Title: Multiple Sclerosis Genomic Map implicates peripheral immune cells & microglia in susceptibility

Authors: International Multiple Sclerosis Genetics Consortium*

The full list of names and affiliations is listed at the end of the paper.

Article type: Research Article

Abstract word count: 122

Word count: 5,065

References: 42

Figures: 6

Subheadings: 4

*Correspondence to: Philip L. De Jager, MD PhD

Center for Translational & Computational Neuroimmunology

Multiple Sclerosis Center

Department of Neurology

Columbia University Medical Center

630 W 168th Street P&S Box 16

New York, NY 10032

T: 212.305.3609

pld2115@cumc.columbia.edu

Abstract:

We analyzed genetic data of 47,429 multiple sclerosis (MS) and 68,374 control subjects and establish a reference map of the genetic architecture of MS that includes 200 autosomal susceptibility variants outside the major histocompatibility complex (MHC), one chromosome X variant, and 32 within the extended MHC. We used an ensemble of methods to prioritize 551 putative susceptibility genes, that implicate multiple innate and adaptive pathways distributed across the cellular components of the immune system. Using expression profiles from purified human microglia, we observe enrichment for MS genes in these brain-resident immune cells, suggesting that these may have a role in targeting an autoimmune process to the central nervous system, although MS is most likely initially triggered by perturbation of peripheral immune responses.

One Sentence Summary: We report a detailed genetic and genomic map of multiple sclerosis, and describe the role of putatively affected genes in the peripheral immune system and brain resident microglia.

Introduction

Over the last decade, elements of the genetic architecture of multiple sclerosis (MS) susceptibility have gradually emerged from genome-wide and targeted studies.(*1-6*) The role of the adaptive arm of the immune system, particularly its CD4+ T cell component has become clearer, with multiple different T cell subsets being implicated.(*4*) While the T cell component plays an important role, functional and epigenomic annotation studies have begun to suggest that other elements of the immune system may be involved as well.(*7, 8*) Here, we assemble available genome-wide MS data to perform a meta-analysis followed by a systematic, comprehensive replication effort in large independent sets of subjects. This effort has yielded a detailed genome-wide genetic map that includes the first successful evaluation of the X chromosome in MS, and provides a powerful platform for the creation of a detailed genomic map, outlining the functional consequences of most variants and their assembly into susceptibility networks (Fig. S1).

Discovery and replication of genetic associations

We organized available (*1, 2, 4, 5*) and newly genotyped genome-wide data in 15 data sets, totaling 14,802 subjects with MS and 26,703 controls for our discovery study (*9*) (Tables S1-S3). After rigorous per data set quality control, we imputed all samples using the 1000 Genomes European panel resulting in an average of 7.8 million imputed single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF) of at least 1% (*9*). We then performed a meta-analysis, penalized for within-data set residual genomic inflation, to a total of 8,278,136 SNPs with data in at least two data sets (*9*). Of these, 26,395 SNPs reached genome-wide significance (p-value $< 5 \times 10^{-8}$; fixed effects inverse-variance meta-analysis) and another 576,204 SNPs had at least nominal evidence of association ($5 \times 10^{-8} >$ p-value < 0.05; fixed effects inverse-variance meta-analysis). In order to identify statistically independent SNPs in the discovery set and to prioritize variants for replication, we applied a genome partitioning approach (*9*). Briefly, we first excluded an extended region of ~12Mb around the major histocompatibility complex (MHC) locus to scrutinize this unique region separately (see below), and we then applied an iterative method to discover statistically independent SNPs in the rest of the genome using conditional

modeling. We partitioned the genome into regions by extracting ±1Mbs on either side of the most statistically significant SNP and repeating this procedure until there were no SNPs with a p-value<0.05 (fixed effects inverse-variance meta-analysis) left in the genome. Within each region we applied conditional modeling to identify statistically independent effects (Fig. S2). As a result, we identified 1,961 non-MHC autosomal regions that included 4,842 presumably statistically independent SNPs. We refer to these 4,842 prioritized SNPs as "*effects"*, assuming that these SNPs tag a true causal genetic effect. Of these, 82 effects were genome-wide significant in the discovery analysis, and another 125 had a p-value < $1x10^{-5}$ (fixed effects inverse-variance meta-analysis).

In order to replicate these 4,842 effects, we analyzed two large-scale independent sets of data. First, we designed the MS Chip to directly replicate each of the prioritized effects (*9*) and, after stringent quality check (*9*) (Table S4), analyzed 20,360 MS subjects and 19,047 controls, which were organized in 9 data sets. Second, we incorporated targeted genotyping data generated using the ImmunoChip platform on an additional 12,267 MS subjects and 22,625 control subjects that had not been used in either the discovery or the MS Chip subject sets (Table S5).(*3*) Overall, we jointly analyzed data from 47,429 MS cases and 68,374 control subjects to provide the largest and most comprehensive genetic evaluation of MS susceptibility to date.

For 4,311 of the 4,842 effects (89%) that were prioritized in the discovery analysis, we could identify at least one tagging SNP in the replication data (*9*) (Table S6). 156 regions had at least one genome-wide effect, and, overall, 200 prioritized effects reached a level of genome-wide significance (GW) in these 156 regions (Fig. 1). 62 of these 200 effects represent secondary, independent, effects that emerged from conditional modeling within a given locus (*9*) (Table S7, Fig. S3). The odds ratios (ORs) of these genome-wide effects ranged from 1.06 to 2.06, and the allele frequencies of the respective risk allele from 2.1% to 98.4% in the European samples of the 1000 Genomes reference (mean: 51.3%, standard deviation: 24.5%; Table S8 and Fig. S4). 19.9% of regions (31 out of 156) harbored more than one statistically independent GW effect. One of the most complex regions was the one harboring the *EVI5* gene that has been the subject of several

reports with contradictory results.(*10-13*) In this locus, we identified four statistically independent genome-wide effects, three of which were found under the same association peak (Fig. 2A), illustrating how our approach and the large sample size clarifies associations described in smaller studies and can facilitate functional follow-up of complex loci.

We also performed a joint analysis of available data on sex chromosome variants (*9*), and we identified rs2807267 as genome-wide significant (OR_T=1.07, p-value=$6.86x10^{-9}$; fixed effects inverse-variance meta-analysis; Tables S9-S10). This variant lies within an enhancer peak specific for T cells and is 948bps downstream of the RNA U6 small nuclear 320 pseudogene (*RNU6-320P*), a component of the U6 small nuclear ribonucleoprotein (snRNP) that is part of the spliceosome and is responsible for the splicing of introns from pre-mRNA (*14*) (Fig. 2B). The nearest gene is *VGLL1* (27,486bps upstream) that has been proposed to be a co-activator of mammalian transcription factors.(*15*) No variant in the Y chromosome had a p-value lower than 0.05 (fixed effects inverse-variance meta-analysis).

The MHC was the first MS susceptibility locus to be identified, and prior studies have found that it harbors multiple independent susceptibility variants, including interactions within the class II HLA genes.(*16, 17*) We undertook a detailed modeling of this region to account for its long-range linkage disequilibrium and allelic heterogeneity using SNP data as well as imputed classical alleles and amino acids of the human leukocyte antigen (HLA) genes in the assembled data. We confirm prior MHC susceptibility variants (including a non-classical HLA effect located in the *TNFA/LST1* long haplotype), and we extend the association map to uncover a total of 31 statistically independent effects at the genome-wide level within the MHC (Fig. 3, Table S11). An interesting finding is that multiple HLA and nearby non-HLA genes have several independent effects that can now be identified due to our large sample, e.g. the *HLA-DRB1* locus has six statistically independent effects. Another exciting finding involves *HLA-B* that also appears to harbor 6 independent effects on MS susceptibility. The role of the non-classical HLA and non-HLA genome in the MHC is also highlighted. One third (9 out of 31) of the identified variants lie within either intergenic regions or in a long-range haplotype that contains several non-classical

HLA and other non-HLA genes.(*17*) Recently, we reported an interaction between HLA-DRB1*15:01 and HLA-DQA1*01:01 by analyzing imputed HLA alleles.(*16*) Here we reinforce this analysis by analyzing SNPs, HLA alleles, and respective amino acids. We replicate the presence of interactions among class II alleles but note that the second interaction term, besides HLA-DRB1*15:01, can vary depending on the other independent variants that are included in the model. First, we found that there are interaction models of HLA-DRB1*15:01 with other variants in MHC that explain better the data than our previously reported HLA-DRB1*15:01/HLA-DQA1*01:01 interaction term (Fig. S5). Second, we observe that there is a group of *HLA*DQB1* and *HLA*DQA1* SNPs, alleles, and amino acids that consistently rank amongst the best models with HLA-DRB1*15:01 interaction terms (Fig. S6). This group of HLA-DRB1*15:01-interacting variants is consistently identified regardless of the marginal effects of other statistically independent variants that are added in the model, implying that these interaction terms capture a different subset of phenotypic variance and can be explored after the identification of the marginal effects. Finally, we performed a sensitivity analysis by including interaction terms of HLA-DRB1*15:01 in each step and selecting the model with the lowest Bayesian information criterion (BIC), instead of testing only the marginal results of the variants as we did in the main analysis (classical model MHC analysis; Table S12). This sensitivity analysis also resulted in 32 statistically independent effects with a genome-wide significant p-value (fixed effects inverse-variance meta-analysis; Table S12), of which one third (9 out of 32) were different than the ones in classical model MHC analysis. The main differences between the results of the two approaches were the inclusion of interaction of HLA-DRB1*15:01 and rs1049058 in step 3 and the stronger association of *HLA*DPB1/2* effects over *HLA*DRB1* effects in the sensitivity model (Tables S12-S13 and Fig. S6). Thus, overall, our MHC results are not strongly affected by the analytic model that we have selected.

Characterization of non-genome wide effects

The commonly used threshold of genome-wide significance (p-value = $5x10^{-8}$) has played an important role in making human genetic study results robust; however, several studies have

demonstrated that non-genome-wide effects explain an important proportion of the effect of genetic variation on disease susceptibility. (*18, 19*) More importantly, several such effects are eventually identified as genome-wide significant, given enough sample size and true effects.(*3*) Thus, we also evaluated the non-genome-wide effects that were selected for replication, have available replication data (n=4,111), but do not meet a standard threshold of genome-wide significance ($p<5x10^{-8}$). Specifically, we decided to stratify these 4,111 effects into 2 main categories (*9*): (1) suggestive effects (S, n=416), and (2) non-replicated effects (NR, n= 3,695). We used these categories in downstream analyses to further characterize the prioritized effects from the discovery study in terms of potential to eventually be replicated. We also included a third category: effects for which there were no data for replication in any of the replication sets (no data, ND, n=532). Furthermore, to add granularity in each category, we sub-stratified the suggestive effects into 2 groups: (1a) strongly suggestive ($5 x10^{-8} > p\text{-value} <1x10^{-5}$; sS, n= 117; fixed effects inverse-variance meta-analysis) and (1b) underpowered suggestive (unS, n=299). Of these two categories of suggestive effects, the ones in the sS category have a high probability of reaching genome-wide significance as we increase our sample size in future studies (*9*) (Table S14).

Heritability explained

To estimate the extent to which we have characterized the genetic architecture of MS susceptibility with our 200 genome-wide non-MHC autosomal MS effects, we calculated the narrow-sense heritability captured by common variation (*h2g*), i.e. the ratio of additive genetic variance to the total phenotypic variance ().(*18, 20*) Only the 15 strata of data from the discovery set had true genome-wide coverage, and hence we used these 14,802 MS subjects and 26,703 controls for the heritability analyses. The overall heritability estimate for MS susceptibility in the discovery set of subjects was 19.2% (95%CI: 18.5-19.8%). Heritability partitioning using minor allele frequency or p-value thresholds has led to significant insights in previous studies,(*21*) and we therefore applied a similar partitioning approach but in a fashion that took into consideration the study design and the existence of replication information from the 2 large-scale replication

cohorts. First, we partitioned the autosomal genome into 3 components: i) the super extended MHC (SE MHC, see above), ii) a component with the 1,961 regions prioritized for replication (Regions), and iii) the rest of the genome that had a p-value>0.05 (fixed effects inverse-variance meta-analysis) in the discovery study (Non-associated regions). Then, we estimated the *h2g* that can be attributed to each component as a proportion of the overall narrow-sense heritability observed. The SE MHC explained 21.4% of the *h2g*, with the remaining 78.6% being captured by the second component (Fig. 4A). Then, we further partitioned the non-MHC component into one that captured all 4,842 statistically independent effects (Prioritized for replication), which explained the vast majority of the overall estimated heritability: 68.3%. The "Non-prioritized" SNPs in the 1,961 regions explained 11.6% of the heritability, which suggests that there may be residual LD with prioritized effects or true effects that have not yet been identified (Fig. 4B).

