

ASSOCIATION STUDIES ARTICLE

Impact of genetic risk loci for multiple sclerosis on expression of proximal genes in patients

Tojo James^{1,2,*,†}, Magdalena Lindén^{1,2,3,†}, Hiromasa Morikawa^{2,4,6}, Sunjay Jude Fernandes^{2,4}, Sabrina Ruhrmann^{1,2}, Mikael Huss⁵, Maya Brandi⁵, Fredrik Piehl^{1,2}, Maja Jagodic^{1,2}, Jesper Tegnér^{2,4,6,7}, Mohsen Khademi^{1,2}, Tomas Olsson^{1,2,‡}, David Gomez-Cabrero^{4,8,9,‡} and Ingrid Kockum^{1,2,‡}

¹Neuroimmunology Unit, Department of Clinical Neuroscience, ²Center for Molecular Medicine, L8: 05, ³Experimental Rheumatology Unit, Department of Medicine, Solna, ⁴Unit of Computational Medicine, Department of Medicine, Solna, Karolinska Institutet, 171 76 Stockholm, Sweden, ⁵Department of Biochemistry and Biophysics, Science for Life Laboratory, Stockholm University, Stockholm, Sweden, ⁶Biological and Environmental Sciences and Engineering Division, Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Saudi Arabia, ⁷Science for Life Laboratory, Karolinska Institutet, 171 76 Stockholm, Sweden, ⁸Mucosal and Salivary Biology Division, King's College London Dental Institute, London, UK and ⁹Translational Bioinformatics Unit, Navarrabiomed, Complejo Hospitalario de Navarra (CHN), Universidad Pública de Navarra (UPNA), IdiSNA, Pamplona, Spain

*To whom correspondence should be addressed. Tel: +46 851776258; Fax: +46 851775562; Email: tojo.james@ki.se

Abstract

Despite advancements in genetic studies, it is difficult to understand and characterize the functional relevance of disease-associated genetic variants, especially in the context of a complex multifactorial disease such as multiple sclerosis (MS). As a large proportion of expression quantitative trait loci (eQTLs) are context-specific, we performed RNA-Seq in peripheral blood mononuclear cells from MS patients ($n = 145$) to identify eQTLs in regions centered on 109 MS risk single nucleotide polymorphisms and 7 associated human leukocyte antigen variants. We identified 77 statistically significant eQTL associations, including pseudogenes and non-coding RNAs. Thirty-eight out of 40 testable eQTL effects were colocalized with the disease association signal. As many eQTLs are tissue specific, we aimed to detail their significance in different cell types. Approximately 70% of the eQTLs were replicated and characterized in at least one major peripheral blood mononuclear cell-derived cell type. Furthermore, 40% of eQTLs were found to be more pronounced in MS patients compared with non-inflammatory neurological diseases patients. In addition, we found two single nucleotide polymorphisms to be significantly associated with the proportions of three different cell types. Mapping to enhancer histone marks and predicted transcription factor binding sites added additional functional evidence for eight eQTL regions. As an example, we found that rs71624119, shared with three other autoimmune diseases and located in a primed enhancer (H3K4me1) with potential binding for STAT transcription factors, significantly associates with ANKRD55 expression. This study provides many novel and validated targets for future functional characterization of MS and other diseases.

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

[‡]The authors wish it to be known that, in their opinion, the last three authors should be regarded as joint Last Authors.

Received: November 1, 2017. Revised: December 22, 2017. Accepted: December 29, 2017

© The Author(s) 2018. Published by Oxford University Press. All rights reserved.

For Permissions, please email: journals.permissions@oup.com

Introduction

Multiple sclerosis (MS) is a chronic inflammatory disease of the central nervous system (CNS), which leads to demyelination and neuronal loss. Genetics play a significant role in this disease, the heritability being 0.64 (1). Large efforts have been made to identify genetic features that contribute to the pathogenesis of MS, resulting in a growing list of established disease-associated loci, including human leukocyte antigen (HLA) variants and single nucleotide polymorphisms (SNPs) across the genome (2–4). In addition, some lifestyle exposures affect the susceptibility of MS such as smoking, body mass index, sun-exposure and viral infections (5). Several genetic regions previously known to be associated with MS and other autoimmune or chronic inflammatory diseases, have been fine-mapped with a dense set of SNPs and tested for association with MS within the ImmunoChip Project (3). More than 100 non-HLA SNPs are currently established as MS susceptibility SNPs. Many of these may be causal variants, because fine-mapping compared to conventional genome-wide association studies (GWAS) increases the probability of pinpointing disease-causing variants rather than variants in linkage disequilibrium (LD). Interestingly, there is an overlap of MS-associated variants with other autoimmune diseases ranging from 0.9% for psoriasis to 9.1% for inflammatory bowel disease, Crohn's disease and primary biliary cirrhosis (3).

Given the non-coding nature of the majority of non-HLA MS SNPs (6), the widespread presence of expression quantitative trait loci (eQTLs) in the genome (7), and the potential impact of gene expression on biological pathways involved in disease, it is likely that a number of MS-associated loci would be involved in regulation of gene expression, acting as eQTLs. A large proportion of the eQTLs present in the human genome seem to be context-specific, i.e. they are active in specific cell types and/or as a response to molecular signaling induced by intrinsic or extrinsic stimuli (6,8,9). Most eQTL studies performed to date, including those focusing on MS variants (10), have used cells from healthy individuals, while studies in the disease context, using patient samples, are still limited. Importantly, in MS chronic inflammation is found in the CNS, with increased numbers of inflammatory cytokines and immune cells in the cerebrospinal fluid, but the inflammatory state is also reflected in the peripheral blood (11) which is more easily accessible.

Therefore, in order to study MS susceptibility variants as eQTLs in their disease and cell-type specific context, we performed whole transcriptome RNA sequencing (RNA-Seq) in peripheral blood mononuclear cells (PBMCs) from 145 patients with MS or clinically isolated syndrome (CIS) that had previously been genotyped on the ImmunoChip SNP microarray (3). Additionally, we included 36 non-inflammatory neurological disease (NINDs) controls, i.e. patients with other neurological diseases in the analysis in order to compare and to identify potential MS-specific effects. Following the eQTL analysis, we characterized the significance of the identified MS eQTLs in lymphoblastic cell lines (LCLs) and primary immune cells (CD4⁺, CD8⁺, B cells and monocytes) sorted from PBMCs obtained from different cohorts of healthy individuals or MS patients. And finally, we added functional evidence for most relevant eQTLs by mapping to enhancer histone marks and predicted transcription factor binding sites. The schematic overview of the approach followed in this study and key results are shown in Figure 1.

This is, to our knowledge, the first study where the established MS susceptibility variants have been tested for cis-eQTL

effects in the disease-specific context by using PBMCs from patients.

Results

Non-HLA and HLA eQTL effects in MS patients

We sequenced the transcriptome of PBMCs from 145 patients with MS or CIS (Cohort 1, [Supplementary Material](#), Table S1), previously genotyped on the ImmunoChip (3), and quantified gene expression using RNA-Seq. Then, we identified pairs of MS SNP–genes by selecting expressed genes within a window of 400 kb on each side of 109 MS-associated non-HLA SNPs ([Supplementary Material](#), Table S2) and 7 MS-associated HLA variants (presence of HLA-DRB1*03: 01, HLA-DRB1*15: 01 and HLA-DRB1*14: 01/14: 02 and absence of HLA-A*02: 01, HLA-B*44 and/or 45, HLA-DRB1*07 and HLA-DRB1*01). The window size was selected based on the increased probability of finding cis-regulated genes at closer distances (12,13) and to limit the number of comparisons ([Supplementary Material](#), Tables S3 and S4). For each MS SNP–gene pair identified, we conducted an eQTL analysis. We first regressed out the effects of the relevant confounders [identified in a principal component analysis (PCs), [Supplementary Material](#), Table S5] of gene expression and next, we computed the non-parametric correlation between residuals and the genotype using an additive model. The total number of MS SNP–gene pairs tested was 2223.

We defined eQTL significance using a permutation-based False Discovery Rate (FDR) threshold of 0.01. [Table 1](#) (column A) lists 49 significant non-HLA eQTLs identified in Cohort 1. [Table 2](#) (column A) lists the 28 significant HLA-associated eQTLs from 7 HLA MS risk alleles ([Supplementary Material](#), Table S4 shows the full HLA eQTL analysis results for Cohort 1).

Filtering eQTL candidate SNPs based on cell proportions

Because the genetic variants themselves may regulate the composition of immune cells within PBMCs (14), the differences in cell-type proportions may explain an observed eQTL effect. To account for that possibility, we first estimated the proportion of different cell types in each sample from the RNA-Seq data using the described method (15). Next, we tested the association of SNPs involved in significant eQTLs ([Tables 1 and 2](#)) with the cell proportions (monocytes, NK cells, CD4⁺ and CD8⁺ T cells, and B cells) using Spearman correlation. We identified significant associations to cell proportions (P-value < 0.05) for seven non-HLA SNPs and one HLA variant ([Supplementary Material](#), Table S6). Subsequently, we re-ran the eQTL analysis for these SNPs, while regressing out the cell proportion effects. As a result, two eQTLs were significant associated with cell proportions ([Supplementary Material](#), Table S7); rs2726518 (associated with *TET2*) was associated with the proportions of monocytes, and rs6881706 (associated with *SPEF2*) was correlated with the proportions of B cells, CD8⁺ cells and monocytes.

Thus, 47 non-HLA MS SNP–gene pairs and 28 HLA MS SNP–genes pairs remained significant eQTLs after correcting for cell proportions. Out of these, some eQTL-associated genes, e.g. *CPT1B*, *MANBA*, *PLEK*, *METTTL21B*, *AHI1*, *TNFRSF14*, *MERTK*, *IQCB1* and *CLECL1*, have been reported previously in studies from healthy individuals (8,10,16). Importantly, we also identified eQTLs associated with, to our knowledge, novel genes (mostly non-coding RNAs, antisense transcripts and pseudogenes). Thirty-one out of 47 non-HLA MS-eQTLs (66%) are located within 100 kb of the gene transcription start site from the SNP,

