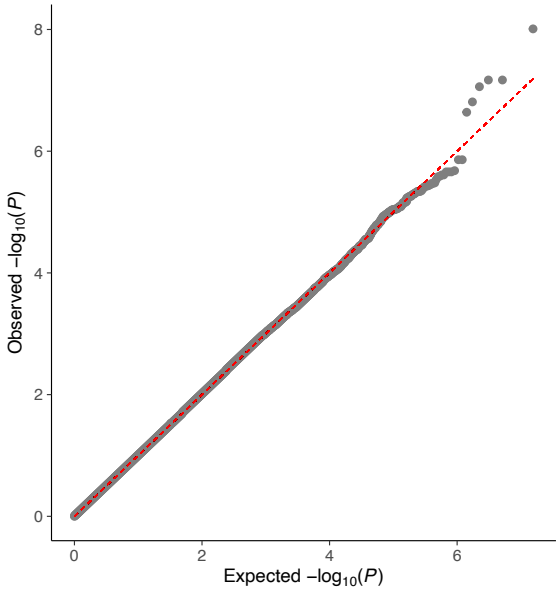
Supplementary Figure 1 | Flowchart of study participants. Each chart details the number of participants submitted to quality control and total number of cases included in the analysis for the discovery (**a**) and replication (**b**) populations. Note that the replication population included MS cases and controls genotyped as part of the same effort. To maximize subsequent phasing and imputation accuracy, all individuals were included in the quality control process. MS cases without measures of disability and controls were excluded prior to analysis, resulting in the final sample count.



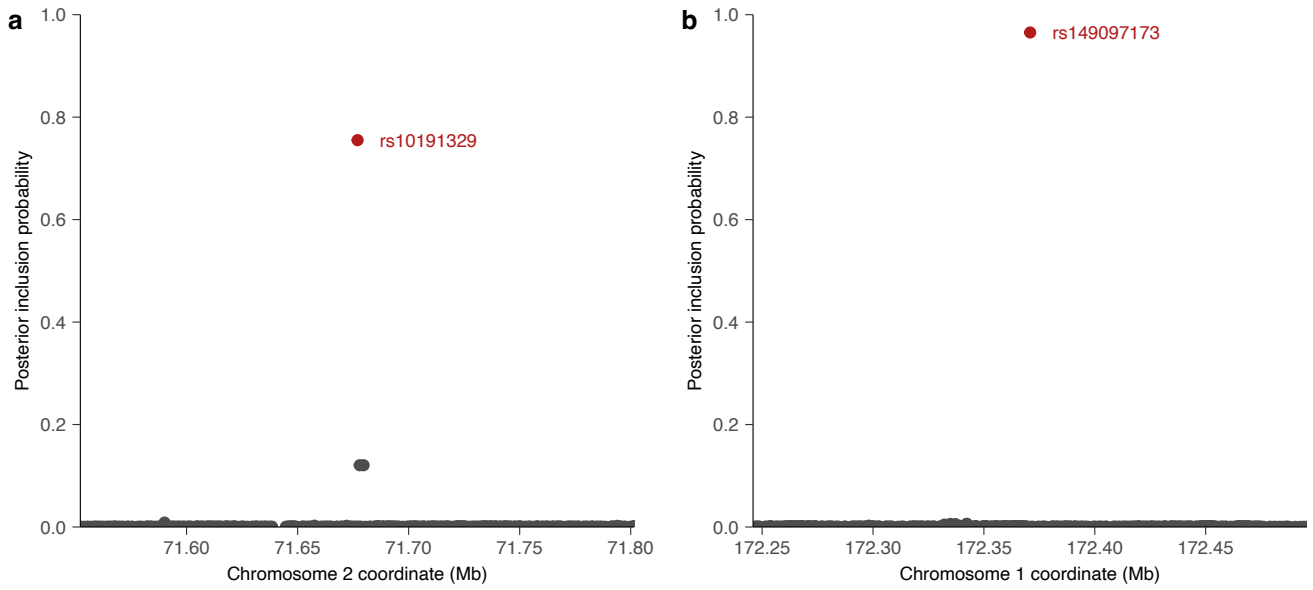
Supplementary Figure 2 | Principal component projections per stratum including 1000 Genomes reference populations. a, Principal component scores for African (AFR), East Asian (EAS) and European (EUR) samples in 1000 Genomes phase 3. **b-f**, Using a common set of high-quality variants, individuals included in the GWAS following quality control are plotted along their projected principal components (red) based on the same 1000 Genomes reference populations (gray). Stratum 1 (**b**) corresponds to the discovery population, whereas strata 2 through 5 (**c-f**) represent the replication population. Cases consistently overlap with the European cluster in 1000 Genomes.