We then used the replication-based categories described above to further partition the "Prioritized" heritability component, namely "GW", "S", "NR", "ND" (Fig. 4C). The genome-wide effects (GW) captured 18.3% of the overall heritability. Thus, along with the contribution of the SE MHC (20.2% in the same model), we can now explain ~39% of the genetic predisposition to MS with the validated susceptibility alleles. This can be extended to ~48% if we include the suggestive (S) effects (9.0%). Interestingly the non-replicated (NR) effects captured 38.8% of the heritability, which could imply that some of these effects might be falsely non-replicated, i.e. that these are true effects that need further data to emerge robustly or that their effect may be true and present in only a subset of the data. However, few of the 3,695 NR effects would fall in either of the above two cases; the vast majority of these effects are likely to be false positive results.

Functional implications of the MS loci, enriched pathways and gene-sets

Next, we began to annotate the MS effects. To prioritize the cell types or tissues in which the 200 non-MHC autosomal effects may exert their effect, we used two different approaches: one that leverages atlases of gene expression patterns and another that uses a catalog of epigenomic features such as DNase hypersensitivity sites (DHSs).(*8, 9, 22-24*) Significant enrichment for MS susceptibility loci was apparent in many different immune cell types and tissues, whereas there

was an absence of enrichment in tissue-level central nervous system (CNS) profiles (Fig. 5). An important finding is that the enrichment is observed not only in immune cells that have long been studied in MS, e.g. T cells, but also in B cells whose role has emerged more recently.(*25*) Furthermore, while the adaptive immune system has been proposed to play a predominant role in MS onset,(*26*) we now demonstrate that many elements of innate immunity, such as natural killer (NK) cells and dendritic cells also display strong enrichment for MS susceptibility genes. Interestingly, at the tissue level, the role of the thymus is also highlighted, possibly suggesting a role of genetic variation in thymic selection of autoreactive T cells in MS.(*27*) Public tissue-level CNS data – which are derived from a complex mixture of cell types - do not show an excess of MS susceptibility variants in annotation analyses. However, since MS is a disease of the CNS, we extended the annotation analyses by analyzing data generated from human iPSC-derived neurons as well as from purified primary human astrocytes and microglia (*9*). As seen in Fig. 6, enrichment for MS genes is seen in human microglia ($p=5x10^{-14}$) but not in astrocytes or neurons, suggesting that the resident immune cells of the brain may also play a role in MS susceptibility.

We repeated the enrichment analyses for the "S" and "NR" effects aiming to test whether these have a similar enrichment pattern with the 200 "GW" effects. The "S" effects exhibited a pattern of enrichment that is similar to the "GW" effects, with only B cell expression reaching a threshold of statistical significance (Fig. S7). This provides additional circumstantial evidence that this category of variants may harbor true causal associations. On the other hand, the "NR" enrichment results seem to follow a rather random pattern, suggesting that most of these effects are indeed not truly MS-related (Fig. S7).

The strong enrichment of the GW effects in immune cell types motivated us to prioritize candidate MS susceptibility genes by identifying those susceptibility variants, which affect RNA expression of nearby genes (*cis* expression quantitative trait loci effect, *cis* -eQTL) (±500Kbps around the effect SNP) (*9*). Thus, we interrogated the potential function of MS susceptibility variants in naive CD4+ T cells and monocytes from 211 healthy subjects as well as peripheral blood mononuclear cells (PBMCs) from 225 remitting relapsing MS subjects. Thirty-six out of

the 200 GW MS effects (18%) had at least one tagging SNP (r2>=0.5) that altered the expression of 46 genes (false discovery rate; FDR<5%) in CD4+ naïve T cells (Table S15-S16), and 36 MS effects (18%; 10 common with the CD4+ naïve T cells) influenced the expression of 48 genes in monocytes (11 genes in common with T cells). In MS PBMC, 30% of the GW effects (60 out of the 200) were cis-eQTLs for 92 genes in the PBMC MS samples, with several loci being shared with those found in healthy T cells and monocytes (26 effects and 27 genes in T cells, and 21 effects and 24 genes in monocytes, respectively; Table S15-S16).

Since MS is a disease of the CNS, we also investigated a large collection of dorsolateral prefrontal cortex RNA sequencing profiles from two longitudinal cohort studies of aging (n=455), which recruit cognitively non-impaired individuals (*9*). This cortical sample provides a tissue-level profile derived from a complex mixture of neurons, astrocytes, and other parenchymal cells such as microglia and occasional peripheral immune cells. In these data, we found that 66 of the GW MS effects (33% of the 200 effects) were *cis*-eQTLs for 104 genes. Over this CNS and the three immune sets of data, 104 GW effects were *cis*-eQTLs for 203 unique genes (n=211 *cis*-eQTLs), with several appearing to be seemingly specific for one of the cell/tissue type (Table S16). Specifically, 21.2**%** (45 out of 212 *cis*-eQTLs) of these cortical cis-eQTLs displayed no evidence of association (p-value>0.05, for linear regression(*9*), with any SNP with r^2>0.1) in the immune cell/tissue results and are less likely to be immune-related (Table S16-S17).

To further explore the challenging and critical question of whether some of the MS variants have an effect that is primarily exerted through a non-immune cell, we performed a secondary analysis of our cortical RNAseq data in which we attempted to ascribe a brain cis-eQTL to a particular cell type. Specifically, we assessed our tissue-level profile and adjusted each cis-eQTL analysis for the proportion of neurons, astrocytes, microglia, and oligodendrocytes estimated to be present in the tissue: the hypothesis was that the effect of a SNP with a cell type-specific cis-eQTL would be stronger if we adjusted for the proportion of the target cell type (Fig. 6; Fig. S8). As anticipated, almost all of the MS variants present in cortex remain ambiguous: it is likely that many of them influence gene function in multiple immune and non-immune cell types. However,

the *SLC12A5* locus is different: here, the effect of the SNP is significantly stronger when we account for the proportion of neurons (Fig. 6A and 6B), and the *CLECL1* locus emerges when we account for the proportion of microglia. *SLC12A5* is a potassium/chloride transporter that is known to be expressed in neurons, and a rare variant in *SLC12A5* causes a form of pediatric epilepsy (*28, 29*). While this MS locus may therefore appear to be a good candidate to have a primarily neuronal effect, further evaluation finds that this MS susceptibility haplotype also harbors susceptibility to rheumatoid arthritis (*30*) and a cis-eQTL in B cells for the *CD40* gene (*31*). Thus, the same haplotype harbors different functional effects in very different contexts, illustrating the challenge in dissecting the functional consequences of autoimmune variants in immune function as opposed to the tissue targeted in autoimmune disease. On the other hand, *CLECL1* represents a simpler case of a known susceptibility effect that has previously been linked to altered *CLECL1* RNA expression in monocytes (*26, 32*); its enrichment in microglial cells, which share many molecular pathways with other myeloid cells, is more straightforward to understand. *CLECL1* is expressed at low level in our cortical RNAseq profiles because microglia represent just a small fraction of cells at the cortical tissue level, and its expression level is 20-fold greater when we compare its level of expression in purified human cortical microglia to the bulk cortical tissue (Fig. 6). *CLECL1* therefore suggests a potential role of microglia in MS susceptibility, which is under-estimated in bulk tissue profiles that are available in epigenomic and transcriptomic reference data. Overall, many genes that are eQTL targets of MS variants in the human cortex are most likely to affect multiple cell types. These brain eQTL results and the enrichment found in analyses of our purified human microglia data therefore highlight the need for more targeted, cell-type specific data for the CNS to adequately determine the extent of its role in MS susceptibility.

These eQTL studies begin to transition our genetic map into a resource outlining the likely MS susceptibility gene(s) in a locus and the potential functional consequences of certain MS variants. To assemble these single-locus results into a higher-order perspective of MS susceptibility, we turned to pathway analyses to evaluate how the extended list of genome-wide

effects provides new insights into the pathophysiology of the disease. Acknowledging that there is no available method to identify all causal genes following GWAS discoveries, we prioritized genes for pathway analyses while allowing several different hypotheses for mechanisms of actions (*9*). In brief, we prioritized genes that: (i) were *cis*-eQTLs in any of the eQTL data sets outlined above, (ii) had at least one exonic variant at r^2>=0.1 with any of the 200 effects**, (**iii) had high score of regulatory potential using a cell specific network approach, (iv) had a similar co-expression pattern as identified using DEPICT.(*33*) Sensitivity analyses were performed including different combinations of the above categories, and including genes with intronic variants at r^2>=0.5 with any of the 200 effects (*9*). Overall, we prioritized 551 candidate MS genes (Table S18; Table S19 for sensitivity analyses) to test for statistical enrichment of known pathways. Approximately 39.6% (142 out of 358) of the Ingenuity Pathway Analysis (IPA) canonical pathways,(*34*) that had overlap with at least one of the identified genes, were enriched for MS genes at an FDR<5% (Table S20). Sensitivity analyses including different criteria to prioritize genes revealed a similar pattern of pathway enrichment (*9*) (Table S21). Interestingly, the extensive list of susceptibility genes, that more than doubles the previous knowledge in MS, captures processes of development, maturation, and terminal differentiation of several immune cells that potentially interact to predispose to MS. In particular, the role of B cells, dendritic cells and natural killer cells has emerged more clearly, broadening the prior narrative of T cell dysregulation that emerged from earlier studies.(*4*) Given the over-representation of immune pathways in these databases, ambiguity remains as to where some variants may have their effect: neurons and particularly astrocytes repurpose the component genes of many “immune” signaling pathways, such as the ciliary neurotrophic factor (CNTF), nerve growth factor (NGF), and neuregulin signaling pathways that are highly significant in our analysis (Table S20). These results – along with the results relating to microglia – emphasize the need for further dissection of these pathways in specific cell types to resolve where a variant is exerting its effect; it is possible that multiple, different cell types could be involved in disease since they all experience the effect of the variant.

Pathway and gene-set enrichment analyses can only identify statistically significant connections of genes in already reported, and in some cases validated, mechanisms of action. However, the function of many genes is yet to be uncovered and, even for well-studied genes, the full repertoire of possible mechanisms is still incomplete. To complement the pathway analysis approach and to explore the connectivity of our prioritized GW genes, we performed a protein-protein interaction (PPI) analysis using GeNets.(*9, 35*) About a third of the 551 prioritized genes (n=190; 34.5%) were connected (p-value = 0.052; permutation-based p-value) and these could be organized into 13 communities, i.e. sub-networks with higher connectivity (p-value: < 0.002; permutation-based p-value; Table S22). This compares to 9 communities that could be identified by the previously reported MS susceptibility list (81 connected genes out of 307; Table S23).(*3*) Next, we leveraged GeNets to predict candidate genes based on network connectivity and pathway membership similarity and test whether our previous known MS susceptibility list could have predicted any of the genes prioritized by the newly identified effects. Of the 244 genes prioritized by novel findings (out of the 551 overall prioritized genes) only five could be predicted given the old results (out of 70 candidates that emerge from the extrapolation of earlier data; Fig. S9 and Table S24). In a similar fashion we estimated that the list of 551 prioritized genes could predict 102 new candidate genes, four of which can be prioritized since they are in the list of suggestive effects. (Fig. 1; Fig. S10 and Table S25).

Discussion

This detailed genetic map of MS is a powerful substrate for annotation and functional studies and provides a new level of understanding for the molecular events that contribute to MS susceptibility. Although the exact amount of MS's heritability varies given the data and method used, (*36-38*) we report that our findings can explain up to 48% of the heritability that can be estimated using large-scale GWAS data. It is clear that these events are widely distributed across the many different cellular components of both the innate and adaptive arms of the immune system: every major immune cell type is enriched for MS susceptibility genes. An important caveat is that many of the implicated molecular pathways, such as response to TNFα and type I

interferons, are repurposed in many different cell types, leading to an important ambiguity: is risk of disease driven by altered function of only one of the implicated cell types or are all of them contributing to susceptibility equally? This question highlights the important issue of the context in which these variants are exerting their effects. We have been thorough in our evaluation of available reference epigenomic data, but many different cell types and cell states remain to be characterized and could alter our summary. Further, inter-individual variability has not been established in such reference data that are typically produced from one or a handful of individuals; thus, this issue is better evaluated in the eQTL data where we have examined a range of samples and states in large numbers of subjects. Overall, while we have identified putative functional consequences for the identified MS variants, the functional consequence of most of these MS variants requires further investigation.