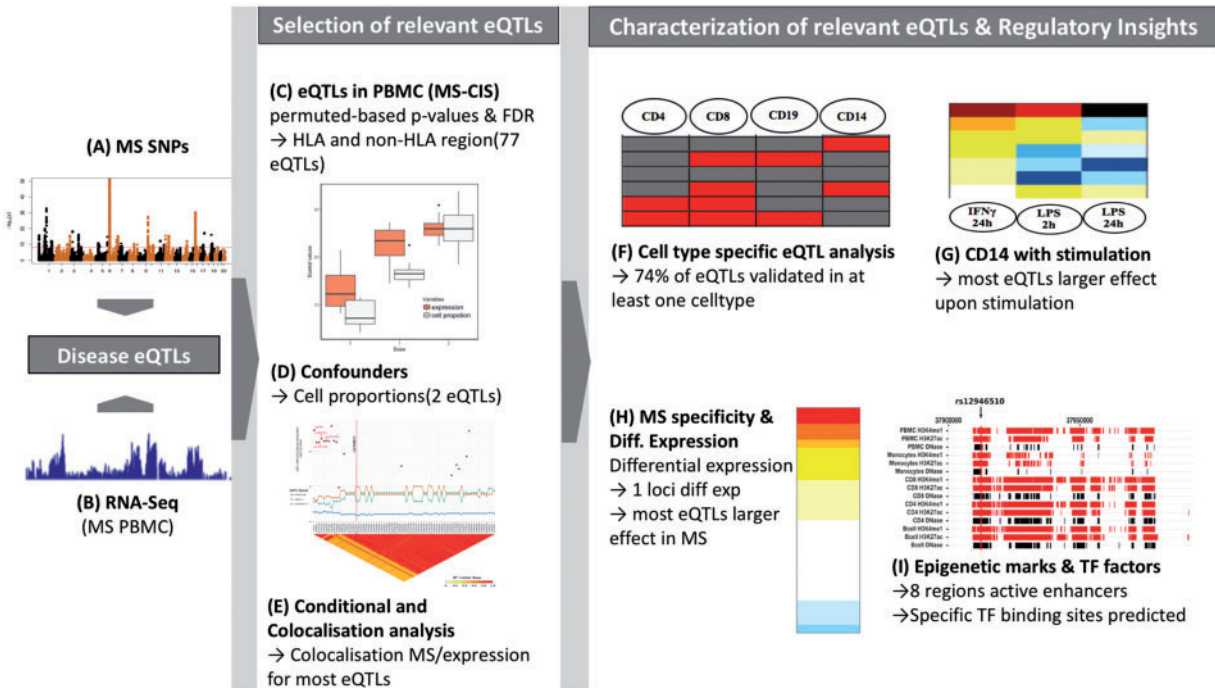


Figure 1. Schematic overview of the study and key results. (A) ImmunoChip study resulted in 110 non-HLA and 7 HLA markers associated to MS (Supplementary Material, Table S2). (B) We determined expression levels in PBMC from 181 patients with RNAseq (Supplementary Material, Table S1) (C) MS-CIS patients eQTL analysis were performed for 109 non-HLA and 5 HLA markers and resulted in 77 significant eQTLs (29 eQTLs associated to HLA markers) (Tables 1 and 2 and Supplementary Material, Tables S3 and S4). (D) Two eQTLs associated to cell proportions were identified: rs2726518 (associated with *TET2*) was associated with the proportions of monocytes, and rs6881706 (associated with *SPEF2*) was correlated with the proportions of B cells, CD8⁺ cells and monocytes (Supplementary Material, Table S6). (E) Colocalization tests performed in 22 MS susceptibility loci. In total, 38 SNP-gene eQTLs were colocalized with the MS locus signal and two SNP-gene eQTLs—rs201202118-XRCC6BP1 and rs533646-AP002954.4.1, were not colocalized (Fig. 2 and Supplementary Material, Figs S2–S5 and Tables S9 and S10). (F) Validation of MS eQTLs in cell-type specific datasets—CD4, CD8 cells (RNA-Seq) and B cells and monocytes (microarray). Of the non-HLA SNP-gene eQTLs found in PBMCs, 40% were significant in at least three additional cell types and 74% were significant in at least one cell type (Figs 3 and 4 and Supplementary Material, Tables S11 and S12). (G) MS eQTL effect changes in monocytes obtained from healthy individuals activated by different stimuli (LPS, 2 and 24 h; IFN- γ , 24 h) compared to unstimulated. eQTL pairs such as rs8042861-LQGAP1 and rs941816-ETV7 showed increased effects in monocytes stimulated with IFN- γ and LPS (2 h), rs2288904-SLC44A2 displayed an increased effect for IFN- γ , and rs2523822 (HLA-A*02)-HLA-H for each of the three stimuli, while rs11052877-CLECL1 showed a decreased eQTL effects after the three stimulations (Fig. 5 and Supplementary Material, Table S13). (H) Disease specificity of eQTL effects—32 out of 73 of the eQTL effects had a slope change <5%, with HLA-B*44 or 45-HCP5 showing the largest (41%) eQTL effect increase. FCRL3 (eQTL gene), FCRL2 and FCRL5 genes in the MS defined locus of rs706015 were significantly downregulated in MS cohort compared to NINDs (Supplementary Material, Fig. S1 and Table S19). (I) Epigenomic mapping of MS-associated variants—Out of the 26 non-HLA eQTLs of MS patients eight overlapped either with an H3K27ac or H3K4me1 active enhancer histone marks or with a DNaseI peak in monocytes, B, CD4⁺ T and CD8⁺ T cells (Fig. 2 and Supplementary Material, Figs S2–S5 and Table S20).

although 70.2% of the total eQTLs did not affect the gene closest to the SNP (Supplementary Material, Table S8).

MS eQTLs and linkage disequilibrium

Observed eQTL effects could potentially be caused by SNPs in linkage disequilibrium (LD) with the established MS-associated SNPs. Using fine-mapped genotypic information from the ImmunoChip study, we investigated all SNPs with $r^2 > 0.5$ (referred to as LD SNPs) in relation to the established top-associated MS SNPs showing eQTL effects in our data. Many of the LD SNPs showed similar or more significant eQTL effects on the same genes as the established MS SNPs (Supplementary Material, Table S9). About 49 out of 71 SNP-gene (including HLA) loci were found to have at least one SNP with better eQTL P-value compared to that of MS SNP-gene. To identify an appropriate candidate eQTL SNP in each locus, we first performed a conditional analysis where, for each MS SNP and for each associated LD SNP, the effect of each associated LD SNP was first regressed out and then residuals were correlated with the MS SNP. None of the MS SNP eQTL effects remained significant

when conditioned on any of the SNPs with $r^2 > 0.5$. Subsequently, we conducted the reverse analysis for each eQTL. For each MS SNP and for each associated LD SNP, we first regressed out the eQTL effect of the MS SNP and then residuals were correlated with the LD SNP as described previously.

Five regions contained LD SNPs with significantly associated ($P < 0.05$) eQTL effects (Fig. 2 and Supplementary Material, Figs S2–S5). These LD SNPs were associated with the expression of *ORMDL3/GSDMB*, *TNFRSF14*, *METTL21B/XRCC6BP1*, *FCRL3* and *HOTAIRM1* suggesting additional independent eQTL effects. With the exception of *FCRL3*, all these genes were located in regions that had been fine-mapped on the ImmunoChip.

For the remaining 21 LD regions included in the conditional analysis, none of the LD SNP eQTL effects remained significant when conditioned on the MS SNP (Supplementary Material, Table S9). We conclude that with few exceptions our results are consistent with ImmunoChip top-associated MS SNPs affecting MS risk by controlling expression levels. However, with an expression cohort with moderate sample size, we cannot exclude the possibility that the eQTL effect we observed for the MS-associated SNPs is due to SNPs in LD.

Table 2. List of MS-associated HLA variants showing significant eQTL effects

| HLA allele | Gene | A | | | B | | |
|------------------------------|-----------------------|----------|---------------------------------------|------------------------|---------------|---------------------------------------|------------------------|
| | | Cohort 1 | | | Pooled cohort | | |
| | | Rho | Bootstrap Rho mean (5%, 95% range) | FDR | Rho | Bootstrap Rho mean (5%, 95% range) | FDR |
| HLA-A*02:01 (absence) | IFITM4P | -0.50 | -0.5 (-0.68, -0.29) | <1.87×10 ⁻⁷ | -0.46 | -0.45 (-0.65, -0.23) | <1.87×10 ⁻⁷ |
| | HLA-F-AS1 | -0.33 | -0.33 (-0.55, -0.09) | 4.1×10 ⁻⁶ | -0.32 | -0.32 (-0.54, -0.07) | 8.1×10 ⁻⁶ |
| | HLA-H | -0.27 | -0.27 (-0.48, -0.03) | 1.8×10 ⁻⁴ | -0.24 | -0.24 (-0.47, 0.01) | 9.0×10 ⁻⁴ |
| | MICD | -0.25 | -0.25 (-0.47, -0.02) | 4.8×10 ⁻⁴ | -0.27 | -0.26 (-0.49, -0.02) | 2.3×10 ⁻⁴ |
| | HLA-K | -0.25 | -0.24 (-0.48, 0) | 7.0×10 ⁻⁴ | -0.25 | -0.25 (-0.48, 0) | 4.8×10 ⁻⁴ |
| | HCG4 ^a | -0.24 | -0.24 (-0.47, 0.01) | 1.0×10 ⁻³ | -0.25 | -0.25 (-0.48, 0) | 5.2×10 ⁻⁴ |
| | HLA-T | -0.22 | -0.22 (-0.47, 0.05) | 2.0×10 ⁻³ | -0.24 | -0.23 (-0.47, 0.03) | 1.2×10 ⁻³ |
| | ZDHHC20P1 | -0.25 | -0.25 (-0.49, 0) ^b | 5.8×10 ⁻⁴ | -0.2 | -0.2 (-0.43, 0.05) | 6.3×10 ⁻³ |
| HLA-DRB1*03:01 (presence) | HLA-DQA1 | -0.44 | -0.44 (-0.61, -0.24) | <1.87×10 ⁻⁷ | -0.47 | -0.47 (-0.64, -0.26) | <1.87×10 ⁻⁷ |
| | HLA-DQB1 | -0.41 | -0.41 (-0.58, -0.2) | <1.87×10 ⁻⁷ | -0.43 | -0.43 (-0.61, -0.21) | <1.87×10 ⁻⁷ |
| | PPP1R2P1 ^a | 0.26 | 0.25 (-0.01, 0.49) ^b | 4.0×10 ⁻⁴ | 0.19 | 0.18 (-0.1, 0.44) | 0.01 |
| | HLA-DRB5 | -0.25 | -0.25 (-0.47, -0.01) | 5.1×10 ⁻⁴ | -0.26 | -0.26 (-0.48, -0.01) | 2.6×10 ⁻⁴ |
| | HLA-DQB1-AS1 | -0.37 | -0.36 (-0.56, -0.14) | 3.7×10 ⁻⁷ | -0.4 | -0.39 (-0.58, -0.17) | 1.9×10 ⁻⁷ |
| | HLA-DQB2 | 0.36 | 0.35 (0.12, 0.55) | 9.4×10 ⁻⁷ | 0.33 | 0.33 (0.11, 0.53) | 3.7×10 ⁻⁶ |
| | PSMB9 | 0.22 | 0.21 (-0.03, 0.44) ^b | 3.1×10 ⁻³ | 0.15 | 0.15 (-0.11, 0.38) | 0.05 |
| | HLA-DRB5 | 0.78 | 0.78 (0.64, 0.87) | <1.87×10 ⁻⁷ | 0.77 | 0.77 (0.62, 0.86) | <1.87×10 ⁻⁷ |
| HLA-DRB1*15:01 (presence) | HLA-DRB1 | 0.68 | 0.68 (0.53, 0.79) | <1.87×10 ⁻⁷ | 0.66 | 0.65 (0.49, 0.78) | <1.87×10 ⁻⁷ |
| | HLA-DQB1 | 0.61 | 0.61 (0.41, 0.76) | <1.87×10 ⁻⁷ | 0.56 | 0.55 (0.34, 0.72) | <1.87×10 ⁻⁷ |
| | HLA-DQB2 | -0.51 | -0.51 (-0.69, -0.28) | <1.87×10 ⁻⁷ | -0.46 | -0.45 (-0.65, -0.23) | <1.87×10 ⁻⁷ |
| | HLA-DRB6 | -0.48 | -0.48 (-0.67, -0.25) | <1.87×10 ⁻⁷ | -0.48 | -0.47 (-0.67, -0.24) | <1.87×10 ⁻⁷ |
| | HLA-DQB1-AS1 | 0.46 | 0.46 (0.24, 0.65) | <1.87×10 ⁻⁷ | 0.42 | 0.41 (0.19, 0.61) | <1.87×10 ⁻⁷ |
| | HLA-DQA1 | 0.39 | 0.38 (0.14, 0.59) | 1.9×10 ⁻⁷ | 0.34 | 0.34 (0.1, 0.56) | 2.1×10 ⁻⁶ |
| | HLA-DQA2 | -0.36 | -0.35 (-0.56, -0.11) | 9.4×10 ⁻⁷ | -0.36 | -0.36 (-0.57, -0.13) | 5.6×10 ⁻⁷ |
| | TAP2 | 0.29 | 0.29 (0.05, 0.52) ^b | 5.0×10 ⁻⁵ | 0.21 | 0.2 (-0.04, 0.43) | 5.1×10 ⁻³ |
| | BTNL2 ^a | 0.21 | 0.21 (-0.05, 0.46) | 3.5×10 ⁻³ | 0.19 | 0.19 (-0.08, 0.44) | 8.9×10 ⁻³ |
| | NOTCH4 | 0.22 | 0.22 (-0.04, 0.45) | 2.7×10 ⁻³ | 0.2 | 0.19 (-0.07, 0.44) | 7.8×10 ⁻³ |
| HLA-B*44 and or 45 (absence) | HLA-B | 0.23 | 0.23 (-0.02, 0.46) | 1.34×10 ⁻³ | 0.26 | 0.26 (0.01, 0.47) | 2.45×10 ⁻⁴ |
| | HCP5 ^a | 0.22 | 0.22 (-0.04, 0.44) ^b | 2.44×10 ⁻³ | 0.13 | 0.13 (-0.12, 0.38) | 0.07 |

Cohort 1, MS specific cohort; Pooled Cohort, Both MS specific cohort and non-inflammatory neurological disease (NINDs) cohort.