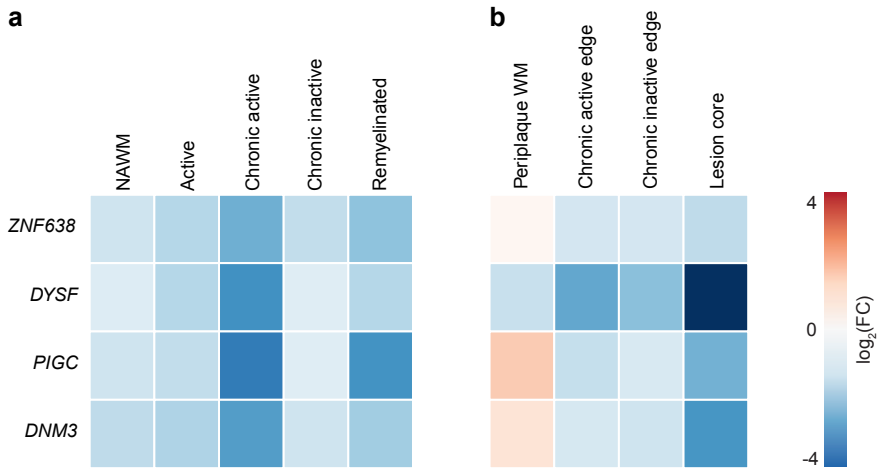
Supplementary Figure 3 | Imputation quality and accuracy. Imputation accuracy was estimated by examining the empirical correlation (EmpR) between true genotypes and imputed dosages for directly genotyped variants. Genome-wide imputation quality was evaluated using the square correlation (Rsq) between imputed dosages and true, unobserved genotypes. Left, EmpR and Rsq metrics are presented for variants with MAF > 0.01; Right, Rsq distribution for imputed variants with MAF > 0.01 and Rsq ≥ 0.3. **a**, Imputation metrics for the discovery population. **b-e**, Imputation metrics for each of the four strata part of the replication population. The number of variants used to construct the EmpR and Rsq boxplots for each panel were respectively: (**a**) 456,478 and 7,947,413; (**b**) 455,932 and 7,917,572; (**c**) 458,542 and 7,914,432; (**d**) 579,565 and 7,888,480; and (**e**) 719,165 and 7,987,450. Box plots show median, first, and third quartiles; whiskers represent the smallest and largest values within 1.5-times the interquartile range; outliers are not depicted. MAF, minor allele frequency.



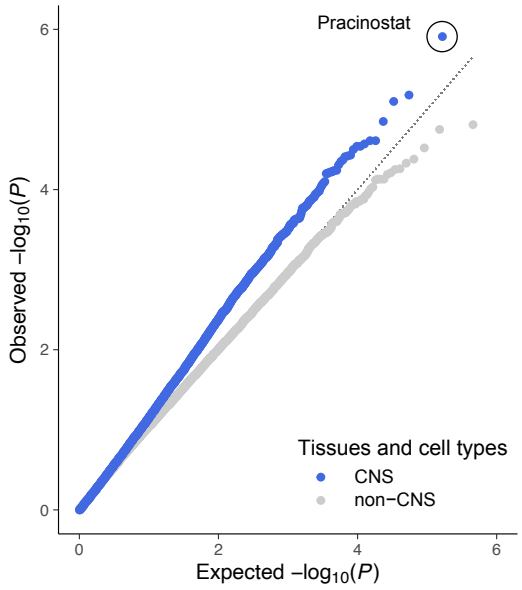
Supplementary Figure 4 | Quantile-quantile plot for the ARMSS score discovery GWAS. Quantile-quantile plot of the observed two-sided -$\log_{10}(P)$ from linear regression analyses versus the expectation under the null hypothesis. The genomic control inflation factor (λ_{GC}) was 1.016, while the LD score regression intercept was not significantly different from 1 (1.009, 95% CI 0.996 to 1.022). ARMSS, age-related MS severity; CI, confidence interval; LD, linkage disequilibrium.