Even where a function is reported, further work is needed to demonstrate that the effect is the causal functional change. This is particularly true of the role of the CNS in MS susceptibility: we mostly have data at the level of the human cortex, a complex tissue with many different cell types, including resident microglia and a small number of infiltrating macrophage and lymphocytes. MS variants clearly influence gene expression in this tissue, and we must now: (1) resolve the implicated cell types and whether pathways shared with immune cells are having their MS susceptibility effect in the periphery or in the brain and; (2) more deeply identify additional functional consequences that may be present in only a subset of cells, such as microglia or activated astrocytes, that are obscured in the cortical tissue level profile. A handful of loci are intriguing in that they alter gene expression in the human cortex but not in the sampled immune cells; these MS susceptibility variants deserve close examination to resolve the important question of the extent to which the CNS is involved in disease onset. Thus, our study suggests that while MS is a disease whose origin may lie primarily within the peripheral immune compartment where dysregulation of all branches of the immune system leads to organ specific autoimmunity, there is subset of loci with a key role in directing the tissue specific autoimmune response. This is similar to our previous examination of ulcerative colitis, where we observed

enrichment of genetic variants mapping to colon tissue.(*7*) This view is consistent with our understanding of the mechanism of important MS therapies such as natalizumab and fingolimod that sequester pathogenic immune cell populations in the peripheral circulation to prevent episodes of acute CNS inflammation. It also has important implications as we begin to consider prevention strategies to block the onset of the disease by early targeting peripheral immune cells.

An important step forward in MS genetics, for a disease with a 3:1 preponderance of women being affected, is robust evidence for a susceptibility locus on the X chromosome. Although chromosome X associations cannot be the sole explanation for the preponderance of women among MS patients, the discovery of an MS locus on the X chromosome is an exciting first step towards understanding the genetic contributions of this strong sex bias. This result also highlights the need for additional, dedicated genetic studies of the sex chromosomes in MS as existing data have not been fully leveraged. (*39*) Future studies will also need to incorporate the interaction of the autosomal genome with factors that can affect the sex bias, e.g. hormones. (*40*)

This genomic map of MS – the genetic map and its integrated functional annotation - is a foundation on which the next generation of projects will be developed. It is an important substrate with which to further dissect the genetic architecture of MS by accounting for the contribution of sex, evaluating the possibility of interaction among loci and assessing other important factors such as heterogeneity of effects across human populations or certain subsets of patients given the heterogeneity of this disease. In the current study we have included individuals with either the relapsing remitting or the progressive form of MS as they are currently conceptualized to belong to the same disease spectrum. Further investigation may lead to the identification of variants that have an effect on the neurodegenerative component of MS, which is largely genetically distinct from MS susceptibility. (*41*) Beyond the characterization of the molecular events that trigger MS, this map will also inform the development of primary prevention strategies since we can leverage this information to identify the subset of individuals who are at greatest risk of developing MS. While insufficient by itself, an MS Genetic Risk Score has a role to play in guiding the management of the population of individuals “at risk” of MS (such as family members) when

deployed in combination with other measures of risk and biomarkers that capture intermediate phenotypes along the trajectory from health to disease.(*42*) We thus report an important milestone in the investigation of MS and share a roadmap for future work: the establishment of a map with which to guide the development of the next generation of studies with high-dimensional molecular data to explore both the initial steps of immune dysregulation across both the adaptive and innate arms of the immune system and secondly the translation of this auto-immune process to the CNS where it triggers a neurodegenerative cascade.

Materials and Methods

Detailed materials and methods are listed in the Supplementary Material (*9*). In brief, we analyzed genetic data from 15 GWAS of MS. For the autosomal non-MHC genome, we applied a partitioning approach to create regions of ±1Mbps around the most statistically significant SNP. Then we performed stepwise conditional analyses within each region to identify statistically independent effects (n=4,842). We replicated these effects in two large-scale replication cohorts: i) nine (9) data sets genotyped with the MS Replication Chip, and ii) eleven (11) data sets genotyped with the ImmunoChip. Chromosomes X and Y were analyzed jointly across all the data sets, i.e. the discovery and replication. The extended MHC region was also analyzed jointly across all data sets. We further imputed HLA class I and II alleles and corresponding amino acids. Statistically independent effects in the autosomal non-MHC genome were group into 4 categories post-replication: i) genome-wide effects (GW), ii) suggestive effects (S), iii) non-replicated (NR), and iv) no replication data (ND). Narrow sense heritability was estimated for various combinations of these effects, and the extended MHC region, to quantify the amount of the heritability our findings could explain. Next, we leveraged enrichment methods and tissue/cell reference data sets to characterize the potential involvement of the identified MS effects in the immune and central nervous system, at the tissue and cellular level. We developed an ensemble approach to prioritize genes putative associated with the identify effects, leveraging cell-specific eQTL studies, network approaches, and genomic annotations. Pathway analyses were performed

to characterize canonical pathways statistically enriched for the putative causal genes. Finally, we leveraged protein-protein interaction networks to quantify the degree of connectivity of the putative causal genes and identify new mechanisms of action.

References and Notes

1. 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).
2. P. L. De Jager *et al.*, Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. *Nat Genet* **41**, 776-782 (2009).
3. International Multiple Sclerosis Genetics Consortium *et al.*, Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet* **45**, 1353-1360 (2013).
4. International Multiple Sclerosis Genetics Consortium *et al.*, Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214-219 (2011).
5. N. A. Patsopoulos *et al.*, Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Ann Neurol* **70**, 897-912 (2011).
6. T. F. Andlauer *et al.*, Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation. *Sci Adv* **2**, e1501678 (2016).
7. K. K. Farh *et al.*, Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337-343 (2015).
8. M. T. Maurano *et al.*, Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-1195 (2012).
9. See supplementary materials.
10. A. Alcina *et al.*, Tag-SNP analysis of the GFI1-EVI5-RPL5-FAM69 risk locus for multiple sclerosis. *Eur J Hum Genet* **18**, 827-831 (2010).
11. A. Didonna *et al.*, A non-synonymous single-nucleotide polymorphism associated with multiple sclerosis risk affects the EVI5 interactome. *Hum Mol Genet* **24**, 7151-7158 (2015).
12. I. A. Hoppenbrouwers *et al.*, EVI5 is a risk gene for multiple sclerosis. *Genes Immun* **9**, 334-337 (2008).
13. D. Martin *et al.*, Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes. *Nat Struct Mol Biol* **18**, 708-714 (2011).
14. D. M. Fortner, R. G. Troy, D. A. Brow, A stem/loop in U6 RNA defines a conformational switch required for pre-mRNA splicing. *Genes Dev* **8**, 221-233 (1994).
15. P. Vaudin, R. Delanoue, I. Davidson, J. Silber, A. Zider, TONDU (TDU), a novel human protein related to the product of vestigial (vg) gene of Drosophila melanogaster interacts with vertebrate TEF factors and substitutes for Vg function in wing formation. *Development* **126**, 4807-4816 (1999).
16. L. Moutsianas *et al.*, Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nat Genet* **47**, 1107-1113 (2015).
17. N. A. Patsopoulos *et al.*, Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet* **9**, e1003926 (2013).
18. J. Yang *et al.*, Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565-569 (2010).
19. H. Shi, G. Kichaev, B. Pasaniuc, Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *Am J Hum Genet* **99**, 139-153 (2016).
20. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
21. J. Yang *et al.*, Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet* **43**, 519-525 (2011).

22. X. Hu *et al.*, Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. *Am J Hum Genet* **89**, 496-506 (2011).
23. K. Slowikowski, X. Hu, S. Raychaudhuri, SNPsea: an algorithm to identify cell types, tissues and pathways affected by risk loci. *Bioinformatics* **30**, 2496-2497 (2014).
24. P. Shooshtari, H. Huang, C. Cotsapas, Integrative genetic and epigenetic analysis uncovers regulatory mechanisms of autoimmune disease. *bioRxiv*, (2016).
25. S. L. Hauser *et al.*, B-cell depletion with rituximab in relapsing-remitting multiple sclerosis. *N Engl J Med* **358**, 676-688 (2008).
26. T. Raj *et al.*, Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* **344**, 519-523 (2014).
27. A. Pugliese *et al.*, The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. *Nat Genet* **15**, 293-297 (1997).
28. T. Stodberg *et al.*, Mutations in SLC12A5 in epilepsy of infancy with migrating focal seizures. *Nat Commun* **6**, 8038 (2015).
29. M. Puskarjov *et al.*, A variant of KCC2 from patients with febrile seizures impairs neuronal Cl- extrusion and dendritic spine formation. *EMBO Rep* **15**, 723-729 (2014).
30. G. Orozco *et al.*, Association of CD40 with rheumatoid arthritis confirmed in a large UK case-control study. *Ann Rheum Dis* **69**, 813-816 (2010).
31. G. Li *et al.*, Human genetics in rheumatoid arthritis guides a high-throughput drug screen of the CD40 signaling pathway. *PLoS Genet* **9**, e1003487 (2013).
32. C. Wallace *et al.*, Statistical colocalization of monocyte gene expression and genetic risk variants for type 1 diabetes. *Hum Mol Genet* **21**, 2815-2824 (2012).
33. T. H. Pers *et al.*, Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).
34. Ingenuity Pathways Analysis, www.ingenuity.com.
35. GeNets, http://apps.broadinstitute.org/genets.
36. P. A. Gourraud, H. F. Harbo, S. L. Hauser, S. E. Baranzini, The genetics of multiple sclerosis: an up-to-date review. *Immunol Rev* **248**, 87-103 (2012).
37. N. A. Patsopoulos, Genetics of Multiple Sclerosis: An Overview and New Directions. *Cold Spring Harb Perspect Med*, (2018).
38. S. Sawcer, R. J. Franklin, M. Ban, Multiple sclerosis genetics. *Lancet Neurol* **13**, 700-709 (2014).
39. Z. Xia *et al.*, Genes and Environment in Multiple Sclerosis project: A platform to investigate multiple sclerosis risk. *Ann Neurol* **79**, 178-189 (2016).
40. R. R. Voskuhl, S. M. Gold, Sex-related factors in multiple sclerosis susceptibility and progression. *Nat Rev Neurol* **8**, 255-263 (2012).
41. M. F. George *et al.*, Multiple sclerosis risk loci and disease severity in 7,125 individuals from 10 studies. *Neurol Genet* **2**, e87 (2016).
42. P. L. De Jager *et al.*, Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. *Lancet Neurol* **8**, 1111-1119 (2009).
43. C. M. Poser *et al.*, New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann Neurol* **13**, 227-231 (1983).
44. W. I. McDonald *et al.*, Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol* **50**, 121-127 (2001).
45. A. L. Price *et al.*, Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-909 (2006).
46. S. R. Browning, B. L. Browning, Population structure can inflate SNP-based heritability estimates. *Am J Hum Genet* **89**, 191-193; author reply 193-195 (2011).
47. X. Jia *et al.*, Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One* **8**, e64683 (2013).

48. A. T. Dilthey, L. Moutsianas, S. Leslie, G. McVean, HLA*IMP--an integrated framework for imputing classical HLA alleles from SNP genotypes. *Bioinformatics* **27**, 968-972 (2011).
49. D. Clayton, Testing for association on the X chromosome. *Biostatistics* **9**, 593-600 (2008).
50. D. Speed, G. Hemani, M. R. Johnson, D. J. Balding, Improved heritability estimation from genome-wide SNPs. *Am J Hum Genet* **91**, 1011-1021 (2012).
51. P. R. Loh *et al.*, Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284-290 (2015).
52. A. I. Su *et al.*, A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A* **101**, 6062-6067 (2004).
53. G. Hyatt *et al.*, Gene expression microarrays: glimpses of the immunological genome. *Nat Immunol* **7**, 686-691 (2006).
54. B. E. Bernstein *et al.*, The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**, 1045-1048 (2010).
55. Roadmap Epigenomics Consortium *et al.*, Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330 (2015).
56. S. John *et al.*, Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet* **43**, 264-268 (2011).
57. S. E. Sullivan *et al.*, Candidate-based screening via gene modulation in human neurons and astrocytes implicates FERMT2 in Abeta and TAU proteostasis. *Hum Mol Genet*, (2018).
58. Y. Zhang *et al.*, Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78**, 785-798 (2013).
59. M. Olah *et al.*, A transcriptomic atlas of aged human microglia. *Nat Commun* **9**, 539 (2018).
60. A. S. Lim *et al.*, Diurnal and seasonal molecular rhythms in human neocortex and their relation to Alzheimer's disease. *Nat Commun* **8**, 14931 (2017).
61. B. Ng *et al.*, An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat Neurosci* **20**, 1418-1426 (2017).
62. P. L. De Jager *et al.*, A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci Data* **5**, 180142 (2018).
63. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
64. W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127 (2007).
65. B. E. Stranger *et al.*, Population genomics of human gene expression. *Nat Genet* **39**, 1217-1224 (2007).
66. A. S. Dimas *et al.*, Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325**, 1246-1250 (2009).
67. K. Wang, M. Li, H. Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
68. Encode Project Consortium An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
69. C. M. Lill *et al.*, Genome-wide significant association with seven novel multiple sclerosis risk loci. *J Med Genet* **52**, 848-855 (2015).