^aNovel eQTL genes found in PBMC.

^beQTL genes which are enriched in MS specific cohort.

MS eQTLs and colocalization with MS risk

For the refinement of causal variants based on disease-associated genetic variants and gene expression, we performed colocalization tests in 22 MS susceptibility loci including HLA loci (the number was obtained after applying the criteria described in Materials and Methods), corresponding to 40 eQTL-gene pairs. The colocalization method 'COLOC' provides an approximate Bayes factor to estimate a posterior probabilities that a SNP is causal in both immunochip disease association and eQTL analysis. Thirty-eight SNP-gene eQTLs were colocalized with the MS locus signal and two SNP-gene eQTLs—rs201202118-XRCC6BP1 (Fig. 2) and rs533646-AP002954.4.1, were not colocalized with disease association signal from ImmunoChip study (Supplementary Material, Table S10).

Validation and exploration of MS eQTLs in cell-type specific datasets

Considering (a) the context-specific nature of eQTLs, and (b) that most eQTLs are expected to be significant in healthy individuals as well, we aimed to characterize significant eQTLs activity in different cell types using published and novel datasets. We characterized the identified eQTLs in lymphoblastic cell lines (LCLs), naïve B cells, naïve monocytes, as well as

monocytes stimulated with LPS (2 and 24 h) or IFN- γ (24 h) sampled from healthy volunteers for which we obtained genotypes and expression data from public datasets (7,8,16) (Supplementary Material, Table S11). We also used a new small cohort consisting of sorted CD4⁺ and CD8⁺ T cells from MS patients and healthy controls (Table 3).

We performed eQTL analyses with the same methodology described earlier. Gene expression in the LCLs and primary T cells had been analyzed using RNA-Seq, which allowed us to detect and validate several of the non-coding RNA transcripts or transcribed pseudogenes found in our initial analysis in the PBMCs. The gene expression data from sorted primary immune cells from healthy individuals had been obtained through microarray analyses, where probes were either lacking or expressed below detection limit for some of the eQTL-affected genes. For confirmation of an eQTL effect on a given gene-locus pair, we accepted a nominal P-value of <0.05 if the effect shows the same direction as observed in Cohort 1. More than 40% of the non-HLA SNP-gene eQTLs found in PBMCs was identified significant in at least three additional cell types. Furthermore, 74% were identified significant in at least one cell type (P-value 3.55×10⁻⁷). Some eQTLs were general for all cell types, while most of them were specific for one or a few cell types or conditions (Figs 3 and 4 and Supplementary Material, Tables S12–S16). Importantly, we

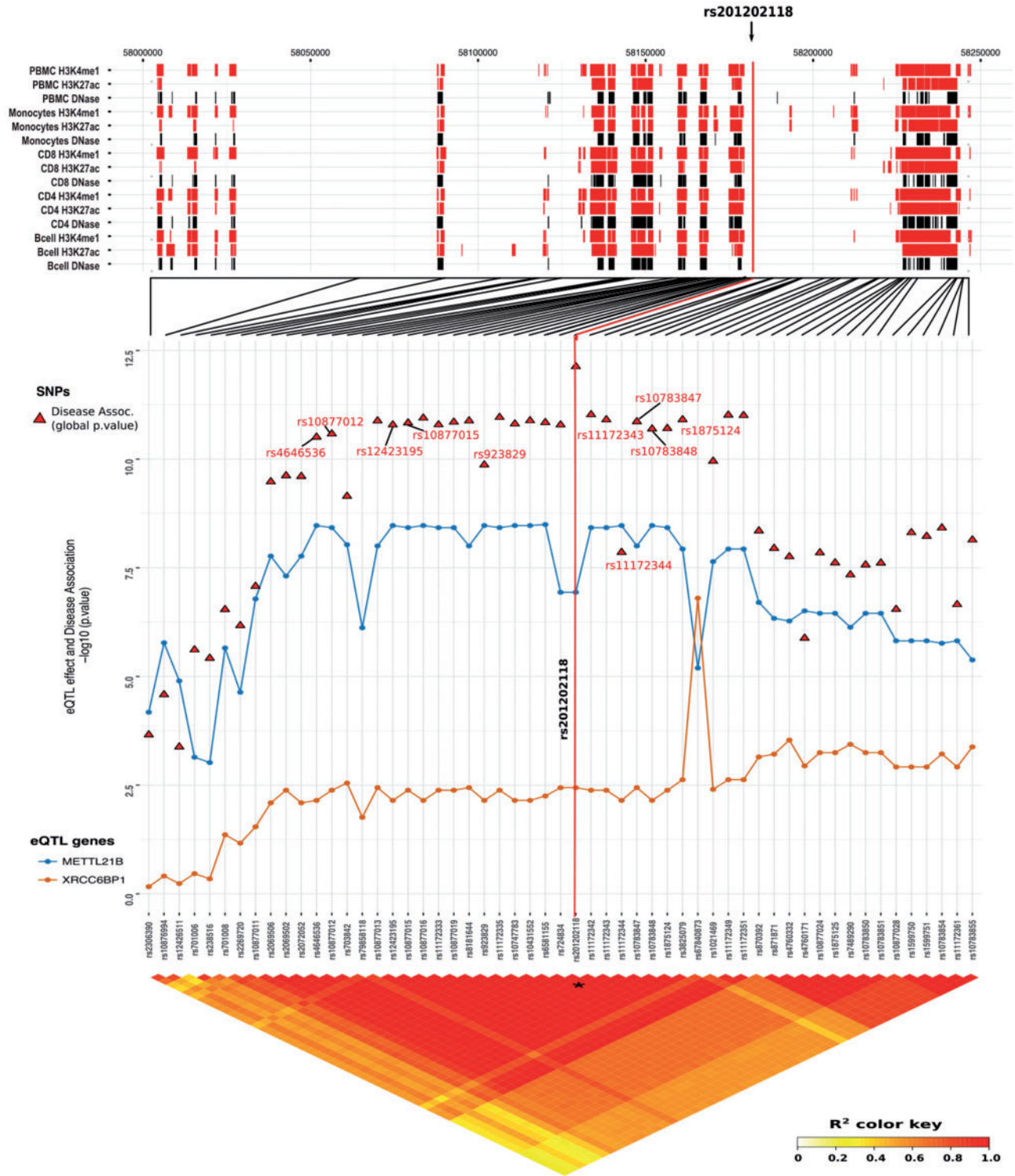


Figure 2. MS associations, LD and mapping to DNaseI peaks, active enhancer chromatin marks and TFBS of ImmunoChip SNPs in the *METTL21B/XRCC6BP1* region. The ImmunoChip association analysis results (red dots) are shown for each of the SNPs. SNPs labelled in red have transcription factor binding site based on RegulomeDB—STAT1 (rs10783848), NTAf (rs1875124), BCL6B (rs11172344), SRF (rs11172343), IRF3 (rs923829), HELIOSA (rs10877015), FOXP (rs12423195), SOX4 (rs12423195), FOXL1 (rs10877012) and E2F2 (rs4646536). The top panel shows regions with H3K4me1 and H3K27ac chromatin mark peaks (red) and DNaseI peaks (black) in the different cell types analyzed in this study (PBMCs, Monocytes, CD8+ and CD4+ T cells and B cells). SNP–gene correlation P-value for two genes—*METTL21B* and *XRCC6BP1* are marked as solid lines. *XRCC6BP1* eQTL association signal is not colocalized with the SNP disease association signals.

Table 3. Summary of different datasets used for genotyping and screening phase (PBMC) and validation phase (PBMC-derived immune cell types)

| Cohort | Tissue | MS | NINDs | HC | Expression type (source) | Analysis results and data |
|--|--------------------------------|-------|-----------------|-------|--|--|
| Pooled Cohort [Cohort 1 (MS)+Cohort 2 (NINDs)] | PBMC | 145 | 36 ^a | 0 | RNA-Seq | Tables 1 and 2 and Supplementary Material, Tables S1, S3, S4 and S15–S18 |
| Cell-type specific | CD4 ^a | 19 | 0 | 8 | RNA-Seq | Figures 3 and 4 and Supplementary Material, Tables S12, S13 and S23 |
| Cell-type specific | CD8 ^a | 17 | 0 | 7 | RNA-Seq | Figures 3 and 4 and Supplementary Material, Tables S12, S13 and S23 |
| Cell-type specific | B-cells | 0 | 0 | 233 | Microarray (16) | Figures 3 and 4 and Supplementary Material, Tables S12 and 13 |
| Cell-type specific | Monocytes | 0 | 0 | 414 | Microarray (8) | Figures 3–5 and Supplementary Material, Tables S12–16 |
| Cell-type specific (stimulation) | Monocytes (IFN gamma) | 0 | 0 | 367 | Microarray (8) | Figures 3–5 and Supplementary Material, Tables S11–S13 |
| Cell-type specific (stimulation-time point) | Monocytes (LPS—2 h) | 0 | 0 | 261 | Microarray (8) | Figures 3–5 and Supplementary Material, Tables S11–S13 |
| Cell-type specific (stimulation-time point) | Monocytes (LPS—24 h) | 0 | 0 | 322 | Microarray (8) | Figures 3–5 and Supplementary Material, Tables S11–S13 |
| Cell-type specific (1000 genome project) | EBV-transformed LCLs (B cells) | 0 | 0 | 234 | RNA-Seq (7) | Figures 3 and 4 and Supplementary Material, Tables S11 and S12 |
| Additional PBMC dataset (MS and NINDs) | PBMC | 23 | 7 | 0 | Human Genome U133 plus 2.0 arrays (18) | Supplementary Material, Table S19 |
| Global Cohort Association study (Cases and Controls) | Blood (DNA) | 17465 | | 30385 | ImmunoChip custom genotyping array (3) | Supplementary Material, Table S10 |
| Joint Analysis cohort (Cases and Controls) | Blood (DNA) | 29300 | | 50794 | ImmunoChip custom genotyping array (3) | Supplementary Material, Table S10 |

LCLs, lymphoblastoid cell lines.

^aSample size estimation and power analysis done for these cohorts.

acknowledge that the analysis using the T cell cohort had low statistical power due to its size; however, the number of replicated eQTLs (P -value < 0.05) is higher or equal to the expected numbers from a power analysis.

In the analysis of eQTL effects in the HLA region, we found that the *HLA-DRB1*15:01* allele or its tag SNP, rs3135388, significantly influenced expression of the *DRB1*, *DRB5* and *DQB1* genes throughout the tested datasets except T cells (Fig. 4 and Supplementary Material, Table S12).

Influence of LPS and IFN- γ stimulation on MS eQTLs in monocytes

To further characterize the context-specific nature of the eQTLs in monocytes, which are thought to be involved in MS pathogenesis (17), we analyzed the behavior of the MS eQTLs in monocytes obtained from healthy individuals (8) activated by different stimuli (LPS, 2 and 24 h; IFN- γ , 24 h) compared to unstimulated, by comparing the changes in effects, measured as rho-values (Fig. 5 and Supplementary Material, Table S13). Most increases in MS eQTL effects (ρ) were observed upon IFN- γ stimulation, while a certain predominance of decreased eQTL effects was seen upon stimulations with LPS as compared with unstimulated cells. In accordance with the pattern observed in PBMCs from MS patients, eQTL pairs such as rs8042861-IQGAP1 and rs941816-ETV7 showed increased effects in monocytes stimulated with IFN- γ and LPS (2 h), rs2288904-SLC44A2 displayed an increased effect for IFN- γ , and rs2523822 (*HLA-A*02*)-*HLA-H* for each of the three stimuli, while rs11052877-CLECL1 showed a decreased eQTL effects after the three stimulations.