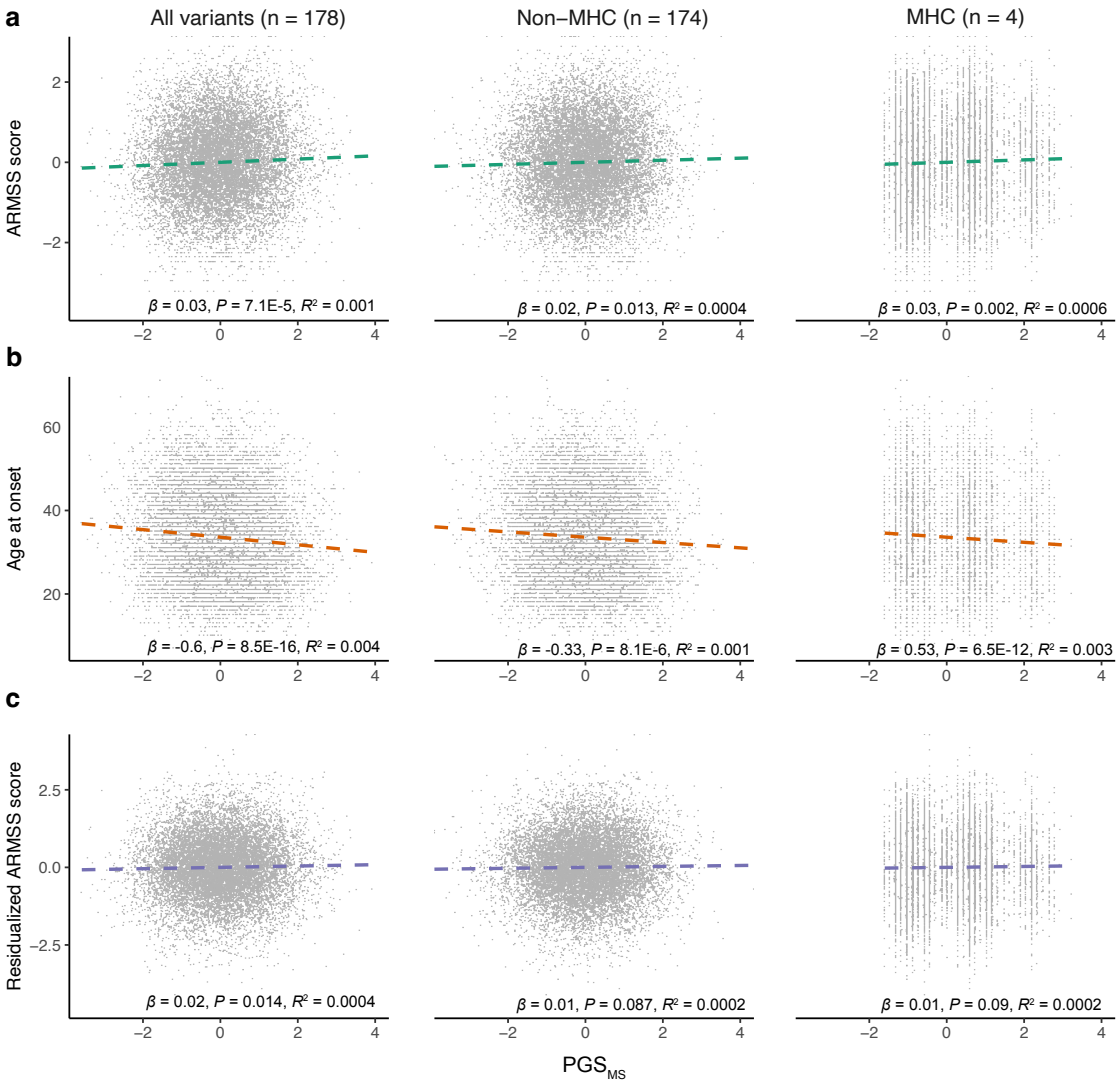
Supplementary Figure 5 | Fine mapping of the MS severity loci. a, Statistical fine-mapping at the *DYSF–ZNF638* locus identified the rs10191329 as the most likely causal variant with a high posterior inclusion probability (PIP). **b**, Statistical fine-mapping at the suggestive *DNM3-PIGC* locus found rs149097173 as the most likely causal variant. MS, multiple sclerosis.



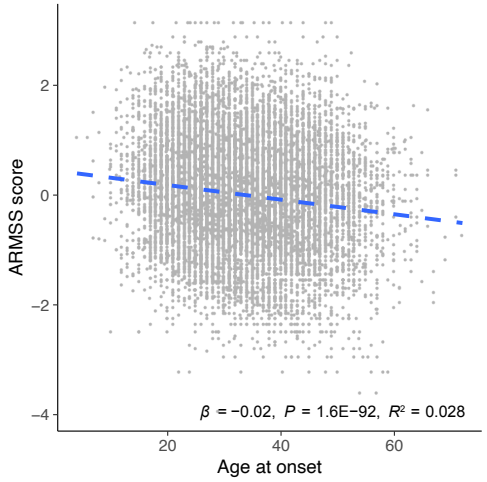
Supplementary Figure 6 | Differential expression of prioritized severity genes in MS lesions. a, Heatmap of the $\log_2$-fold change in expression of the four prioritized severity genes across different MS lesion types and normal appearing white matter (NAWM) compared to control white matter. Samples originated from a study of four individuals with progressive MS and five controls[11]. **b,** Repeated analysis in a study of five individuals with progressive MS and three controls[12]. All comparisons were statistically significant ($P < 0.05/36$) using a Wilcoxon Rank Sum two-tailed test adjusted for the number of genes (N = 4) and sampling conditions across studies (N = 9). FC, fold-change; WM, white matter.