Acknowledgments

We thank the Harvard Aging Brain Study (HABS - P01AG036694). We thank the Biorepository Facility and the Center for Genome Technology laboratory personnel (specifically Sandra West,

Simone Clarke, Daniela Martinez, and Patrice Whitehead) within the John P. Hussman Institute for Human Genomics at the University of Miami for centralized DNA handling and genotyping for this project. The IMSGC wishes to acknowledge William and Lois Edgerly, John and Elaine Carlos, Martha Crowninshield, William and Cindy Fowler, whose enduring commitments were critical in creation of the Consortium. We thank the volunteers from the Oxford Biobank (www.oxfordbiobank.org.uk) and the Oxford NIHR Bioresource for their participation. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. German Ministry for Education and Research, German Competence Network MS (BMBF KKNMS).

Funding: This investigation was supported in part by a Postdoctoral Fellowship from the National Multiple Sclerosis Society (FG 1938-A-1) and a Career Independence Award from the National Multiple Sclerosis Society (TA 3056-A-2) to Nikolaos A. Patsopoulos, and National Multiple Sclerosis Society award AP3758-A-16. Nikolaos A. Patsopoulos has been supported by Harvard NeuroDiscovery Center and an Intel Parallel Computing Center award. The US National Multiple Sclerosis Society (grants RG 4680-A-1) and the NIH/NINDS (grant R01NS096212). Till Andlauer was supported by the German Federal Ministry of Education and Research (BMBF) through the Integrated Network IntegraMent, under the auspices of the e:Med Programme (01ZX1614J). Swedish Medical Research Council; Swedish Research Council for Health, Working Life and Welfare, Knut and Alice Wallenberg Foundation, AFA insurance, Swedish Brain Foundation, the Swedish Association for Persons with Neurological Disabilities. This study makes use of data generated as part of the Wellcome Trust Case Control Consortium 2 project (085475/B/08/Z and 085475/Z/08/Z), including data from the British 1958 Birth Cohort DNA collection (funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02) and the UK National Blood Service controls (funded by the Wellcome Trust). The study was supported by the Cambridge NIHR Biomedical Research Centre, UK Medical Research Council (G1100125) and the UK MS society (861/07). NIH/NINDS: R01 NS049477, NIH/NIAID: R01 AI059829, NIH/NIEHS: R01 ES0495103. Research Council of

Norway grant 196776 and 240102. NINDS/NIH R01NS088155. Oslo MS association and the Norwegian MS Registry and Biobank and the Norwegian Bone Marrow Registry. Research Council KU Leuven, Research Foundation Flanders. AFM, AFM-Généthon, CIC, ARSEP, ANR-10-INBS-01 and ANR-10-IAIHU-06. Research Council KU Leuven, Research Foundation Flanders. Inserm ATIP-Avenir Fellowship and Connect-Talents Award. German Ministry for Education and Research, German Competence Network MS (BMBF KKNMS). Dutch MS Research Foundation. TwinsUK is funded by the Wellcome Trust, Medical Research Council, European Union, the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. The recall process was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre Programme. Italian Foundation of Multiple Sclerosis (FISM grants, Special Project "Immunochip" 2011/R/1, 2015/R/10)). NMSS (RG 4680A1/1). MultipleMS EU project. Lundbeck Foundation and Benzon Foundation for support (THP). This research was supported by grants from the Danish Multiple Sclerosis Society, the Danish Council for Strategic Research [grant number 2142-08-0039], Novartis, Biogen Biogen (Denmark) A/S, and the Sofus Carl Emil Friis og Hustru Olga Doris Friis Foundation and the Foundation for Research in Neurology. The Observatoire Français de la Sclérose en Plaques (OFSEP) is supported by a grant provided by the French State and handled by the "Agence Nationale de la Recherche," within the framework of the "Investments for the Future" program, under the reference ANR-10-COHO-002, by the Eugène Devic EDMUS Foundation against multiple sclerosis and by the ARSEP Foundation.

Competing interests: The authors declare no competing interests

Data availability: The genome wide associations studies used in the discovery phase are available in the following repositories: i) dbGAP: phs000275.v1.p1, phs000139.v1.p1, phs000294.v1.p1, and phs000171.v1.p1, and ii) European Genome-phenome Archive database: EGAD00000000120, EGAD00000000022, and EGAD00000000021. The ANZGENE GWAS data is available via request to the ANZGENE Consortium; a direct request can be made via

MSRA.org.au with an application regarding MS-related research: the request should state the purpose of the study and a list of the data being requested. The request is then reviewed by the steering committee of ANZGENE for scientific merit. The Rotterdam MS data can be requested via e-mailing k.kreft@erasmusmc.nl; a formal application regarding MS-related research is required. Rotterdam Study control data can be made available to interested researchers upon request. Requests can be directed to data manager Frank J.A. van Rooij (f. vanrooij@erasmusmc.nl) or visit the following website for more information http://www.ergo-onderzoek.nl/wp/contact. We are unable to place data in a public repository due to legal and ethical restraints. Sharing of individual participant data was not included in the informed consent of the Rotterdam MS study, and there is potential risk of revealing participants' identities as it is not possible to completely anonymize the data. The genetic data from multiple sclerosis (MS) case and controls recruited through Kaiser Permanente Division of Research and UC Berkeley cannot be shared publicly because their use, as per informed consent and Institutional Review Board approval at each research site, is restricted to MS-research only; data are available from the Institutional Data Access / Ethics Committee at UC Berkeley (contact Richard Harris, rharris@berkeley.edu, for researchers who meet the criteria for access to confidential data. Please reference the manuscript title and corresponding author in your communication). Corresponding summary statistics for these three GWAS studies (ANZGENE, Rotterdam, and Berkeley) are available upon request. The MS Chip and ImmunoChip data are available in the respective EGA accession codes: EGAS00001003216 and EGAS00001003219. The ImmVar data are available in GEO: GSE56035. The MS PBMC data are available in GEO: GSE16214. Human Gene Atlas: http://snpsea.readthedocs.io/en/latest/data.html#geneatlas2004-gct-gz. ImmGen: http://snpsea.readthedocs.io/en/latest/data.html#immgen2012-gct-gz. The brain related data are available in Synapse: https://www.synapse.org/#!Synapse:syn2580853/wiki/409844. The list of putative associated MS genes is available for public access in GeNets. (https://apps.broadinstitute.org/genets).

Fig. legends:

Fig. 1. The genetic map of multiple sclerosis. The circos plot displays the 4,842 prioritized autosomal non-MHC effects and the associations in chromosome X. Joint analysis (discovery and replication) p-values are plotted as lines (fixed effects inverse-variance meta-analysis). The green inner layer displays genome-wide significance (p-value<$5x10^{-8}$), the blue inner layer suggestive p-values ($1x10^{-5}$<p-value>$5x10^{-8}$), and the grey p-values > $1x10^{-5}$. Each line in the inner layers represents one effect. 200 autosomal non-MHC and one in chromosome X genome-wide effects are listed. The vertical lines in the inner layers represent one effect and the respective color displays the replication status (see main text and Online Methods): green (genome-wide), blue (potentially replicated), red (non-replicated). 551 prioritized genes are plotted on the outer surface. The inner circle space includes protein-protein interactions (PPI) between genome-wide genes (green), and genome-wide genes and potentially replicated genes (blue) that are identified as candidates using protein-protein interaction networks (see main text) (*9*).

Fig. 2. Multiple independent effects in the *EVI5* locus and chromosome X associations. A) Regional association plot of the *EVI5* locus. Discovery p-values (fixed effects inverse-variance meta-analysis) are displayed. The layer tagged "Marginal" plots the associations of the marginal analysis, with most statistically significant SNP being rs11809700 (OR_T=1.16; p-value= $3.51x10^{-15}$). The "Step 1" plots the associations conditioning on rs11809700; rs12133753 is the most statistically significant SNP (OR_C=1.14; p-value= $8.53x10^{-09}$). "Step 2" plots the results conditioning on rs11809700 and rs12133753, with rs1415069 displaying the lowest p-value (OR_G=1.10; p-value= $4.01x10^{-5}$). Finally, "Step 3" plots the associations conditioning on rs11809700, rs12133753, and rs1415069, identifying rs58394161 as the most-statistically significant SNP (OR_C=1.10; p-value= $8.63x10^{-4}$). All 4 SNPs reached genome-wide significance

in the respective joint, discovery plus replication, analyses (Table S7). Each of the independent 4 SNPs, i.e. lead SNPs, are highlighted using a triangle in the respective layer. **B**) Regional association plot for the genome-wide chromosome X variant. Joint analysis p-values (fixed effects inverse-variance meta-analysis) are displayed. Linkage disequilibrium, in terms of r^2 based on the 1000 Genomes European panel, is indicated using a combination of color grade and symbol size (see legend for details). All positions are in human genome 19.

Fig. 3. Independent associations in the major histocompatibility locus. Regional association plot in the MHC locus. Only genome-wide statistically independent effects are listed. The order of variants in the X-axis represents the order these were identified. The size of the circle represents different values of –log10(p-value) (fixed effects inverse-variance meta-analysis). Different colors are used to depict class I, II, III, and non-HLA effects. Y-axis displays million base pairs.

Fig. 4. Heritability partitioning. Proportion of the overall narrow-sense heritability under the liability model (~19.2%) explained by different genetic components. (A) The overall heritability is partitioned in the super extended MHC (SE MHC), the 1,962 Regions that include all SNPs with p-value<0.05 (Regions; fixed effects inverse-variance meta-analysis), and the rest of genome with p-values>0.05 (Non-associated regions). (B) The Regions are further partitioned to the seemingly statistically independent effects (Prioritized) and the residual (Non-prioritized). (C) The Prioritized component is partitioned based on the replication knowledge to genome-wide effects (GW), suggestive (S), non-replicated (ND), and no data (ND). The lines connecting the pie charts depict the component that is partitioned. All values are estimated using the discovery data-sets (n= 4,802 cases and 26,703 controls).

Fig. 5. Tissue and cell type enrichment analyses. (A) Gene Atlas tissues and cell types gene expression enrichment. (B) DNA hypersensitivity sites (DHS) enrichment for tissues and cell types from the NIH Epigenetic Roadmap. Rows are sorted from immune cells/tissues to central nervous system related ones. Both X axes display –log10 of Benjamini & Hochberg p-values (false discovery rate).

Fig. 6. Dissection of cortical RNAseq data. In (A), we present a heatmap of the results of our analysis assessing whether a cortical eQTL is likely to come from one of the component cell types of the cortex: neurons, oligodendrocytes, endothelial cells, microglia and astrocytes (in rows). Each column presents results for one of the MS brain eQTLs. The color scheme relates to the p-value of the interaction term (linear regression), with red denoting a more extreme result. (B) We present the same results in a different form, comparing results of assessing for interaction with neuronal proportion (y axis) and microglial proportion (x-axis): the *SLC12A5* eQTL is significantly stronger when accounting for neuronal proportion, and *CLECL1* is significantly stronger when accounting for microglia. The Bonferroni-corrected threshold of significance is highlighted by the dashed line. (C) Locus view of the *SLC12A5/CD40* locus, illustrating the distribution of MS susceptibility and the *SLC12A5* brain eQTL in a segment of chromosome 20 (x axis); the y axis presents the p-value of association with MS susceptibility (top panel; fixed effects inverse-variance meta-analysis) or *SLC12A5* RNA expression (bottom panel; linear regression). The lead MS SNP is denoted by a triangle, other SNPs are circles, with the intensity of the red color denoting the strength of LD with the lead MS SNP in both panels. (D) Here we plot the level of expression, transcriptome-wide, for each measured gene in our cortical RNAseq dataset (n=455)(y-axis) and purified human microglia (n=10)(x-axis) from the same cortical region. In blue, we highlight those genes with > 4 fold increased expression in microglia relative to bulk cortical tissue and are expressed at a reasonable level in microglia. Each dot is one gene. Gray dots denote the 551 putative MS genes from our integrated analysis. *SLC12A5* and *CLECL1*

are highlighted in red; in blue, we highlight a selected subset of the MS genes – many of them well-validated – which are enriched in microglia. For clarity, we did not include all of the MS genes that fall in this category.