Disease-specific genes in MS, CIS and NINDs

To understand the role of MS-SNP regulated genes in the disease, gene expression levels of MS were compared to those of NINDs and CIS (Supplementary Material, Table S14). Thirty-nine out of 45 differentially expressed genes were downregulated in MS compared to NINDs (adjusted P -value < 0.05) and no genes were significantly differentially expressed between MS and CIS samples (Supplementary Material, Table S14). *FCRL3* (eQTL gene), *FCRL2* and *FCRL5* genes in the MS defined locus of rs706015 were significantly downregulated in MS cohort compared to NINDs (Supplementary Material, Fig. S5).

Disease specificity of eQTL effects

While we only observe one eQTL gene, *FCRL3*, to be differentially expressed, we aimed to investigate if the eQTL effects are disease specific by displaying an increased association in MS, or rather general and independent of the disease status.

To this end, we pooled RNA-Seq and genotyping data from 36 NINDs (Cohort 2) and MS (Cohort 1) and regressed out the effect of disease type (MS/NINDs) in the gene expression profiles. Then, we compared results from this Pooled cohort with those from Cohort 1 [Tables 1 and 2 (panel B)]; the complete analysis results for the Pooled Cohort are found in Supplementary Material, Tables S15 and S16]. All non-HLA and HLA eQTLs except two (*HLA-DRB1*03:01*-*PSMB9* and *HLA-B*44/45*-*HCP5*) remained significant in the Pooled Cohort. We calculated the percentage of relative increase or decrease of eQTL effects in the MS group (Cohort 1) as compared to when patients with all diagnoses are included (Pooled Cohort) (Supplementary Material, Fig. S1A and B and Tables S17 and S18). The results

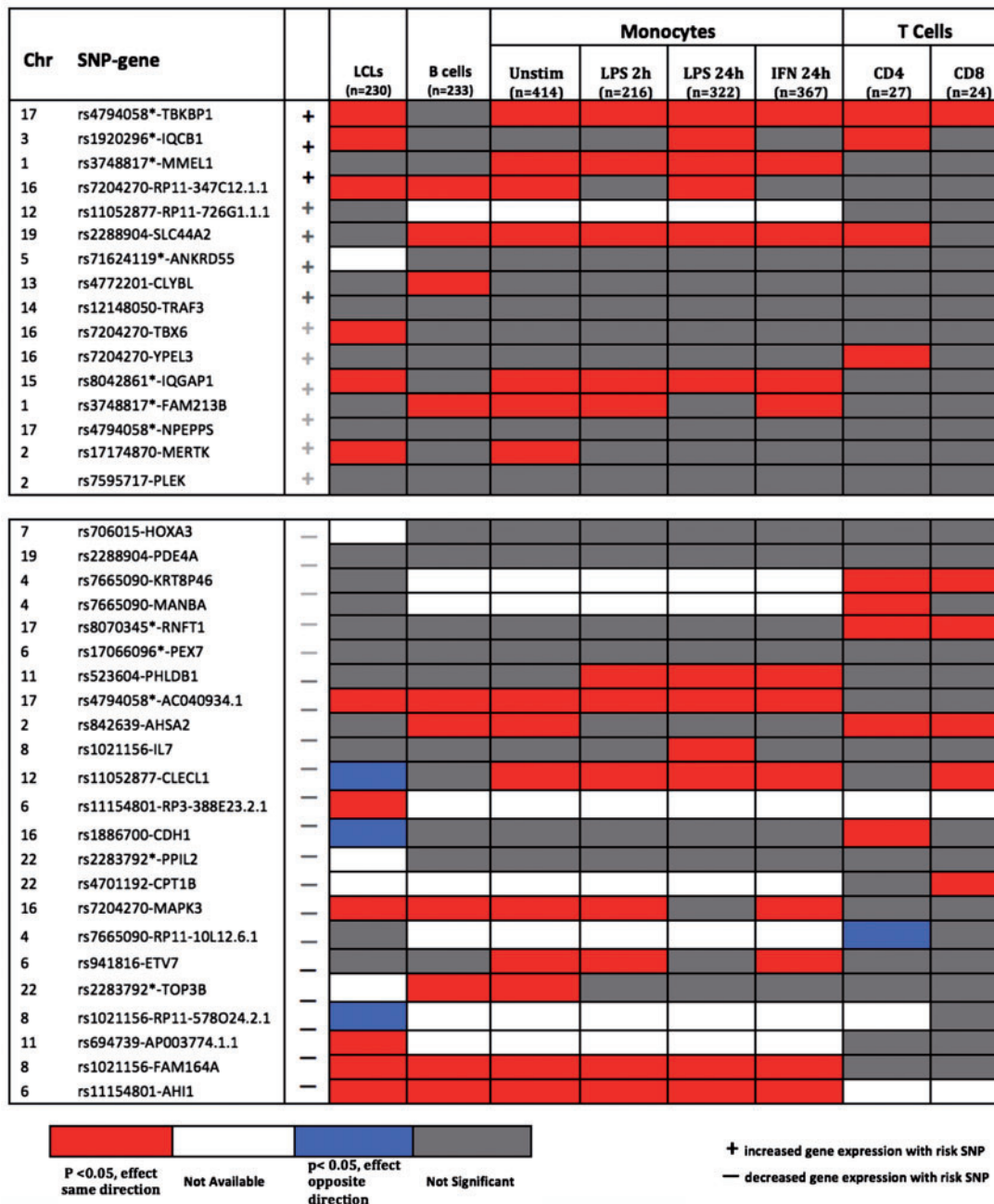


Figure 3. Graphical overview of MS-associated non-HLA eQTLs in immune cell types. Validation and exploration of MS non-HLA eQTLs in LCLs, primary B cells, primary monocytes [unstimulated, stimulated with IFN- γ (24 h), LPS (2 h), LPS (24 h)], from healthy individuals, as well as CD4⁺ and CD8⁺ T cells from MS patients and healthy controls. A nominal P-value of <0.05 was accepted as validation of a significant eQTL in PBMCs. *A proxy SNP was used in the analysis of primary B cells and monocytes (listed in [Supplementary Material](#), Table S24). eQTL SNPs (rs2050568, rs2726518 rs12946510 and rs201202118) were excluded as original snps or tag snps were not obtained for the corresponding studies. +/- indicates the gene expression increase/decrease with respect to the risk allele.

indicated that 32 out of 73 of the eQTL effects had a slope change <5%, with HLA-B*44 or 45-HCP5 ([Supplementary Material](#), Fig. S1B) showing the largest (41%) eQTL effect increase. We lack power to quantify accurately the change of the slope per eQTL, however, we observed that for a set of HLA eQTLs the association strength (absolute value of ρ) is reduced when adding non-MS individuals ([Supplementary Material](#), Table S17) despite the increase of the number of individuals (P-value <0.05). In non-HLA eQTLs, we also observed for many

eQTLs a reduction of association strength, but it was not statistically significant. However, to further test the validity of increase or decrease of eQTL effects in MS compared to other patients, we computed eQTLs in a small previously analyzed cohort with MS and NINDs (OND) patient samples taken from the same study population as the Pooled Cohort (18) and we found that the change of the effect between MS and NINDs was in the same direction for a significant set of eQTLs (P-value <0.05) ([Supplementary Material](#), Table S19).

| HLA allele | Tag SNP | gene | | LCLs (n=234) | B cells (n=233) | Monocytes | | | | T cells | |
|-------------------------------------|-----------|---------------------|---|-----------------|--------------------|-------------------|-------------------|--------------------|----------------|---------------|---------------|
| | | | | | | unstim (n=414) | LPS 2h (n=216) | LPS 24h (n=322) | IFN (n=367) | CD4 (n=27) | CD8 (n=24) |
| A*02:01 (absence) | rs2970533 | <i>IFITM4P</i> | - | | | | | | | | |
| | | <i>HLA-F-AS1</i> | - | | | | | | | | |
| | | <i>HLA-H</i> | - | | | | | | | | |
| | | <i>MICD</i> | - | | | | | | | | |
| | | <i>HLA-K</i> | - | | | | | | | | |
| | | <i>ZDHHC20P1</i> | - | | | | | | | | |
| | | <i>HCG4</i> | - | | | | | | | | |
| | | <i>HLA-T</i> | - | | | | | | | | |
| DRB1*03:01 (presence) | rs2187668 | <i>HLA-DQA1</i> | - | | | | | | | | |
| | | <i>HLA-DQB1</i> | - | | | | | | | | |
| | | <i>HLA-DQB1-AS1</i> | - | | | | | | | | |
| | | <i>HLA-DRB5</i> | - | | | | | | | | |
| | | <i>PSMB9</i> | + | | | | | | | | |
| | | <i>PPP1R2P1</i> | + | | | | | | | | |
| | | <i>HLA-DQB2</i> | + | | | | | | | | |
| DRB1*15:01 (presence) | rs9271366 | <i>HLA-DQB2</i> | - | | | | | | | | |
| | | <i>HLA-DRB6</i> | - | | | | | | | | |
| | | <i>HLA-DQA2</i> | - | | | | | | | | |
| | | <i>BTNL2</i> | + | | | | | | | | |
| | | <i>NOTCH4</i> | + | | | | | | | | |
| | | <i>TAP2</i> | + | | | | | | | | |
| | | <i>HLA-DQA1</i> | + | | | | | | | | |
| | | <i>HLA-DQB1-AS1</i> | + | | | | | | | | |
| | | <i>HLA-DQB1</i> | + | | | | | | | | |
| | | <i>HLA-DRB1</i> | + | | | | | | | | |
| | | <i>HLA-DRB5</i> | + | | | | | | | | |
| B*44 and/or 45 (absence) | rs2596551 | <i>HLA-B</i> | + | | | | | | | | |
| | | <i>HCP5</i> | + | | | | | | | | |

P < 0.05, effect same direction

Not Available

p < 0.05, effect opposite direction

Not Significant

+ increased gene expression with risk SNP
- decreased gene expression with risk SNP

Figure 4. Graphical overview of MS-associated HLA eQTLs in immune cell types. Validation and exploration of MS HLA eQTLs in LCLs, primary B cells, primary monocytes [unstimulated, stimulated with IFN- γ (24 h), LPS (2 h), LPS (24 h)], from healthy individuals, as well as CD4+ and CD8+ T cells from MS patients and healthy controls. *rsids used as tag SNPs for the HLA alleles in LCLs, B cells and monocytes. +/- indicates the gene expression increase/decrease with respect to the risk allele.

In order to address if our results were affected by realapse-remission status, we investigated the eQTLs reported in Tables 1 and 2, stratifying the MS and CIS patients based on realapse-remission status. Among patients in remission ($n=101$) three of the eQTLs (*HLA-DRB1*03:01-PSMB9*, rs11154801- *RP3-388E23.2.1* and rs8070345- *RNFT1*) were not significant while all the others remained significant. However, these three eQTLs were all significant among patients in remission ($n=29$), indicating that they may be specific to situation of active inflammatory response (data not shown).

Chromatin marks

To functionally characterize the MS eQTLs identified, we performed a bioinformatics screening of non-HLA MS eQTL loci using chromatin marks and DNaseI peaks, as well as analysis of predicted TFBS mapping from ENCODE data (19) and Roadmap

Epigenomics data (20) from healthy individuals. Out of the 26 non-HLA eQTLs in PBMCs of MS patients eight overlapped either with an H3K27ac or H3K4me1 active enhancer histone marks or with a DNaseI peak in monocytes, B, CD4+ T and CD8+ T cells. Two non-HLA eQTLs also overlapped with a predicted TFBS for PIT-1 and STAT1, while NKX2-3, NKX2-8, *Irx5*, *Irx3* and *GRHL1* were predicted to bind SNPs without any other overlapping marks (Supplementary Material, Table S20).