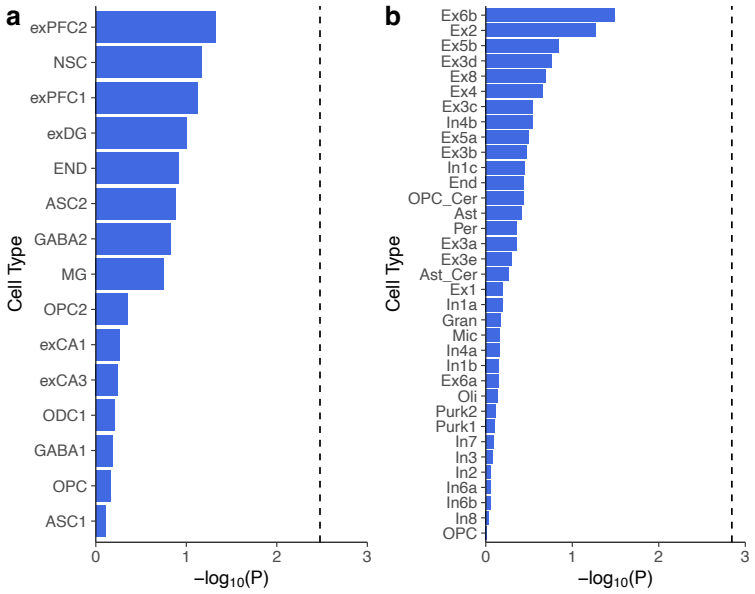
Supplementary Figure 7 | Quantile-quantile plot for correlation tests between genetically determined expression associated with MS severity and compound-perturbed gene expression. Each data point represents a cell-type-compound pair, with those in CNS tissues showing a departure from the expected distribution under the null hypothesis. One-sided *P* values were calculated using negative Spearman's rank correlation.



Supplementary Figure 8 | ARMSS score and MS susceptibility PGS (PGS_{MS}). a, An increased burden of MS susceptibility variants, as reflected by the PGS_{MS}, modestly increased ARMSS scores. This effect was similar for MHC and non-MHC variants. **b**, The PGS_{MS} had a stronger inverse association with age at onset. **c**, After adjusting for age of onset (in addition to baseline covariates), the effect of the PGS_{MS} on ARMSS scores was attenuated. The regression coefficients and two-sided *P* values were calculated using linear regression models. The R^2 values represent the gain in coefficient of determination (incremental R^2). ARMSS and PGS_{MS} scores were standardized to facilitate comparison. ARMSS, age-related multiple sclerosis severity score (rank-based inverse-normal transformed); MHC, major histocompatibility complex; PGS, polygenic score.



Supplementary Figure 9 | ARMSS score and age at onset. Earlier age at onset is associated with increased MS severity when comparing individuals of the same age. The regression coefficient and its corresponding two-sided *P* value were calculated using a linear regression model adjusted for age, sex, date of birth, EDSS source, center, genotyping batch, and the first ten principal components. The R^2 value represents the gain in coefficient of determination (incremental R^2) when age at onset is added as a variable to a regression of ARMSS score on the baseline covariates. ARMSS, age-related MS severity score (rank-based inverse-normal transformed).



Supplementary Figure 10 | Association between MS severity and CNS cell types by heritability enrichment. **a,** Enrichment of MS severity associations in genes with high expression, by single-nucleus RNA sequencing, in 15 human cell types from human prefrontal cortex and the hippocampus. exPFC, glutamatergic neurons from the prefrontal cortex; GABA, GABAergic interneurons; exCA, pyramidal neurons from the hippocampus Cornu Ammonis region; exDG, granule neurons from the hippocampus dentate gyrus region; ASC, astrocytes; MG, microglia; ODC, oligodendrocytes; OPC, oligodendrocyte precursor cells; NSC, neuronal stem cells; END, endothelial cells. **b,** Repeated analysis in 24 cell types from the human frontal and visual cortex, as well as the cerebellum. Ex, cortical excitatory neuronal subtypes; In, cortical inhibitory neuronal subtypes; Gran, cerebellar granule cells; Purk, Purkinje neurons; End, endothelial cells; Per, pericytes; Ast, astrocytes; Oli, oligodendrocytes; OPC, oligodendrocytes their precursor cells; Mic, microglia; Ast_Cer, cerebellar-specific Ast; OPC_Cer, cerebellar-specific OPC. For both panels, one-sided *P* values for cell type heritability enrichment were obtained using partitioned LD score regression.