Supplementary Materials:

Materials and Methods

Supplementary Text

Consortium Memberships

Figs. S1 to S40

Tables S1 to S53

Supplementary text files with detailed *cis*-eQTL results (n=4)

References (42–69)

Extended Author list and Affiliations:

Nikolaos A Patsopoulos,[1-4] Sergio E Baranzini,[5] Adam Santaniello,[5] Parisa Shoostari,[4,6-7] Chris Cotsapas,[4,6-7] Garrett Wong,[1,3] Ashley H Beecham,[8] Tojo James,[9] Joseph Replogle,[10] Ioannis S Vlachos,[1,3-4] Cristin McCabe,[4] Tune H Pers,[11] Aaron Brandes,[4] Charles White,[4,10] Brendan Keenan,[12] Maria Cimpean,[10] Phoebe Winn,[10] Ioannis-Pavlos Panteliadis,[1,4] Allison Robbins,[10] Till FM Andlauer,[13-15] Onigiusz Zarzycki,[1,4] Bénédicte Dubois,[16] An Goris,[16] Helle Bach Søndergaard,[17] Finn Sellebjerg,[17] Per Soelberg Sorensen,[17] Henrik Ullum,[18] Lise Wegner Thoerner,[18] Janna Saarela,[19] Isabelle Cournu-Rebeix,[20] Vincent Damotte,[20,21] Bertrand Fontaine,[20, 22] Lena Guillot-Noel,[20] Mark Lathrop,[23-25] Sandra Vukusic,[26-28] Achim Berthele, [14-15] Viola Pongratz, [14-15] Dorothea Buck, [14-15] Christiane Gasperi, [14-15] Christiane Graetz, [15,29] Verena Grummel, [14-15] Bernhard Hemmer,[14-15,30,] Muni Hoshi, [14-15] Benjamin Knier, [14-15] Thomas Korn,[14-15,30] Christina M Lill, [15,31-32] Felix Luessi, [15,31] Mark Mühlau, [14-15] Frauke Zipp, [15,31] Efthimios Dardiotis,[33] Cristina Agliardi,[34] Antonio Amoroso,[35] Nadia Barizzone,[36] Maria D Benedetti,[37-38] Luisa Bernardinelli,[39] Paola Cavalla,[40] Ferdinando Clarelli,[41] Giancarlo Comi,[41-42] Daniele Cusi,[43] Federica Esposito,[41,44] Laura Ferrè,[44] Daniela Galimberti,[45-46] Clara Guaschino,[41,44] Maurizio A Leone,[47] Vittorio Martinelli,[44] Lucia Moiola,[44] Marco Salvetti,[48-49] Melissa Sorosina,[41] Domizia Vecchio,[50] Andrea Zauli,[41] Silvia Santoro,[41] Nicasio Mancini,[51] Miriam Zuccalà,[52] Julia Mescheriakova,[53] Cornelia van Duijn,[53-54] Steffan D Bos,[55] Elisabeth G Celius,[55-56] Anne

Spurkland,[57] Manuel Comabella,[58] Xavier Montalban,[58] Lars Alfredsson,[59] Izaura L Bomfim,[60] David Gomez-Cabrero,[60] Jan Hillert,[60] Maja Jagodic,[60] Magdalena Lindén,[60] Fredrik Piehl,[60] Ilijas Jelčić,[61-62] Roland Martin,[61-62] Mirela Sospedra,[61-62] Amie Baker,[63] Maria Ban,[64] Clive Hawkins,[64] Pirro Hysi,[65] Seema Kalra,[66] Fredrik Karpe,[66] Jyoti Khadake,[67] Genevieve Lachance,[65] Paul Molyneux,[65] Matthew Neville,[66] John Thorpe,[68] Elizabeth Bradshaw,[10] Stacy J Caillier,[5] Peter Calabresi,[69] Bruce AC Cree,[5] Anne Cross,[70] Mary Davis,[71] Paul WI de Bakker,[2-4†] Silvia Delgado,[72] Marieme Dembele,[69] Keith Edwards,[73] Kate Fitzgerald,[69] Irene Y Frohlich,[10] Pierre-Antoine Gourraud,[5,74] Jonathan L Haines,[75] Hakon Hakonarson,[76-77] Dorlan Kimbrough,[78] Noriko Isobe,[5,79] Ioanna Konidari,[8] Ellen Lathi,[80] Michelle H Lee,[10] Taibo Li,[81] David An,[81] Andrew Zimmer,[81] Lohith Madireddy,[5] Clara P Manrique,[8] Mitja Mitrovic,[4,6-7] Marta Olah,[10] Ellis Patrick,[10,82-83] Margaret A Pericak-Vance,[8] Laura Piccio,[69] Cathy Schaefer,[84] Howard Weiner,[85] Kasper Lage,[80] ANZgene, IIBDGC, WTCCC2, Alastair Compston,[62] David Hafler, [4,86] Hanne F Harbo,[54-55] Stephen L Hauser,[5] Graeme Stewart,[87] Sandra D'Alfonso,[88] Georgios Hadjigeorgiou,[33] Bruce Taylor,[89] Lisa F Barcellos,[90] David Booth,[91] Rogier Hintzen,[92] Ingrid Kockum,[9] Filippo Martinelli-Boneschi,[41-42] Jacob L McCauley,[8] Jorge R Oksenberg,[5] Annette Oturai,[16] Stephen Sawcer,[62] Adrian J Ivinson,[93] Tomas Olsson,[9] Philip L De Jager,[4,10]

1. Systems Biology and Computer Science Program, Ann Romney Center for Neurological Diseases, Department of Neurology, Brigham & Women's Hospital, Boston, 02115 MA, USA

2. Division of Genetics, Department of Medicine, Brigham & Women's Hospital, Harvard Medical School, Boston, MA, USA.

3. Harvard Medical School, Boston, MA 02115, USA.

4. Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA, USA

5. Department of Neurology, University of California at San Francisco, Sandler Neurosciences Center, 675 Nelson Rising Lane, San Francisco, CA 94158, USA.

6. Department of Neurology, Yale University School of Medicine, New Haven, CT 06520, USA

7. Department of Genetics, Yale School of Medicine, New Haven, CT 06520, USA.

8. John P. Hussman Institute for Human Genomics, University of Miami, Miller School of Medicine, Miami, FL 33136, USA.

9. Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden.

10. Center for Translational & Computational Neuroimmunology, Multiple Sclerosis Center, Department of Neurology, Columbia University Medical Center, New York, NY, USA.

11. The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, 2100, Denmark.

12. Center for Sleep and Circadian Neurobiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA.

13. Max Planck Institute of Psychiatry, 80804 Munich, Germany.

14. Department of Neurology, Klinikum rechts der Isar, Technical University of Munich, 81675 Munich, Germany.

15. German competence network for multiple sclerosis.

16. KU Leuven Department of Neurosciences, Laboratory for Neuroimmunology, Herestraat 49 bus 1022, 3000 Leuven, Belgium.

17. Danish Multiple Sclerosis Center, Department of Neurology, Rigshospitalet, University of Copenhagen, Section 6311, 2100 Copenhagen, Denmark.

18. Department of Clinical Immunology, Rigshospitalet, University of Copenhagen, Section 2082, 2100 Copenhagen, Denmark.

19. Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland.

20. ICM-UMR 1127, INSERM, Sorbonne University, Hôpital Universitaire Pitié-Salpêtrière 47 Boulevard de l'Hôpital, F-75013 Paris.

21. UMR1167 Université de Lille, Inserm, CHU Lille, Institut Pasteur de Lille.

22. CRM-UMR974 Department of Neurology Hôpital Universitaire Pitié-Salpêtrière 47 Boulevard de l'Hôpital F-75013 Paris.

23. Commissariat à l'Energie Atomique, Institut Genomique, Centre National de Génotypage, Evry, France.

24. Fondation Jean Dausset - Centre d'Etude du Polymorphisme Humain, Paris, France.

25. McGill University and Genome Quebec Innovation Center, Montreal, Canada.

26. Hospices Civils de Lyon, Service de Neurologie, sclérose en plaques, pathologies de la myéline et neuro-inflammation, F-69677 Bron, France.

27. Observatoire Français de la Sclérose en Plaques, Centre de Recherche en Neurosciences de Lyon, INSERM 1028 et CNRS UMR 5292, F-69003 Lyon, France.

28. Université de Lyon, Université Claude Bernard Lyon 1, F-69000 Lyon, France; Eugène Devic EDMUS Foundation against multiple sclerosis, F-69677 Bron, France.

29. Focus Program Translational Neuroscience (FTN), Rhine Main Neuroscience Network (rmn2), Johannes Gutenberg University-Medical Center, Mainz, Germany.

30. Munich Cluster for Systems Neurology (SyNergy), 81377 Munich, Germany.

31. Department of Neurology, Focus Program Translational Neuroscience (FTN), and Immunology (FZI), Rhine-Main Neuroscience Network (rmn2), University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany.

32. Genetic and Molecular Epidemiology Group, Institute of Neurogenetics, University of Luebeck, Luebeck, Germany.

33. Neurology Dpt, Neurogenetics Lab, University Hospital of Larissa, Greece.

34. Laboratory of Molecular Medicine and Biotechnology, Don C. Gnocchi Foundation ONLUS, IRCCS S. Maria Nascente, Milan, Italy.

35. Department of Medical Sciences, Torino University, Turin, Italy.

36. Department of Health Sciences and Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara, Italy.

37. Centro Regionale Sclerosi Multipla, Neurologia B, AOUI Verona, Italy.

38. Fondazione IRCCS Cà Granda, Ospedale Maggiore Policlinico, Italy.

39. Medical Research Council Biostatistics Unit, Robinson Way, Cambridge CB2 0SR, UK.

40. MS Center, Department of Neuroscience, A.O. Città della Salute e della Scienza di Torino & University of Turin, Torino, Italy.

41. Laboratory of Human Genetics of Neurological complex disorder, Institute of Experimental Neurology (INSPE), Division of Neuroscience, San Raffaele Scientific Institute, Via Olgettina 58, 20132, Milan, Italy.

42. Department of Biomedical Sciences for Health, University of Milan, Milan, Italy.

43. University of Milan, Department of Health Sciences, San Paolo Hospital and Filarete Foundation, viale Ortles 22/4, 20139 Milan, Italy.

44. Department of Neurology, Institute of Experimental Neurology (INSPE), Division of Neuroscience, San Raffaele Scientific Institute, Via Olgettina 58, 20132, Milan, Italy.

45. Neurology Unit, Dept. of Pathophysiology and Transplantation, University of Milan, , Dino Ferrari Center, Milan, Italy.

46. Fondazione IRCCS Ca' Granda, Ospedale Policlinico, Milan, Italy.

47. Fondazione IRCCS Casa Sollievo della Sofferenza, Unit of Neurology, San Giovanni Rotondo (FG), Italy.

48. Center for Experimental Neurological Therapy (CENTERS), Neurology and Department of Neurosciences, Mental Health and Sensory Organs, Università La Sapienza, Roma, Italy.

49. IRCCS Istituto Neurologico Mediterraneo (INM) Neuromed, Pozzilli, Italy.

50. Department of Neurology, Ospedale Maggiore, Novara, Italy.

51. Laboratory of Microbiology and Virology, University Vita-Salute San Raffaele, Hospital San Raffaele, Milan, Italy.

52. Department of Health Sciences and Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara, Italy.

53. Department of Neurology, Erasmus MC, Rotterdam, Netherlands.

54. Nuffield Department of Population Health, Big Data Institute, University of Oxford, Li Ka Shing Centre for Health Information and Discovery, Old Road Campus, Oxford OX3 7LF, UK.

55. Department of Neurology, Institute of Clinical Medicine, University of Oslo, Norway.

56. Department of Neurology, Oslo University Hospital, Oslo, Norway.

57. Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway.

58. Servei de Neurologia-Neuroimmunologia, Centre d'Esclerosi Múltiple de Catalunya (Cemcat), Institut de Recerca Vall d'Hebron (VHIR), Hospital Universitari Vall d'Hebron, Spain.

59. Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.

60. Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden.

61. Neuroimmunology and MS Research (nims), Neurology Clinic, University Hospital Zurich, Frauenklinikstrasse 26, 8091 Zurich, Switzerland .

62. Department of Neuroimmunology and MS Research, Neurology Clinic, University Hospital Zürich, Frauenklinikstrasse 26, 8091 Zürich, Switzerland.

63. University of Cambridge, Department of Clinical Neurosciences, Addenbrooke's Hospital, BOX 165, Hills Road, Cambridge CB2 0QQ, UK.

64. Keele University Medical School, University Hospital of North Staffordshire, Stoke-on-Trent ST4 7NY, UK.

65. Department of Twin Research and Genetic Epidemiology, King's College London, London, SE1 7EH, UK.

66. NIHR Oxford Biomedical Research Centre, Diabetes and Metabolism Theme, OCDEM, Churchill Hospital, Oxford UK.

67. NIHR BioResource, Box 299,University of Cambridge and Cambridge University Hospitals NHS Foundation Trust Hills Road, Cambridge CB2 0QQ, UK.

68. Department of Neurology, Peterborough City Hospital, Edith Cavell Campus, Bretton Gate, Peterborough PE3 9GZ, UK.

69. Department of Neurology, Johns Hopkins University School of medicine, Baltimore MD.

70. Multiple sclerosis center, Department of neurology, School of medicine, Washington University St Louis, St Louis MO.

71. Center for Human Genetics Research, Vanderbilt University Medical Center, 525 Light Hall, 2215 Garland Avenue, Nashville, TN 37232, USA.