Discussion

We have performed an eQTL analysis of the top-associated MS risk loci (3) by studying cis-eQTL effects in 800-kb windows in PBMCs from patients with MS. We found 35 MS loci that displayed eQTL effects influencing one or more genes within the analyzed regions. A majority of the eQTL effects remained significant when adding data from non-MS patients, however,

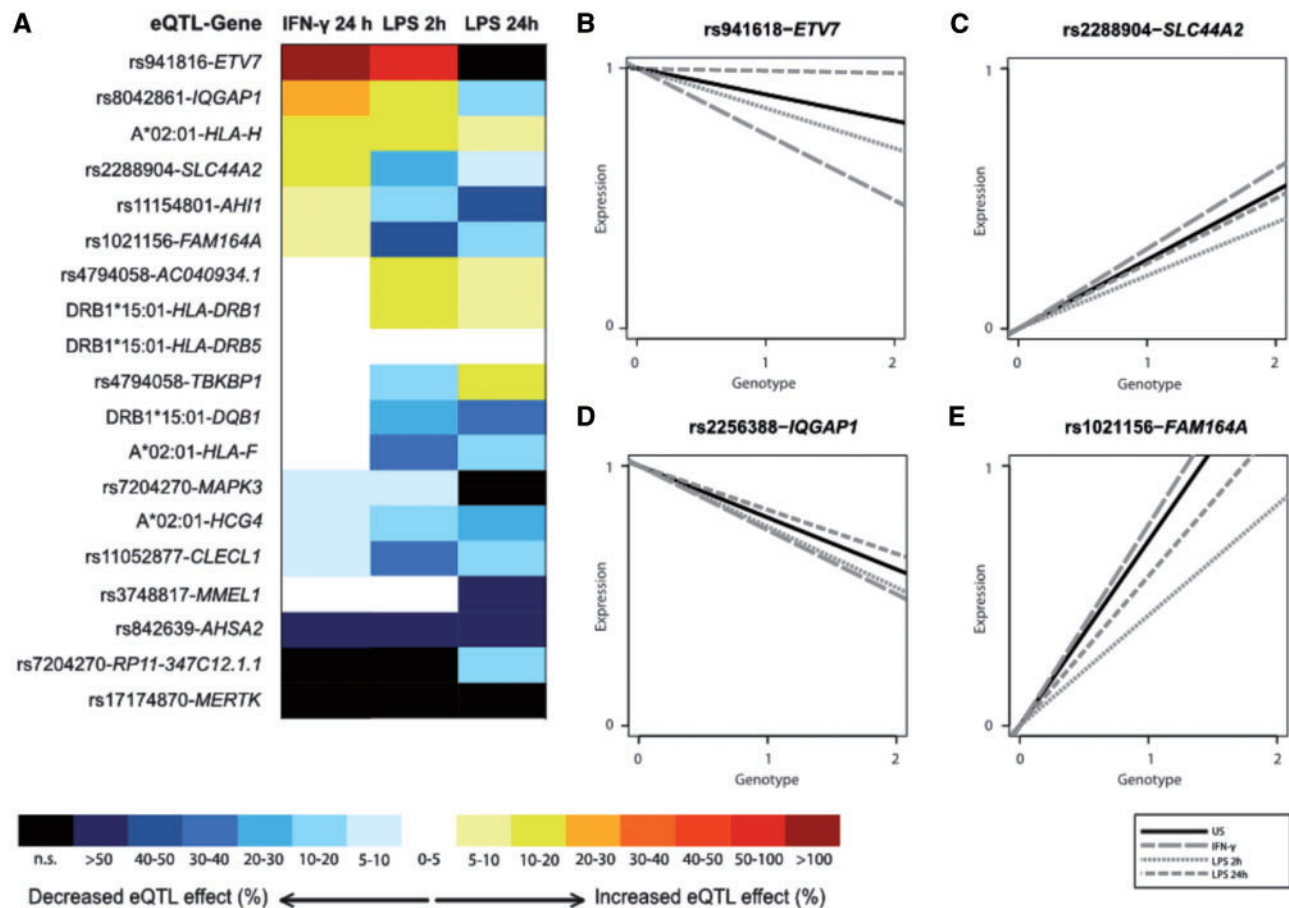


Figure 5. Heat map of changes in MS eQTL effects upon stimulation of monocytes. The changes are measured as rho values and are depicted in percentage relative to the unstimulated (naïve) condition. Only SNP-gene pairs with nominally significant ($P < 0.05$) effects in the naïve state are included in the heatmap. Graphs are illustrating eQTL effect size (beta-value) changes for the different stimuli in monocytes, showing examples of expression of four genes: *ETV7*, *SLC44A2*, *IQGAP1* and *FAM164A*.

$>70\%$ of the effect sizes changed when analyzed in only MS/CIS patients, with a predominance of increased effects in MS cohort for HLA-related eQTLs (Supplementary Material, Fig. S1).

We replicated the eQTLs identified from PBMCs in datasets from LCLs, primary B cells and monocytes from healthy individuals, including monocytes activated by LPS and IFN- γ as well as in CD4 $^+$ and CD8 $^+$ T cells from MS patients and healthy controls. In total, 74% of the non-HLA eQTLs were replicated in at least one cell type. We replicated known observations from healthy individuals (8,10,16) as well as reported previously unidentified eQTLs, especially for HLA genes, long non-coding or anti-sense RNA genes and pseudogenes. Interpretation of comparisons with previously published eQTL studies should however be made with caution, considering the differences in the SNPs or proxy SNPs, expression platforms and number of study subjects included. Importantly, considering the overlap in risk genes between autoimmune diseases, we consider that the results presented here are relevant for other autoimmune diseases (Table 1).

Roles of MS eQTLs in different immune cell types

B cells and monocytes are found in MS lesions and their roles in the disease process are being investigated (21–24), however, there are plenty of studies pointing toward an important role of T cells in the disease (2,3). Therefore, we presented RNA-Seq

based eQTL data from a small cohort of CD4 $^+$ and CD8 $^+$ T cells, which allowed us to validate a few of the eQTL findings from PBMCs. Despite the small sample size, two of the eQTL effects found in CD4 $^+$ and CD8 $^+$ T cells (rs7204270-YPEL3, rs8070345-RNFT1) could not be replicated in any other cell type, thus suggesting T cell specificity (Fig. 3 and Supplementary Material, Table S11). Larger sample sizes will be needed for a more robust exploration of MS SNPs as eQTLs in different T cell subsets.

After integrating chromatin marks and TFBS data, and taking the different validation strategies together, we found that the *AHSA2* eQTL, rs842639, had perfectly matching results for DNase I and histone marks throughout all tested datasets, confirming its eQTL effect. For the remaining MS eQTLs histone marks, DNaseI and TFBS data overlapped with eQTL results in the different tissues to a varying extent.

To describe some examples, rs3748817 was found as an eQTL for *FAM213B* and *MMEL1* in PBMCs from MS patients. However, the behavior of this eQTL effect under different stimulatory conditions of monocytes varied (no eQTL for *FAM213B* in 24-h stimulation with LPS), and it also changed in different PBMC-derived cell types (*MMEL1* has no eQTL effect in B cells, and no eQTL effect for any of the genes in T cells). An active enhancer histone mark was also found for this SNP in monocytes (Supplementary Material, Table S20 and Fig. 4). rs523604, an eQTL for *PHLDB1* in MS PBMCs, was found in healthy monocytes, however not supported by any other data for this cell

type, while it mapped to active enhancer in healthy PBMCs, B cells and CD8⁺ T cells. We confirmed the eQTL for *CLYBL*, rs4772201, in healthy B cells, while active enhancer histone marks and DNase I peaks were found throughout the analyzed cell types (Supplementary Material, Fig. S2 and Table S20). Most of the TFs with predicted binding to eQTL SNPs have described functions during development (mainly neuronal) and cell cycle. Only STAT1 (predicted to bind to rs71624119, eQTL for *ANKRD55*) and NKX2-3 (predicted to bind rs11154801, eQTL for *AHI1* and *RP3-388E23.2.1*) have known functions in immunity.

Detection of shared causal variants

A question of importance is whether the disease-associated SNP in a region is the one driving an observed eQTL effect and therefore possibly responsible for disease risk. We show that for at least 6 of the 24 testable MS SNPs there are other SNPs in the region with stronger eQTL effect in PBMCs from MS patients (Supplementary Material, Table S9). The most significant disease-associated SNPs in the global analyses in these regions may simply have other functional effects than eQTLs in PBMCs, or alternatively, previous association analysis and fine mapping has failed to pinpoint the correct SNP for MS susceptibility (Fig. 2 and Supplementary Material, Figs S2–S5). Additionally, for every gene–SNP pair, we conducted a colocalization analysis in a window span of 800 kb around risk SNP, to test whether the eQTL variants colocalizes with the disease association variant. Our study suggest that 95% of the testable disease loci were colocalized to the eQTL gene signal (Supplementary Material, Table S10). Larger cohorts that enable a more thorough dissection of the LD structure and regions with even coverage of SNPs will be needed in order to conclude exactly which SNPs are necessary for the effect, although it is likely that haplotypes, rather than single SNPs are the drivers (25).

Modification of eQTL effects by context

Current data from eQTL analyses in different immune cell types show that a large proportion of eQTL effects are highly context-specific, i.e. they vary depending on the signaling pathways activated by different kinds of stimuli, such as cytokines affecting the particular cell type (6,9,10). MS is an inflammatory disease and patients present with perturbed peripheral blood cytokine levels (26). It is thus likely that the differences in eQTL effect sizes in MS patients as compared to the full cohort with NINDs included (Supplementary Material, Fig. S1) are at least in part due to the exposure of PBMCs to inflammatory cytokines in MS patients. One of the gene with the highest eQTL effect increase in MS/CIS patients, was *PSMB9*, a proteasome subunit involved in processing of class I MHC peptides (27) and *TAP2* which help in antigen transport and function as a molecular scaffold for the final stage of MHC class I folding (28).

We analyzed the MS eQTLs in monocytes from healthy individuals stimulated with LPS and IFN- γ , which may to a degree mimic an inflammatory context similar to MS (29). Both are potent stimuli that boost cytokine production at 24 hours, thus activating additional signaling pathways in the monocytes. We have used the stimulation-specific eQTL data as a potential approximation of the context of chronic inflammation in MS patients, and for some of the genes with increased eQTL effects in MS, such as *IQGAP1* or *CLECL1*, we could indeed observe a similar pattern upon stimulation of monocytes with LPS and/or IFN- γ .

Gene functions and potential relevance in MS disease process

The non-HLA MS genes eQTLs in PBMCs represent a wide variety of functions and processes. Several are involved in the immune system, such as the well-established *IL7*, *TRAF3* (30,31), *MERTK* (32), *CLECL1* (33) and *TBKBP1* (34,35). A large proportion have unknown or poorly studied functions (e.g. *FAM164A*, *FAM213B*) and another group of genes are the non-translated ones, such as anti-sense RNA genes (*RP3-388E23.2.1*, *AP002954.4.1*, *CTD-232004.2.1*, *AP003774.1.1*) and pseudogenes (*RP11-578024.2.1*, *RP11-726G1.1.1*, *AC040934.1*, *KRT8P46*, *RP11-10L12.6.1*). Even though pseudogenes do not produce functional proteins, there is increasing evidence for regulatory roles for transcribed pseudogenes (36–38). In this study, we have observed significant downregulation of *FCRL3* in MS compared to NINDs. SNPs in the *FCRL3* locus have been identified as risk SNPs for MS and other autoimmune diseases such as autoimmune thyroid disease (Supplementary Material, Table S2). We also identify *FCRL3* as one of the eQTL genes in this study. Furthermore, there is an experimentally validated interaction between *FCRL3* and *PTPN22* (risk gene for rheumatoid arthritis) identified by the Protein–protein interaction (PPI) network provided in STRING database (39). Therefore, as it is differentially expressed it is a relevant candidate to be further characterized at the protein level. Recently, the possible role of *ANKRD55* in neuroinflammation was studied at protein expression level in neuron and microglia cultures based on an animal model of experimental autoimmune encephalomyelitis (EAE) (40). In our study, the eQTL SNP for *ANKRD55* is found to be located in an active regulatory region (H3K4me1) in PBMC and four PBMC-derived immune cell types with a predicted score for binding of STAT1 transcription factor and is also found to be associated to other autoimmune diseases such as Crohn's disease, Rheumatoid Arthritis and Juvenile Idiopathic Arthritis (3,41). Finally, *AHI1* (42,43), *MANBA* and *MMEL1* (44,45) have potential roles in the CNS, but could also be important in other cell types. Some genetic variants regulate the expression of several genes, however, the effects observed for the different genes may vary largely depending on the cell type that is being studied, and therefore it is not clear which gene drives the disease association.

eQTL effects in the *HLA-DRB1*15:01*-containing haplotypes have been investigated previously (46–49). We replicated the differences in expression of *HLA-DRB1*, *DRB5*, *DQB1* (46,48) and *DQA1* (46) in relation to *HLA-DRB1*15:01*, but we also find novel associations (e.g. *HCP5* and *HLA-B* association to *HLA-B*44* or 45). As *HLA-DRB5* is present exclusively on *HLA-DRB1*15* and *HLA-DRB1*16*-containing haplotypes, its expression has served as a positive control for the *HLA-DRB1*15* allele throughout all the tested cell types (Fig. 4). Presence of *HLA-DRB1*15:01* was significantly correlated with increased expression of *DRB1* in PBMCs, B cells and monocytes. Importantly, we also found genes affected by the presence of the *DRB1*03:01* risk allele (*HLA-DQA1*, *HLA-DQB1*, *PPP1R2P1*, *HLA-DRB5*, *HLA-DQB1-AS1*, *HLA-DQB2*, *PSMB9*) and the *A*02:01* protective allele (*IFITM4P*, *HLA-F-AS1*, *HLA-H*, *MICD*, *HLA-K*, *HCG4*, *HLA-T*, *ZDHHC20P1*), which to our knowledge, have not been reported before (Fig. 4).

Extensive efforts are presently being made to map eQTLs in different healthy immune cell subsets and stimulatory contexts. Disease-associated eQTLs in a disease-specific context is an important additional aspect that should be considered in light of our results on increased strength effect (absolute ρ) in HLA-related eQTLs. By using MS patient samples, we provide an

initial step of insight into potential SNP–gene relationships that lay a basis for a more extensive mapping in the context of MS. We present a detailed overview through which we provide a list of putative MS genes, confirm that these genes are not necessarily those located closest to the SNPs, vary in their response depending on the cell subset examined, and are influenced by disease or stimulatory conditions. This study serves as a resource of information for further functional exploration of genes and signaling pathways that are affected by MS-associated polymorphisms.

Materials and Methods

We have tested correlation between genotype and expression levels in several different cohorts which are briefly described in Table 3.

Patient selection in screening phase for PBMCs

For the screening-phase of the project, we obtained RNA and DNA from PBMC of 181 subjects collected between 2001 and 2010, at the Neurology Clinic of the Karolinska University Hospital, Solna, Sweden. The cohort consisted of 117 MS patients fulfilling the 2001 McDonald criteria (for consistency later revisions to these guidelines were not considered) (50), 28 patients with clinically isolated syndrome (CIS) and 36 patients were NINDs, i.e. patients suffering from neurological diseases without an inflammatory state, such as neuralgia, paresthesia, sensory symptoms, vertigo, tension headache (Supplementary Material, Table S1). About 24 out of 28 CIS patients were later diagnosed to have MS.

Preparation of PBMCs, CD4⁺ and CD8⁺ T cells

We isolated PBMCs immediately from blood samples taken with sodium citrate-containing cell preparation tubes (BD Vacutainer™ CPT™ Tube, Becton-Dickinson, Franklin Lakes, NJ), in the screening phase, or using a standard Ficoll (GE Healthcare, Little Chalfont, UK) procedure, in the validation phase. We separated the cells with density gradient centrifugation, collected them from the interphase, washed twice in Dubelcco's phosphate buffered saline (PBS) and either froze as a pellet at -80°C for subsequent RNA extraction, or prepared the cells for sorting. Sorting of the CD4⁺ and CD8⁺ T cell population was performed by adding fluorochrome conjugated antibodies against human CD4 (clone SK3, APC-conjugated, Becton Dickinson), CD8 (clone SK1, FITC-conjugated, Becton Dickinson) and CD3 (clone UCHT1, PE-conjugated, BD Bioscience) to the negative fraction obtained after manual sorting for CD14⁺ monocytes and using a MoFlo high speed cell sorter.

RNA isolation from PBMCs, CD4⁺ and CD8⁺ T cells

We performed extractions of PBMC RNA to be used for sequencing using the RNeasy Mini Kit (Qiagen, Hilden, Germany), and included DNase digestion, following the manufacturer's instructions. We eluted the RNA in 10 mM Tris-HCl, 1 mM EDTA buffer, pH 7.0 (Ambion, Austin, TX) and stored it in -80°C . We measured the RNA concentration on a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE), RNA was extracted from CD4⁺ and CD8⁺ T cells with Trizol (Invitrogen) using the miRNeasy Kit (Qiagen) according to manufacturer's recommendations, and analyzed the RNA

integrity with Bioanalyzer (Agilent Technologies, Santa Clara, CA). All samples had an RNA integrity number (RIN) >8.

DNA extraction and genotyping

For DNA extraction, we collected blood in EDTA tubes (Becton-Dickinson), which were stored at -20°C for subsequent procedures. We extracted DNA from whole blood samples using QIAamp DNA Blood Maxi Kit (Qiagen, Hilden, Germany), according to the manufacturer's protocol, and stored it at -20°C . An Illumina Infinium custom array, designed specifically to fine-map regions with established autoimmune-associated risk loci, as well as to test overlap in association between different autoimmune diseases, was used for genotyping. Genotyping of DNA samples in the screening phase (as a part of the ImmunoChip Project) was performed at the Wellcome Trust Sanger Institute and at the University of Miami (UM), Miller School of Medicine (described in detail in 3). Of relevance for the study described in this paper, we included the 109 established non-MHC risk loci typed on the ImmunoChip and showing the strongest independent association to MS in each region (3) (Supplementary Material, Table S2), in the subsequent eQTL analysis.

Samples used for validation and exploration of MS eQTLs in CD4⁺ and CD8⁺ T cells were genotyped with the Illumina custom array called MS replication chip at Miller School of Medicine, UM as part of an IMSGC project (51).

Imputed HLA types

For the samples from PBMC, we imputed HLA types with HLA*IMP:02 (52), using genotypes obtained as described earlier (ImmunoChip for cohort 1 and 2 and MS replication chip for CD4⁺ and CD8⁺ T cells). Alignment and phasing was done against the HapMap CEU panel. All samples included in subsequent analyses had an imputation quality score >0.7 for the HLA alleles of interest. Both the ImmunoChip and MS replication chip has extensive coverage of SNPs in the MHC region.

Transcriptome analysis in PBMC

cDNA libraries for PBMC samples (Cohort 1 and Cohort 2) were prepared using Illumina TruSeq kit (Illumina, San Diego) and sequenced on an Illumina HiSeq 2000 machine. A total of 100 base long paired-end sequence reads were generated on an average sequence depth of 36 million reads per sample for 181 samples (Table 3). The reads were mapped to the *H. sapiens* reference genome (NCBI v37, hg19) using STAR aligner (53) and then HTSeq tool (54) was used to quantify counts per gene, applying the default parameters in each case. Conditional Quantile Normalization (CQN) method (55) was used to normalize the gene count datasets accounting for the GC content and length biases.

Transcriptome analysis in CD4⁺ and CD8⁺ T cells

Of total RNA, 500 ng with RIN >8.0 was subjected to the Illumina TruSeq Stranded mRNA Library Preparation Protocol with Dual Indexes (Cat. No: RS-122-2013). Libraries were quantified using the Kapa Library Quantification Kit (Cat. No: KK4824). Sequencing was carried out on the Illumina HiSeq 2500 to generate 75 bp paired-end data with an average of 10M reads above Q30. The sequence reads were mapped to hg19 reference with TopHat and HTseq was used to quantify counts-per-gene. The

expression data from CD4⁺ cells (23 samples) and CD8⁺ cells (26 samples) was normalized with CQN method as described in PBMC analysis. Sample size estimation and power analysis was done for both the expression dataset using 'pwr.r.test' function implemented in 'pwr' R package.

Expression and genotype data from LCLs

We downloaded genotypes from the 1000 Genomes Phase 1 dataset from <http://www.1000genomes.org>. A detailed description of the dataset is provided in the study (56) (Table 3). We used tagging SNPs for the three HLA variants to be validated in our analysis (all were estimated in Swedish population; 979 individuals with HLA*IMP: 01 imputed genotypes): rs2970533 A tagging HLA-A*02:01 ($r^2=0.83$), rs2596551 C tagging HLA-B*44 and/or 45 ($r^2=0.89$), rs9271366 G tagging HLA-DRB1*15:01 ($r^2=0.98$) and rs2187668 T tagging HLA-DRB1*03:01 ($r^2=0.96$). rs201202118 was not present in the 1000 Genomes dataset and rs10431552 ($r^2=1$) was used as a proxy.

We downloaded RNA-Seq files in FASTQ format for EBV transformed LCLs from individuals included in the 1000 Genomes Phase 1 dataset <http://www.ebi.ac.uk/arrayexpress/experiments/E-GEUV-1/samples/>. More details about this public RNA-Seq dataset are found in (7). The reads were mapped to the *H. sapiens* reference genome (NCBI v37, hg19) using STAR aligner (53) and normalized using the CQN method. In total, 234 individuals (105 females and 129 males) with genotype and expression data were included in this study.

Expression and genotype data from B cells and monocytes

Both microarray-based mRNA expression and genotype data were obtained from the Fairfax *et al.* studies (8,16) for B cells (233 samples), unstimulated monocytes (414 samples) and stimulated monocytes—IFN- γ (367 samples), LPS 2 h (261 samples) and LPS 24 h (322 samples). All the samples collected in this study were obtained from healthy individuals (Table 3). The microarray expression data was obtained as log-2 transformed normalized data. For the missing MS-associated SNPs in the Fairfax datasets, proxy SNPs with high LD ($r^2>0.8$) were identified using the SNAP proxy search tool (57) and were used for the eQTL analysis (Supplementary Material, Tables S12 and S13). rs22837922, rs7665090 and rs4701192 were not present on the genotyping chip used in the study, and no appropriate proxy SNPs were available.

Expression and genotype data from PBMC array

Expression data for PBMC was obtained from the study (18) and 30 samples were genotyped as part of an IMSGC project (51). The expression data was reanalyzed using the latest array annotation data for HGU133Plus2 and expressed probes (after background correction) were reported in RMA measures after quantile normalization.

eQTL analysis of MS-associated SNPs

Different approaches implemented toward the transcriptome analysis for various cohorts used in this study (Table 3) are described in Supplementary Material. For each gene–SNP association, we did a two-level step analysis to compute the strength and significance of the association. First, we regressed

out the effects of a selected set of variables based on the result of component-based analysis and, depending on the dataset, available meta-information (Supplementary Material, Table S5). Secondly, residuals were correlated (non-parametric Spearman correlation) with genotype information considering a genetic additive model. For each SNP analyzed, we studied gene–SNP associations for the genes that were located 400-kb upstream and downstream of the SNP. The size of the genomic window was chosen based on the increased probability of finding cis-regulated genes at closer distances up to 400 kb (12,13) and with the aim to limit the number of comparisons. Only genes that were expressed in at least 85% of all the samples were included in the analysis.

To estimate the significance of each gene–SNP association, we computed a permutation-based P-value, where 15 000 permutations were done over the residuals of the regress-out model in the CQN transformed expression dataset. False discovery rate (FDR) was computed using a non-parametric method described elsewhere (58); in all cases monotonicity was enforced and FDR was capped at 1.

Gene–SNP associations with FDR < 0.01 were considered significant. Importantly, we selected a conservative selection criteria aiming for high specificity and low sensitivity in order to minimize false positives. In this study, the eQTL analysis was done separately for HLA region and non-HLA region. For a given SNP or allele, the minimum number of samples carrying homozygous minor allele to implement the eQTL analysis was limited to three. HLA alleles which were reported as risk for MS in immunochip study, such as absence of HLA-B*44:02, absence of HLA-B*38:01 and absence of HLA-B*55:01 were removed from the analysis due to low number of samples in one particular allele group.