72. Multiple Sclerosis Division, Department of Neurology, University of Miami, Miller School of Medicine, Miami, FL 33136, USA.

73. MS Center of Northeastern NY 1205 Troy Schenectady Rd, Latham, NY 12110.

74. Université de Nantes, INSERM, Centre de Recherche en Transplantation et Immunologie, UMR 1064, ATIP-Avenir, Equipe 5, Nantes, France.

75. Population & Quantitative Health Sciences, Department of Epidemiology and Biostatistics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-4945 USA.

76. Center for Applied Genomics, The Children's Hospital of Philadelphia, 3615 Civic Center Blvd., Philadelphia, PA 19104, USA.

77. Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia PA, USA.

78. Department of Neurology, Brigham & Women's Hospital, Boston, 02115 MA, USA.

79. Departments of Neurology and Neurological Therapeutics, Neurological Institute, Graduate School of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka City, Fukuoka 812-8582 Japan.

80. The Elliot Lewis Center, 110 Cedar St, Wellesley MA, 02481, USA.

81. Broad Institute of Harvard University and MIT, Cambridge, 02142 MA, USA.

82. School of Mathematics and Statistics, University of Sydney, Sydney, NSW 2006, Australia

83. Westmead Institute for Medical Research, University of Sydney, Westmead, NSW 2145, Australia.

84. Kaiser Permanente Division of Research, Oakland, CA, USA.

85. Ann Romney Center for Neurological Diseases, Department of Neurology, Brigham & Women's Hospital, Boston, 02115 MA, USA.

86. Departments of Neurology and Immunobiology, Yale University School of Medicine, New Haven, CT 06520, USA.

87. Westmead Millennium Institute, University of Sydney, New South Wales, Australia.

88. Department of Health Sciences and Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara, Italy.

89. Menzies Research Institute Tasmania, University of Tasmania, Australia.

90. UC Berkeley School of Public Health and Center for Computational Biology, USA.

91. Westmead Millennium Institute, University of Sydney, New South Wales, Australia.

92. Department of Neurology and Department of Immunology, Erasmus MC, Rotterdam, Netherlands.

93. UK Dementia Research Institute, University College London, Gower Street, London WC1E 6BT, UK.

† Current address: Vertex Pharmaceuticals, 50 Northern Avenue, Boston, MA 02210, USA.

Expanded list of group authorships (to not be included as individual authors)

ANZgene Membership

Rodney J Scott [1], Jeannette Lechner-Scott [1], Rod Lea [1], Pablo Moscato [1], David R Booth [2], Graeme J Stewart [2], Stephen Vucic [2], Grant Parnell [2], Michael Barnett [3], Deborah Mason [4], Lyn Griffiths [5], Simon Broadley [6], Lotti Tajouri [7], Alan Baxter [8], Mark Slee [9], Bruce V Taylor [10], Jac Charlesworth [10], Trevor J Kilpatrick [11], Justin Rubio [11], Vilija Jokubaitis [12], James Wiley [11], Helmut Butzkueven [12], Stephen Leslie [11], Allan Motyer [11], Jim Stankovich [11], William M Carroll [13], Allan G Kermode [13], Marzena Pedrini [13].

Hunter Medical Research Institute, University of Newcastle NSW Australia

1 Westmead Institute for Medical Research, University of Sydney NSW Australia

2 Brain and Mind Centre, University of Sydney NSW Australia

3 Canterbury District Health Board, Christchurch, New Zealand

4 Griffith Institute of Health and Medical Research, Griffith University, Gold Coast QLD Australia

5 School of Medicine, Griffith University, Gold Coast QLD Australia

6 Bond University, Gold Coast QLD Australia

7 James Cook University, Townsville, QLD Australia

8 Flinders University, SA Australia

9 Menzies Research Institute Tasmania, University of Tasmania Australia

10 University of Melbourne, VIC Australia

11 Monash University, VIC Australia

12 Sir Charles Gairdner Hospital, Perth, WA Australia

International IBD Genetics Consortium (IIBDGC) contributing members

Murray Barclay[1], Laurent Peyrin-Biroulet[2], Mathias Chamaillard[3], Jean-Frederick Colombe[4], Mario Cottone[5], Anthony Croft[6], Renata D'Incà[7], Jonas Halfvarson[8,9], Katherine Hanigan[6], Paul Henderson[10,11], Jean-Pierre Hugot[12,13], Amir Karban[14], Nicholas A Kennedy[15], Mohammed Azam Khan[16], Marc Lémann[17], Arie Levine[18], Dunecan Massey[19], Monica Milla[20], Grant W Montgomery[21], Sok Meng Evelyn Ng[22], Ioannis Oikonomou[22], Harald Peeters[23], Deborah D. Proctor[22], Jean-Francois Rahier[24], Rebecca Roberts[2], Paul Rutgeerts[25], Frank Seibold[26], Laura Stronati[27], Kirstin M Taylor[28], Leif Törkvist[29], Kullak Ublick[30], Johan Van Limbergen[31], Andre Van Gossum[32], Morten H. Vatn[33], Hu Zhang[20], Wei Zhang[22], Australia and New Zealand IBDGC*, Belgium Genetic Consortium†, Initiative on Crohn and Colitis, NIDDK IBDGC‡, United Kingdom IBDGC, Wellcome Trust Case Control Consortium§

1 Department of Medicine, University of Otago, Christchurch, New Zealand.

2 Gastroenterology Unit, INSERM U954, Nancy University and Hospital, France.

3 INSERM, U1019, Lille, France.

4 Univ Lille Nord de France, CHU Lille and Lille-2 University,Gastroenterology Unit, France.

5 Division of Internal Medicine, Villa Sofia-V. Cervello Hospital, University of Palermo, Palermo, Italy.

6 Inflammatory Bowel Diseases, Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia.

7 Department of Surgical and Gastroenterological Sciences, University of Padua, Padua, Italy.

8 Department of Medicine, Örebro University Hospital, Örebro, Sweden.

9 School of Health and Medical Sciences, Örebro University, Örebro, Sweden.

10 Royal Hospital for Sick Children, Paediatric Gastroenterology and Nutrition, Edinburgh, UK.

11 Child Life and Health, University of Edinburgh, Edinburgh, UK.

12 INSERM U843, Paris, France.

13 Univ-Paris Diderot Sorbonne Paris-Cité, Paris France.

14 Department of Gastroenterology, Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel.

15 Gastrointestinal Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK.

16 Genetic Medicine, MAHSC, University of Manchester, Manchester, UK.

17 Université Paris Diderot, GETAID group, Paris, France.

18 Pediatric Gastroenterology Unit, Wolfson Medical Center and Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel.

19 Inflammatory Bowel Disease Research Group, Addenbrooke's Hospital, University of Cambridge, Cambridge, UK.

20 Azienda Ospedaliero Universitaria (AOU) Careggi, Unit of Gastroenterology SOD2, Florence, Italy.

21 Molecular Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia.

22 Department of Internal Medicine, Section of Digestive Diseases, Yale School of Medicine, New Haven, Connecticut, USA.

23 Dept Gastroenterology - University hospital Gent - De Pintelaan - 9000 Gent Belgium.

24 Dept Gastroenterology - UCL Mont Godinne Belgium.

25 Division of Gastroenterology, University Hospital Gasthuisberg, Leuven, Belgium.

26 University of Bern, Division of Gastroenterology, Inselspital, Bern,

Membership of Wellcome Trust Case Control Consortium 2 (WTCCC2)

Management Committee

Peter Donnelly (Chair)[1,2], Ines Barroso (Deputy Chair)[3], Jenefer M Blackwell[4, 5], Elvira Bramon[6], Matthew A Brown[7], Juan P Casas[8], Aiden Corvin[9], Panos Deloukas[3], Audrey Duncanson[10], Janusz Jankowski[11], Hugh S Markus[12], Christopher G Mathew[13], Colin NA Palmer[14], Robert Plomin[15], Anna Rautanen[1], Stephen J Sawcer[16], Richard C Trembath[13], Ananth C Viswanathan[17], Nicholas W Wood[18]

Data and Analysis Group

Chris C A Spencer[1], Gavin Band[1], Céline Bellenguez[1], Colin Freeman[1], Garrett Hellenthal[1], Eleni Giannoulatou[1], Matti Pirinen[1], Richard Pearson[1], Amy Strange[1], Zhan Su[1], Damjan Vukcevic[1], Peter Donnelly[1,2]

DNA, Genotyping, Data QC and Informatics Group

Cordelia Langford[3], Sarah E Hunt[3], Sarah Edkins[3], Rhian Gwilliam[3], Hannah Blackburn[3], Suzannah J Bumpstead[3], Serge Dronov[3], Matthew Gillman[3], Emma Gray[3], Naomi Hammond[3], Alagurevathi Jayakumar[3], Owen T McCann[3], Jennifer Liddle[3], Simon C Potter[3], Radhi Ravindrarajah[3], Michelle Ricketts[3], Matthew Waller[3], Paul Weston[3], Sara Widaa[3], Pamela Whittaker[3], Ines Barroso[3], Panos Deloukas[3].

Publications Committee

Christopher G Mathew (Chair)[13], Jenefer M Blackwell[4,5], Matthew A Brown[7], Aiden Corvin[9], Chris C A Spencer[1]

1 Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK; 2 Dept Statistics, University of Oxford, Oxford OX1 3TG, UK; 3 Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; 4 Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, 100 Roberts Road, Subiaco, Western Australia 6008; 5 Cambridge Institute for Medical Research, University of Cambridge School of Clinical Medicine, Cambridge CB2 0XY, UK; 6 Department of Psychosis Studies, NIHR Biomedical Research Centre for Mental Health at the Institute of Psychiatry, King's College London and The South London and Maudsley NHS Foundation Trust, Denmark Hill, London SE5 8AF, UK; 7 University of Queensland Diamantina Institute, Brisbane, Queensland, Australia; 8 Dept Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT and Dept Epidemiology and Public Health, University College London WC1E 6BT, UK; 9 Neuropsychiatric Genetics Research Group, Institute of Molecular Medicine, Trinity College Dublin, Dublin 2, Eire; 10 Molecular and Physiological Sciences, The Wellcome Trust, London NW1 2BE; 11 Department of Oncology, Old Road Campus, University of Oxford, Oxford OX3 7DQ, UK , Digestive Diseases Centre, Leicester Royal Infirmary, Leicester LE7 7HH, UK and Centre for Digestive Diseases, Queen Mary University of London, London E1 2AD, UK; 12 Clinical Neurosciences, St George's University of London, London SW17 0RE; 13 King's College London Dept Medical and Molecular Genetics, King's Health Partners, Guy's Hospital, London SE1 9RT, UK; 14 Biomedical Research Centre, Ninewells Hospital and Medical School, Dundee DD1 9SY, UK; 15 King's College London Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Denmark Hill, London SE5 8AF, UK; 16

University of Cambridge Dept Clinical Neurosciences, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK; 17 NIHR Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London EC1V 2PD, UK; 18 Dept Molecular Neuroscience, Institute of Neurology, Queen Square, London WC1N 3BG, UK.