For the regression model in the MS datasets (PBMC, CD4⁺ and CD8⁺ samples), we selected the variables based on Principal Component Analysis (PCA) from the meta-information available for each dataset (batch of RNA-Seq library preparation, age of patient, gender, disease-type, clinical course of MS and CIS and Interferon treatment). In the MS PBMC dataset (Cohort 1) and NINDs dataset (Cohort 2), we selected batch of RNA-Seq library preparation as covariate (Supplementary Material, Table S5). For the PBMC dataset with both MS and NINDs (Pooled Cohort) we selected batch of RNA-Seq library preparation and Disease-type (NINDs and MS with CIS) as covariates in the regression model. In the CD4⁺ and CD8⁺ datasets, we regressed out gender and disease type (MS and healthy) to compute the residuals (batch of RNA-Seq library preparation was corrected for in a previous step using the ComBat package) (59). For the datasets obtained from other studies of naïve B-cell, monocytes and stimulated monocytes and LCLs (Table 3), we used gender as a covariate in the regression model to compute the residuals. For the Fairfax dataset, residuals were calculated from log-transformed microarray gene expression.

Conditional analysis on LD SNPs

In the PBMC dataset for every significant SNP–gene pair association, we identified the SNPs in LD ($r^2>0.5$) and computed the cis-eQTL beta-value and P-value for the association as described earlier in the eQTL methods. For each new significant eQTL SNP (eQTL FDR < 0.01), a stepwise association model was applied where we regressed out the effect of most-significant LD SNPs along with other covariates (batch and/or disease-type depending availability) from the expression levels (CQN values). The

eQTL analysis was re-run on the remaining significant LD SNPs using the resulting residuals. In this stepwise association model, only the LD SNPs with a beta-value greater than the original MS SNP were included. To further characterize the cis-regulatory signals, we tested significant LD SNPs for the remaining eQTL effect after regressing it out with MS-associated SNP.

eQTL SNPs correlations to PBMC cell-type profile

The cell-type deconvolution method CIBERSORT (15) was applied to predict the proportion of different cell types in PBMC expression data from MS patients (Cohort 1). To understand the effect of different RNA-Seq normalization methods in predicting cell-type proportions, the CIBERSORT algorithm was first tested in RNA-Seq data obtained from CD4⁺ and CD8⁺ cells sorted from PBMCs, using the LM22 signature geneset (1000 iterations).

Subsequently, we applied the CIBERSORT to CQN-transformed expression values in the PBMC samples (Cohort 1), with the similar setting applied in sorted cell expression data. We selected only the cell types with a proportion >5% of the total PBMC mixture and aggregated them into five possible categories: CD4⁺ cell type ('T cells CD4⁺ naïve', 'T cells CD4⁺ memory activated' and 'T cells CD4⁺ memory resting'), CD8⁺ cell type ('T cells CD8⁺'), monocytes, B cells ('B cells naïve' and 'B cells memory') and NK cells ('NK cells resting' and 'NK cells activated'). Aggregated cell-type data was corrected for batch effects using a linear model and the residual values obtained after correction were used to correlate to the 31 non-HLA SNPs and 3 HLA variants associated with significant eQTL effects in the analysis of MS PBMCs. We considered significant cell type-SNP association those with P-value <0.05 for Spearman correlation.

For SNPs that were significantly correlated to one cell type, an association model was applied, where we regressed out the effect of the corresponding cell type along with the RNA-Seq batch covariate from the expression levels (CQN values). In case of SNPs that were associated to more than one cell type a stepwise association model was implemented where corresponding cell-type proportions were added as covariates in the order of cell type-SNP association strength. The eQTL analysis was re-ran on these selected SNPs using the resulting residuals to test the independence of the SNP-gene associations from the cell type-specific effects. A potential limitation of this analysis is that the power of the dataset is not large enough to correct for any non-MS associated variants in those regions that are correlated with cell proportions.

Colocalization analysis

Bayesian-based statistical method for colocalization analysis implemented in R, 'coloc' package was used to test the hypothesis that independent casual variants does not exist and eQTL genes can be attributed to the MS genetic association. The method calculates the posterior probabilities for shared causal signal (H4) and independent signals (H3) from both eQTL association and disease association in a given locus (60). The test was done in loci where a minimum of 10 independent SNPs were available after LD pruning ($LD r^2 < 0.8$), within ± 400 -kb windows of MS-associated SNP and the prior probabilities (p_1/p_2 and p_{12}) were set to the default values (1×10^{-4} and 1×10^{-5} , respectively). The MS disease association P-values were obtained from ImmunoChip study (3) and spearman correlated P-values for eQTL genes were obtained for all the SNPs in the locus using the

method described under eQTL analysis of MS-associated SNPs. For each SNP-gene the approximate bayes factor (abf) was calculated for H3 and H4 using 'coloc.abf' function. Then SNP-gene pairs with H3.abf greater than H4.abf were considered as not colocalized (Supplementary Material, Table S10).

Differential expression analysis

We conducted differential gene expression analysis using DESeq2 tool (61) implemented in R. Batch effect from the RNA-Seq library preparation is added as a covariate in the design (model) and an adjusted P-value cutoff of 0.05 is used to identify significant differentially expressed genes (Supplementary Material, Table S14).

Comparison of eQTL effect sizes

To study the disease specificity, we calculated the bootstrapping derived mean of rho values of the significant eQTLs identified in Cohort 1. Then we compared estimated rho from Cohort 1 and Pooled Cohort by computing the relative change (Tables 1 and 2 and Supplementary Material, Fig. S1). For comparisons, we considered only the eQTLs with sample distribution of minimum five samples in at least two genotype-distribution (Supplementary Material, Tables S21 and S22). Calculations for Supplementary Material, Figure S1 can be found in Supplementary Material, Tables S17 and S18. To test the change in direction of the eQTL effect, we computed the mean rho values (bootstrapping derived) in a small test array dataset (Table 3) from the study (18) and relative change (ratios) were estimated from MS cohort and pooled cohort (23 MS and 7 NIND). Increase or decrease of disease enrichment (direction of change) observed in the PBMC RNA-Seq cohort was compared with the one obtained in this small test cohort and the significance of deviations from original observation (PBMC RNA-Seq cohort) was reported using binomial exact test.

To study the modification of the eQTL effect by context in sorted cells, we compared the effect ratios between stimulated and unstimulated monocytes (Fig. 5). Only effects that were significant in the reference group (unstimulated monocytes) were included in the ratio calculations. Ratios showing an effect difference of less than $\pm 5\%$ were depicted with white color in the corresponding heatmaps, while larger effect differences were depicted according to the color scale. Calculations for Figure 5 are found in Supplementary Material, Table S13.

Epigenomic mapping of MS-associated variants

Narrow contiguous regions of enrichment (peaks) for DNase I hypersensitivity and Enhancer specific marks (H3K27ac and H3K4me1) based on ChromImpute (62) were obtained for PBMC and PBMC-derived immune cells (CD4⁺, CD8⁺, B cells and monocytes) from NIH Roadmap Epigenomics Mapping Consortium database (http://egg2.wustl.edu/roadmap/web_portal/index.html). Transcription factors that potentially bind to SNPs were obtained from the RegulomeDB database (63) and all these epigenetic marks were mapped to the MS-associated region.

Multiple disease association

ImmunoBase web-based tool (www.immunobase.org) was used to map other autoimmune diseases to the MS-associated SNPs.

Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgements

We thank all patients who have been willing to contribute with their blood samples and make this study possible. We acknowledge International Multiple Sclerosis Genetics consortium (IMSGC) for supplying genotypes from ImmunoChip and MS replication chip used in this study. We acknowledge the Science for Life Laboratory in Stockholm for bioinformatic support in the form of a collaborative project. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under project b2011139. We thank the Julian Knight group for sharing genotype and expression data from primary B cells and monocytes. We are grateful to Shahin Aeinehband for performing a part of the cell sorting, and to Izaura Lima Bomfim and Jenny Link for HLA imputation.

Conflict of Interest statement. T.O. has received unrestricted MS research grant, and/or honoraria for lectures and advisory boards from Biogen, Genzyme, Novartis, Merck, TEVA and Roche. All other authors declare that they have no competing interests.

Funding

This work was supported by grants from the Knut and Alice Wallenberg Foundation and the Swedish Association of Persons with Neurological Disabilities (Neuroförbundet), and Astra Zenica (AstraZeneca-Science for Life Laboratory collaboration) grant.

References

1. Westerlind, H., Ramanujam, R., Uvehag, D., Kuja-Halkola, R., Boman, M., Bottai, M., Lichtenstein, P. and Hillert, J. (2014) Modest familial risks for multiple sclerosis: a registry-based study of the population of Sweden. *Brain*, **137**, 770–778.
2. Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C.C. a., Patsopoulos, N. a., Moutsianas, L., Dilthey, A., Su, Z., Freeman, C. and Hunt, S.E. (2011) Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*, **476**, 214–219.
3. Beecham, A.H., Patsopoulos, N.A., Xifara, D.K., Davis, M.F., Kempainen, A., Cotsapas, C., Shah, T.S., Spencer, C., Booth, D., Goris, A. et al. (2013) Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.*, **45**, 1353–1360.
4. Moutsianas, L., Jostins, L., Beecham, A.H., Dilthey, A.T., Xifara, D.K., Ban, M., Shah, T.S., Patsopoulos, N. a., Alfredsson, L., Anderson, C.A. et al. (2015) Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nat. Genet.*, **47**, 1107–1113.
5. Kockum, I., Alferdsson, L. and Olsson, T. (2014) Genetic and environmental risk factors for multiple sclerosis—a role for interaction analysis. In Padyukov L. (ed.), *In between the Lines of Genetic Code, Genetic Interactions in Understanding Disease and Complex Phenotypes*, Academic Press, San Diego, CA, USA, London, UK and Waltham, MA, USA, pp. 124–133.
6. Farh, K.K.-H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shores, N., Whitton, H., Ryan, R.J.H., Shishkin, A.A. et al. (2015) Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, **518**, 337–343.
7. Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P. A., Monlong, J., Rivas, M. A., González-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G. et al. (2013) Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*, **501**, 506–511.
8. Fairfax, B.P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., Jostins, L., Plant, K., Andrews, R., McGee, C. et al. (2014) Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science*, **343**, 1246949.
9. Hu, X., Kim, H., Raj, T., Brennan, P.J., Trynka, G., Teslovich, N., Slowikowski, K., Chen, W.-M., Onengut, S., Baecher-Allan, C. et al. (2014) Regulation of gene expression in autoimmune disease loci and the genetic basis of proliferation in CD4+ effector memory T cells. *PLoS Genet.*, **10**, e1004404.
10. Raj, T., Rothamel, K., Mostafavi, S., Ye, C., Lee, M.N., Replogle, J.M., Feng, T., Lee, M., Asinovski, N., Frohlich, I. et al. (2014) Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science*, **344**, 519–523.
11. Romme Christensen, J., Börnsen, L., Hesse, D., Krakauer, M., Sørensen, P.S., Søndergaard, H.B. and Sellebjerg, F. (2012) Cellular sources of dysregulated cytokines in relapsing-remitting multiple sclerosis. *J. Neuroinflammation*, **9**, 215.
12. Folkersen, L., Hooft, F. v., Chernogubova, E., Agardh, H.E., Hansson, G.K., Hedin, U., Liska, J., Syvänen, A.-C., Paulsson-Berne, G., Franco-Cereceda, A. et al. (2010) Association of genetic risk variants with expression of proximal genes identifies novel susceptibility genes for cardiovascular disease. *Circ. Cardiovasc. Genet.*, **3**, 365 LP–373.
13. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J. et al. (2012) Systematic localization of common disease-associate variation in regulatory DNA. *Science*, **337**, 1190–1195.
14. Orrù, V., Steri, M., Sole, G., Sidore, C., Virdis, F., Dei, M., Lai, S., Zoledziowska, M., Busonero, F., Mulas, A. et al. (2013) Genetic variants regulating immune cell levels in health and disease. *Cell*, **155**, 242–256.
15. Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M. and Alizadeh, A.A. (2016) Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods*, **21**, 193–201.
16. Fairfax, B.P., Makino, S., Radhakrishnan, J., Plant, K., Leslie, S., Dilthey, A., Ellis, P., Langford, C., Vannberg, F.O. and Knight, J.C. (2012) Genetics of gene expression in primary immune cells identifies cell type – specific master regulators and roles of HLA alleles. *Nat. Genet.*, **44**, 1–10.
17. Chuluundorj, D., Harding, S.A., Abernethy, D. and La Flamme, A.C. (2014) Expansion and preferential activation of the CD14(+)CD16(+) monocyte subset during multiple sclerosis. *Immunol. Cell Biol.*, **92**, 509–517.
18. Brynedal, B., Khademi, M., Wallström, E., Hillert, J., Olsson, T. and Duvefelt, K. (2010) Gene expression profiling in multiple sclerosis: a disease of the central nervous system, but with relapses triggered in the periphery? *Neurobiol. Dis.*, **37**, 613–621.
19. Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R. et al. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.

20. Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J. et al. (2015) Integrative analysis of 111 reference human epigenomes. *Nature*, **518**, 317–330.
21. Lovato, L., Willis, S.N., Rodig, S.J., Caron, T., Almendinger, S.E., Howell, O.W., Reynolds, R., O'Connor, K.C. and Hafler, D.A. (2011) Related B cell clones populate the meninges and parenchyma of patients with multiple sclerosis. *Brain*, **134**, 534–541.
22. Von Büdingen, H.C., Kuo, T.C., Sirota, M., Van Belle, C.J., Apeltsin, L., Glanville, J., Cree, B.A., Gourraud, P.A., Schwartzburg, A., Huerta, G. et al. (2012) B cell exchange across the blood-brain barrier in multiple sclerosis. *J. Clin. Invest.*, **122**, 4533–4543.
23. Bar-Or, A., Nuttall, R.K., Duddy, M., Alter, A., Kim, H.J., Ifergan, I., Pennington, C.J., Bourgoin, P., Edwards, D.R. and Yong, V.W. (2003) Analyses of all matrix metalloproteinase members in leukocytes emphasize monocytes as major inflammatory mediators in multiple sclerosis. *Brain*, **126**, 2738–2749.
24. Hauser, S.L., Bar-Or, A., Comi, G., Giovannoni, G., Hartung, H.-P., Hemmer, B., Lublin, F., Montalban, X., Rammohan, K.W., Selmaj, K. et al. (2017) Ocrelizumab versus interferon beta-1a in relapsing multiple sclerosis. *N. Engl. J. Med.*, **376**, 221–234.
25. Corradin, O., Saiakhova, A., Akhtar-Zaidi, B., Myeroff, L., Willis, J., Cowper-Sal{middle dot}lari, R., Lupien, M., Markowitz, S. and Scacheri, P.C. (2014) Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res.*, **24**, 1–13.
26. Navikas, V. and Link, H. (1996) Review: cytokines and the pathogenesis of multiple sclerosis. *J. Neurosci. Res.*, **45**, 322–333.
27. Kelly, A., Powis, S.H., Glynn, R., Radley, E., Beck, S. and Trowsdale, J. (1991) Second proteasome-related gene in the human MHC class II region. *Nature*, **353**, 667–668.
28. de la Salle, H., Hanau, D., Fricker, D., Urlacher, A., Kelly, A., Salamero, J., Powis, S., Donato, L., Bausinger, H. and Laforet, M. (1994) Homozygous human TAP peptide transporter mutation in HLA class I deficiency. *Science*, **265**, 237–241.
29. Jagodic, M., Colacios, C., Nohra, R., Dejean, A.S., Beyeen, A.D., Khademi, M., Casemayou, A., Lamouroux, L., Duthoit, C., Papapietro, O. et al. (2009) A role for VAV1 in experimental autoimmune encephalomyelitis and multiple sclerosis. *Sci. Transl. Med.*, **1**, 10ra21.
30. Zarnegar, B.J., Wang, Y., Mahoney, D.J., Dempsey, P.W., Cheung, H.H., He, J., Shiba, T., Yang, X., Yeh, W.-C. and Mak, T.W. (2008) Noncanonical NF-kappaB activation requires coordinated assembly of a regulatory complex of the adaptors cIAP1, cIAP2, TRAF2 and TRAF3 and the kinase NIK. *Nat. Immunol.*, **9**, 1371–1378.
31. Hu, H., Brittain, G.C., Chang, J.-H., Puebla-Osorio, N., Jin, J., Zal, A., Xiao, Y., Cheng, X., Chang, M., Fu, Y.-X. et al. (2013) OTUD7B controls non-canonical NF-kB activation through deubiquitination of TRAF3. *Nature*, **494**, 371–374.
32. Cabezón, R., Carrera-Silva, E.A., Flórez-Grau, G., Errasti, A.E., Calderón-Gómez, E., Lozano, J.J., España, C., Ricart, E., Panés, J., Rothlin, C.V. et al. (2015) MERTK as negative regulator of human T cell activation. *J. Leukoc. Biol.*, **97**, 751–760.
33. Ryan, E.J., Marshall, A.J., Magaletti, D., Floyd, H., Draves, K.E., Olson, N.E. and Clark, E.A. (2002) Dendritic cell-associated lectin-1: a novel dendritic cell-associated, C-type lectin-like molecule enhances T cell secretion of IL-4. *J. Immunol.*, **169**, 5638–5648.
34. Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P.O., Bergamini, G., Croughton, K., Cruciat, C., Eberhard, D., Gagneur, J., Ghidelli, S. et al. (2004) A physical and functional map of the human TNF-alpha/NF-kappaB signal transduction pathway. *Nat. Cell Biol.*, **6**, 97–105.
35. Li, S., Wang, L., Berman, M., Kong, Y.Y. and Dorf, M.E. (2011) Mapping a dynamic innate immunity protein interaction network regulating type I interferon production. *Immunity*, **35**, 426–440.
36. Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M. et al. (2008) Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature*, **453**, 534–538.
37. Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W.J. and Pandolfi, P.P. (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*, **465**, 1033–1038.
38. Johnsson, P., Ackley, A., Vidarsdottir, L., Lui, W.-O., Corcoran, M., Grandér, D. and Morris, K.V. (2013) A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. *Nat. Struct. Mol. Biol.*, **20**, 440–446.
39. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P. et al. (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.*, **43**, D447–D452.
40. Lopez de Lapuente, A., Feliú, A., Ugidos, N., Mecha, M., Mena, J., Astobiza, I., Riera, J., Carillo-Salinas, F., Comabella, M., Montalban, X. et al. (2016) Novel insights into the multiple sclerosis risk gene ANKRD55. *J. Immunol.*, **196**, 4553–4565.
41. Onengut-Gumuscu, S., Chen, W.M., Burren, O., Cooper, N.J., Quinlan, A.R., Mychaleckyj, J.C., Farber, E., Bonnie, J.K., Szpak, M. and Schofield, E. et al. (2015) Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat. Genet.*, **47**, 381–386.
42. Dixon-Salazar, T., Silhavy, J.L., Marsh, S.E., Louie, C.M., Scott, L.C., Gururaj, A., Al-Gazali, L., Al-Tawari, A.A., Kayserili, H., Sztrika, L. et al. (2004) Mutations in the AHI1 gene, encoding joubertin, cause Joubert syndrome with cortical polymicrogyria. *Am. J. Hum. Genet.*, **75**, 979–987.
43. Zhou, L.L., Zhao, Y., Ringrose, A., DeGeer, D., Kennah, E., Lin, A.E.-J., Sheng, G., Li, X.-J., Turhan, A. and Jiang, X. (2008) AHI-1 interacts with BCR-ABL and modulates BCR-ABL transforming activity and imatinib response of CML stem/progenitor cells. *J. Exp. Med.*, **205**, 2657–2671.
44. Ouimet, T., Facchinetti, P., Rose, C., Bonhomme, M.C., Gros, C., Schwartz, J.C. and Tanja, O. (2000) Nephrilysin II: a putative novel metalloprotease and its isoforms in CNS and testis. *Biochem. Biophys. Res. Commun.*, **271**, 565–570.
45. Huang, J.Y., Hafez, D.M., James, B.D. and Bennett, D.A. and Marr R.A. (2012) Altered NEP2 expression and activity in mild cognitive impairment and Alzheimer's disease. *J. Alzheimers Dis.*, **28**, 433–441.
46. Handel, A.E., Handunnetthi, L., Berlanga, A.J., Watson, C.T., Morahan, J.M. and Ramagopalan, S.V. (2010) The effect of single nucleotide polymorphisms from genome wide association studies in multiple sclerosis on gene expression. *PLoS One*, **5**, e10142.

47. Irizar, H., Muñoz-Culla, M., Zuriarrain, O., Goyenechea, E., Castillo-Triviño, T., Prada, A., Saenz-Cuesta, M., De Juan, D., Lopez de Munain, A., Olascoaga, J. and Otaegui, D. (2012) HLA-DRB1*15: 01 and multiple sclerosis: a female association? *Mult. Scler.*, **18**, 569–577.
48. Alcina, A., de Abad-Grau, M.M., Fedetz, M., Izquierdo, G., Lucas, M., Fernández, Ó., Ndagire, D., Catalá-Rabasa, A., Ruiz, A. and Gayán, J. (2012) Multiple sclerosis risk variant HLA-DRB1*1501 associates with high expression of DRB1 gene in different human populations. *PLoS One*, **7**, e29819.
49. Apperson, M.L., Tian, Y., Stamova, B., Ander, B.P., Jickling, G.C., Agius, M.A. and Sharp, F.R. (2013) Genome wide differences of gene expression associated with HLA-DRB1 genotype in multiple sclerosis: a pilot study. *J. Neuroimmunol.*, **257**, 90–96.
50. McDonald, W.I., Compston, A., Edan, G., Goodkin, D., Hartung, H.-P., Lublin, F.D., McFarland, H.F., Paty, D.W., Polman, C.H., Reingold, S.C. et al. (2001) Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the Diagnosis of Multiple Sclerosis. *Ann. Neurol.*, **50**, 121–127.
51. International Multiple Sclerosis Genetics Consorti, Patsopoulos, N., Baranzini, S.E., Santaniello, A., Shoostari, P., Cotsapas, C., Wong, G., Beecham, A.H., James, T., Replogle, J. et al. (2017) The Multiple Sclerosis Genomic Map: Role of peripheral immune cells and resident microglia in susceptibility. *bioRxiv*, 10.1101/143933.
52. Dilthey, A., Leslie, S., Moutsianas, L., Shen, J., Cox, C., Nelson, M.R. and McVean, G. (2013) Multi-Population Classical HLA Type Imputation. *PLoS Comput. Biol.*, **9**, e1002877.
53. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-Seq aligner. *Bioinformatics*, **29**, 15–21.
54. Anders, S., Pyl, P.T., Huber, W., Camp, J.G., Vernot, B., Köhler, K., Voigt, B., Okita, K., Maricic, T., He, Z. et al. (2015) HTSeq – a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.
55. Hansen, K.D., Irizarry, R.A. and Wu, Z. (2012) Removing technical variability in RNA-Seq data using conditional quantile normalization. *Biostatistics*, **13**, 204–216.
56. McVean, G.A., Altshuler (Co-Chair), D.M., Durbin (Co-Chair), R.M., Abecasis, G.R., Bentley, D.R., Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., Flicek, P. et al. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.
57. Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J. and de Bakker, P.I.W. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*, **24**, 2938–2939.
58. Li, J. and Tibshirani, R. (2013) Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. *Stat. Methods Med. Res.*, **22**, 519–536.
59. Johnson, W.E., Li, C. and Rabinovic, A. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, **8**, 118–127.
60. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., Plagnol, V. and Williams, S.M. (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.*, **10**, e1004383.
61. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol.*, **15**, 550.
62. Ernst, J. and Kellis, M. (2015) Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat. Biotechnol.*, **33**, 364–376.
63. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C. and Weng, S. (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.*, **22**, 1790–1797.