Figs:

Fig. 1

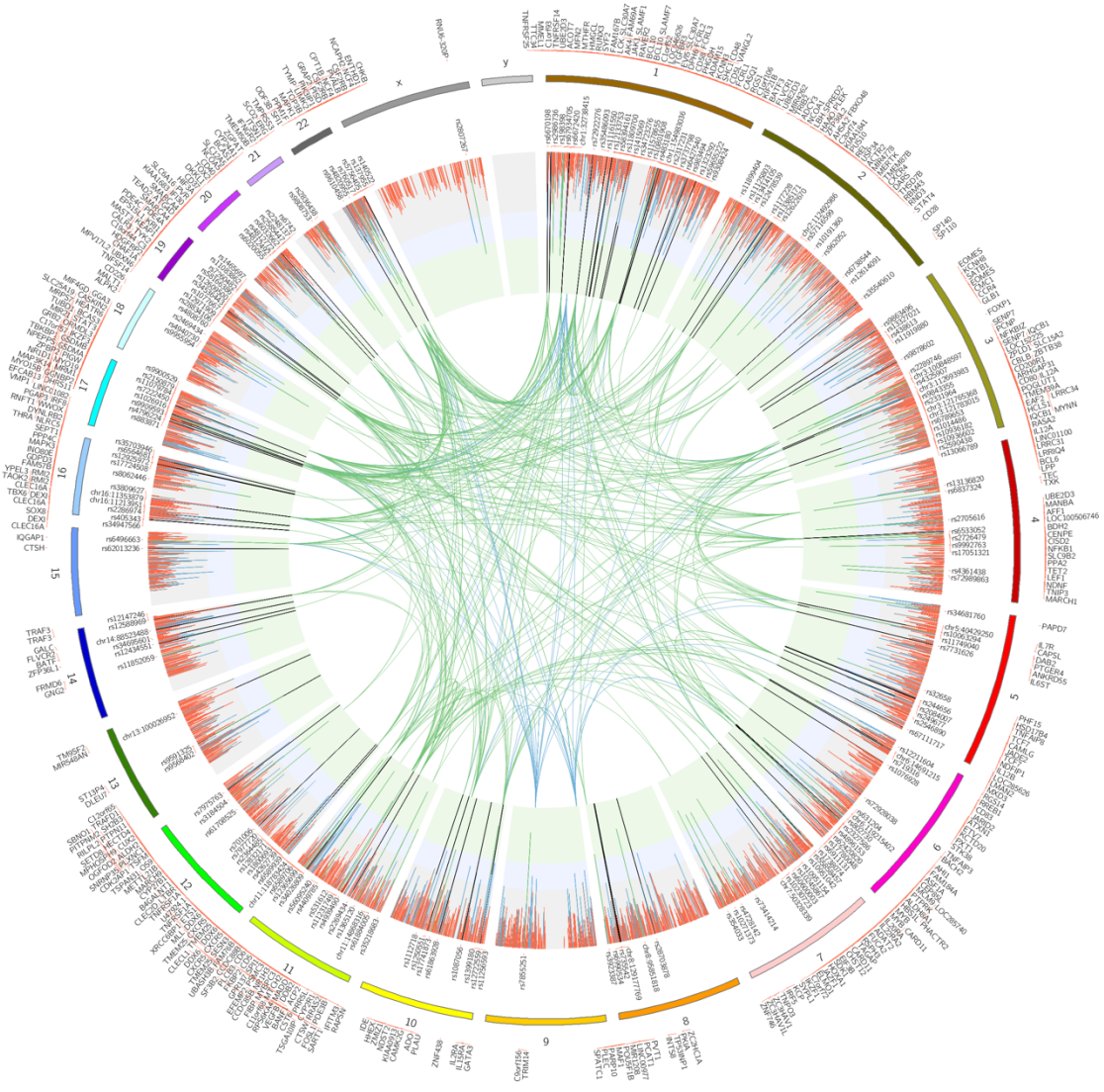


Fig. 2

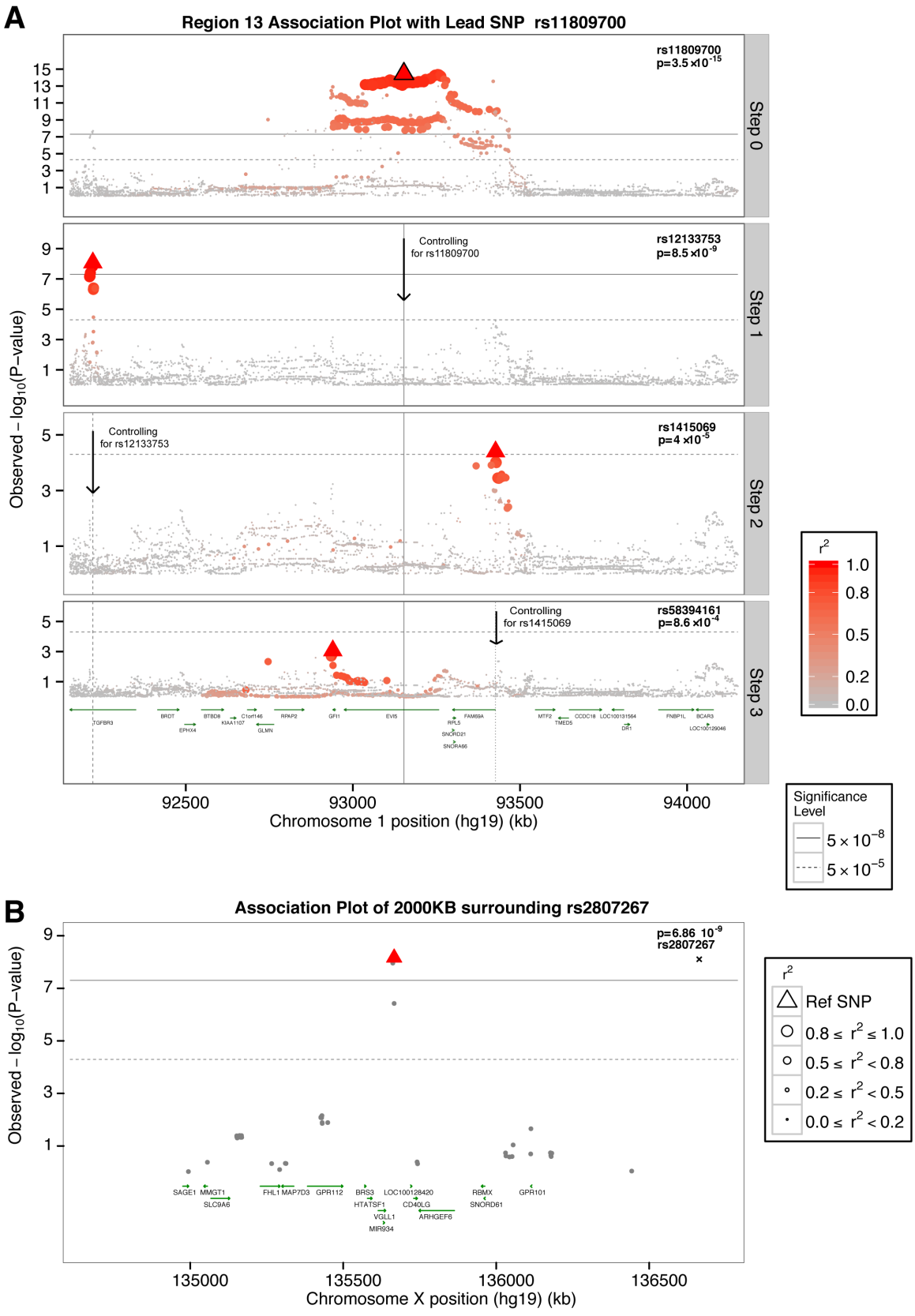

A
Region 13 Association Plot with Lead SNP rs11809700
rs11809700
p=3.5 x10-15
Step 0
Controlling
for rs11809700
rs12133753
p=8.5 x10-9
Step 1
Controlling
for rs12133753
rs1415069
p=4 x10-5
Step 2
Controlling
for rs1415069
rs58394161
p=8.6 x10-4
Step 3
Observed – log10(P–value)
92500
93000
93500
94000
Chromosome 1 position (hg19) (kb)
r2
1.0
0.8
0.5
0.2
0.0
Significance
Level
5 x 10-8
5 x 10-5
B
Association Plot of 2000KB surrounding rs2807267
p=6.86 10-9
rs2807267
Observed – log10(P–value)
135000
135500
136000
136500
Chromosome X position (hg19) (kb)
r2
Ref SNP
0.8 ≤ r2 ≤ 1.0
0.5 ≤ r2 < 0.8
0.2 ≤ r2 < 0.5
0.0 ≤ r2 < 0.2

Fig. 3

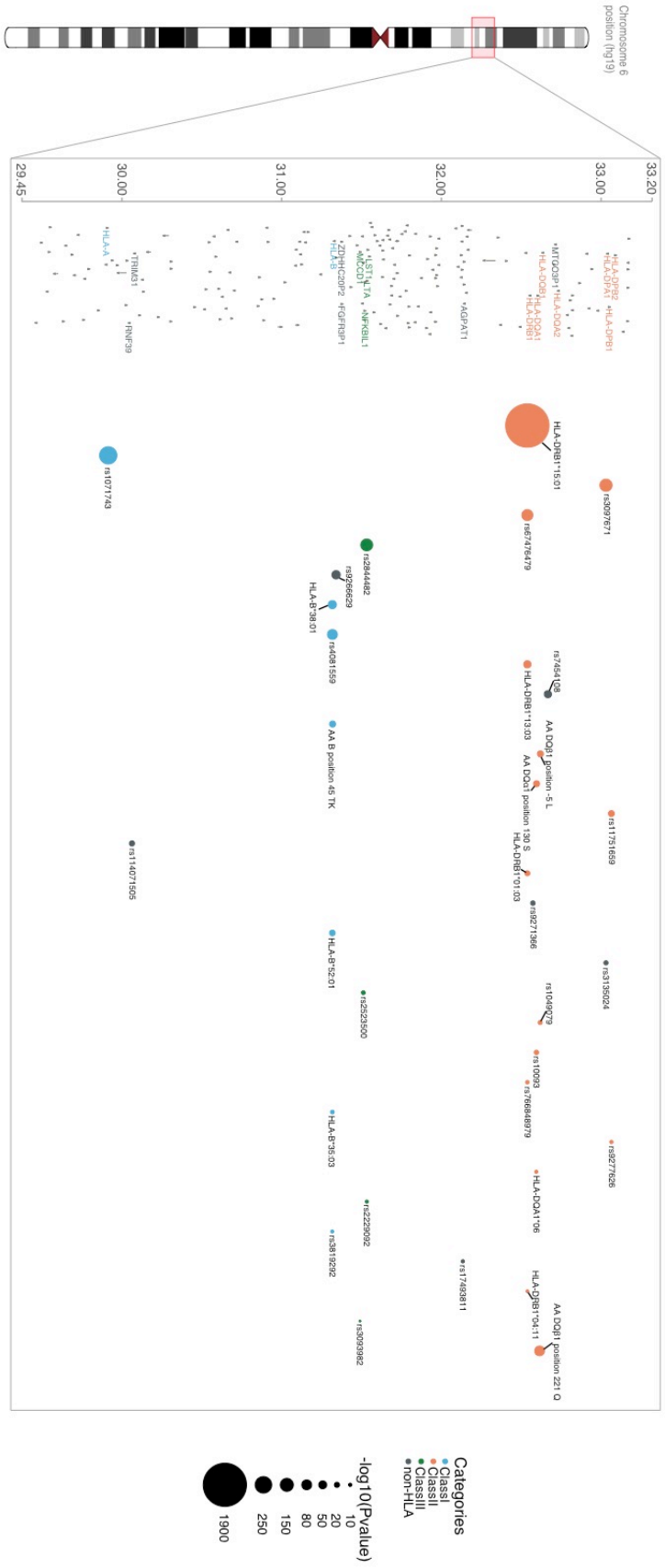

Chromosome 6
position (hg19)
33.20
33.00
32.00
31.00
30.00
29.45
HLA-DPB2
HLA-DPA1
HLA-DPB1
MTG03P
HLA-DQA2
HLA-DQA1
HLA-DRB1
AGPAT1
LST1
NCR3
NFKBIL1
ZDHHC20P2
FGFR3P1
HLA-B
TRIM31
HLA-A
RNF39
HLA-DRB1*15:01
rs3997671
rs6474479
rs2844482
rs9266629
HLA-B*38:01
rs4081559
rs7454128
HLA-DRB1*13:03
AA DQB1 position -5 L
AA DQa1 position 130 S
HLA-DRB1*01:03
rs1751659
rs9271366
rs3135024
rs1049279
rs10093
rs7766848979
rs9277626
HLA-DQA1*06
rs7749311
HLA-DRB1*04:11
AA DQB1 position 221 Q
rs1071743
rs1140715O5
AA B position 45 TK
HLA-B*52:01
rs2523500
HLA-B*35:03
rs2229092
rs3819292
rs3093982
Categories
ClassI
ClassII
ClassIII
non-HLA
-log10(Pvalue)
10
20
50
80
150
250
1900

Fig. 4

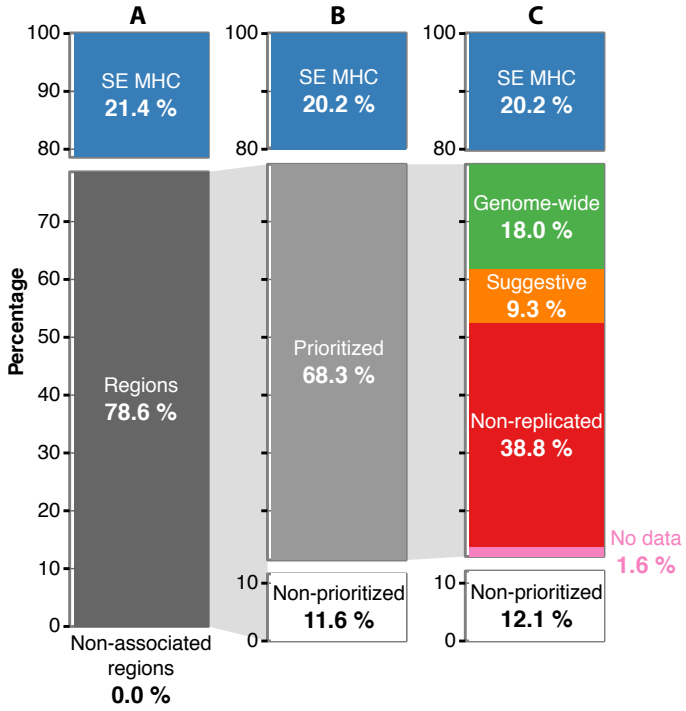

A
B
C
100
90
80
70
60
50
40
30
20
10
0
Percentage
SE MHC
21.4 %
Regions
78.6 %
Non-associated
regions
0.0 %
SE MHC
20.2 %
Prioritized
68.3 %
Non-prioritized
11.6 %
SE MHC
20.2 %
Genome-wide
18.0 %
Suggestive
9.3 %
Non-replicated
38.8 %
No data
1.6 %
Non-prioritized
12.1 %

Fig. 5

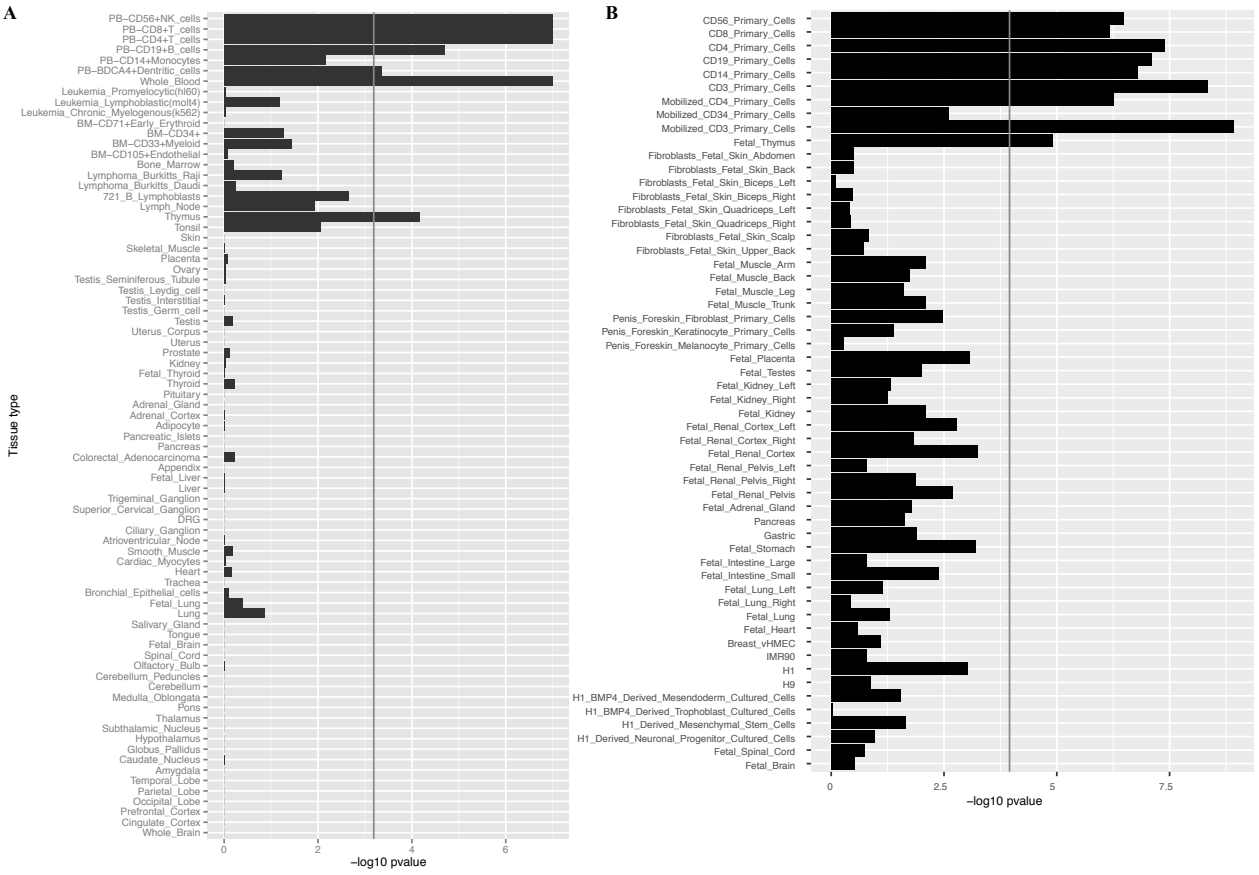
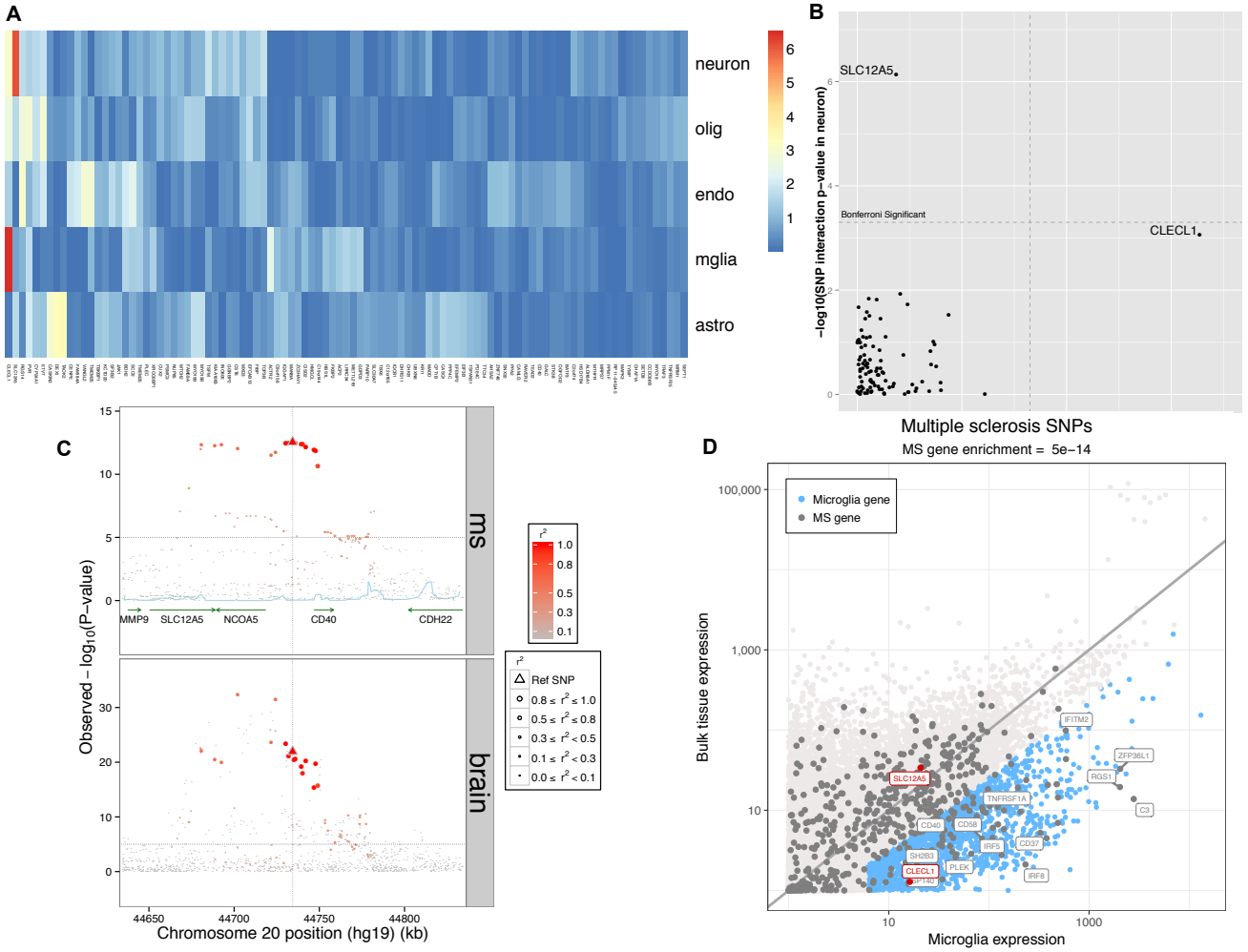

A
Tissue type
PB–CD56+NK_cells
PB–CD8+T_cells
PB–CD4+T_cells
PB–CD19+B_cells
PB–CD14+Monocytes
PB–BDCA4+Dentritic_cells
Whole_Blood
Leukemia_Promyelocytic(hl60)
Leukemia_Lymphoblastic(molt4)
Leukemia_Chronic_Myelogenous(k562)
BM–CD71+Early_Erythroid
BM–CD34+
BM–CD33+Myeloid
BM–CD105+Endothelial
Bone_Marrow
Lymphoma_Burkitts_Raji
Lymphoma_Burkitts_Daudi
721_B_Lymphoblasts
Lymph_Node
Thymus
Tonsil
Skin
Skeletal_Muscle
Placenta
Ovary
Testis_Seminiferous_Tubule
Testis_Leydig_cell
Testis_Interstitial
Testis_Germ_cell
Testis
Uterus_Corpus
Uterus
Prostate
Kidney
Fetal_Thyroid
Thyroid
Pituitary
Adrenal_Gland
Adrenal_Cortex
Adipocyte
Pancreatic_Islets
Pancreas
Colorectal_Adenocarcinoma
Appendix
Fetal_Liver
Liver
Trigeminal_Ganglion
Superior_Cervical_Ganglion
DRG
Ciliary_Ganglion
Atrioventricular_Node
Smooth_Muscle
Cardiac_Myocytes
Heart
Trachea
Bronchial_Epithelial_cells
Fetal_Lung
Lung
Salivary_Gland
Tongue
Fetal_Brain
Spinal_Cord
Olfactory_Bulb
Cerebellum_Peduncles
Cerebellum
Medulla_Oblongata
Pons
Thalamus
Subthalamic_Nucleus
Hypothalamus
Globus_Pallidus
Caudate_Nucleus
Amygdala
Temporal_Lobe
Parietal_Lobe
Occipital_Lobe
Prefrontal_Cortex
Cingulate_Cortex
Whole_Brain
0
2
4
6
–log10 pvalue
B
CD56_Primary_Cells
CD8_Primary_Cells
CD4_Primary_Cells
CD19_Primary_Cells
CD14_Primary_Cells
CD3_Primary_Cells
Mobilized_CD4_Primary_Cells
Mobilized_CD34_Primary_Cells
Mobilized_CD3_Primary_Cells
Fetal_Thymus
Fibroblasts_Fetal_Skin_Abdomen
Fibroblasts_Fetal_Skin_Back
Fibroblasts_Fetal_Skin_Biceps_Left
Fibroblasts_Fetal_Skin_Biceps_Right
Fibroblasts_Fetal_Skin_Quadriceps_Left
Fibroblasts_Fetal_Skin_Quadriceps_Right
Fibroblasts_Fetal_Skin_Scalp
Fibroblasts_Fetal_Skin_Upper_Back
Fetal_Muscle_Arm
Fetal_Muscle_Back
Fetal_Muscle_Leg
Fetal_Muscle_Trunk
Penis_Foreskin_Fibroblast_Primary_Cells
Penis_Foreskin_Keratinocyte_Primary_Cells
Penis_Foreskin_Melanocyte_Primary_Cells
Fetal_Placenta
Fetal_Testes
Fetal_Kidney_Left
Fetal_Kidney_Right
Fetal_Kidney
Fetal_Renal_Cortex_Left
Fetal_Renal_Cortex_Right
Fetal_Renal_Cortex
Fetal_Renal_Pelvis_Left
Fetal_Renal_Pelvis_Right
Fetal_Renal_Pelvis
Fetal_Adrenal_Gland
Pancreas
Gastric
Fetal_Stomach
Fetal_Intestine_Large
Fetal_Intestine_Small
Fetal_Lung_Left
Fetal_Lung_Right
Fetal_Lung
Fetal_Heart
Breast_vHMEC
IMR90
H1
H9
H1_BMP4_Derived_Mesendoderm_Cultured_Cells
H1_BMP4_Derived_Trophoblast_Cultured_Cells
H1_Derived_Mesenchymal_Stem_Cells
H1_Derived_Neuronal_Progenitor_Cultured_Cells
Fetal_Spinal_Cord
Fetal_Brain
0
2.5
5
7.5
–log10 pvalue

Fig. 6


A
neuron
olig
endo
mglia
astro
B
SLC12A5
CLECL1
Bonferroni Significant
−log10(SNP interaction p−value in neuron)
Multiple sclerosis SNPs
C
ms
brain
MMP9
SLC12A5
NCOA5
CD40
CDH22
Observed − log10(P−value)
Chromosome 20 position (hg19) (kb)
Ref SNP
D
MS gene enrichment = 5e−14
Microglia gene
MS gene
Bulk tissue expression
Microglia expression