Published in final edited form as:
Ann Neurol. 2011 December ; 70(6): 897–912. doi:10.1002/ana.22609.

Genomewide meta-analysis identifies novel multiple sclerosis susceptibility loci

Nikolaos A. Patsopoulos, MD, PhD[1,2,3], **the BSP MS Genetics working group and the steering committees of studies evaluating IFNb-1b and a CCR1-antagonist, ANZgene Consortium, GeneMSA, International Multiple Sclerosis Genetics Consortium**, and **Paul I.W. de Bakker, PhD**[2,3,4,5]

[1]Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Department of Neurology, Brigham & Women's Hospital, Boston, MA 02115, USA [2]Division of Genetics, Department of Medicine, Brigham & Women's Hospital, Harvard Medical School, Boston, MA 02115, USA [3]Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA 02139, USA [4]Department of Medical Genetics, Division of Biomedical Genetics, University Medical Center, Utrecht, The Netherlands [5]Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, The Netherlands

Abstract

Objective—To perform a one-stage meta-analysis of genome-wide association studies (GWAS) of multiple sclerosis (MS) susceptibility and explore functional consequences of new susceptibility loci.

Methods—We synthesized 7 MS GWAS. Each dataset was imputed using HapMap phase II and a per-SNP meta-analysis was performed across the 7 datasets. We explored RNA expression data using a quantitative trait analysis in peripheral blood mononuclear cells (PBMCs) of 228 subjects with demyelinating disease.

Results—We meta-analyzed 2,529,394 unique SNPs in 5,545 cases and 12,153 controls. We identified three novel susceptibility alleles: rs170934^{T} at 3p24.1 (OR=1.17, $P = 1.6 \times 10^{-8}$) near *EOMES*, rs2150702^{G} in the second intron of *MLANA* on chromosome 9p24.1 (OR = 1.16, $P = 3.3 \times 10^{-8}$), and rs6718520^{A} in an intergenic region on chromosome 2p21, with *THADA* as the nearest flanking gene (OR = 1.17, $P = 3.4 \times 10^{-8}$). The three new loci do not have a strong "*cis*" effect on RNA expression in PBMCs. Ten other susceptibility loci had a suggestive $P<1\times10^{-6}$, some of which have evidence of association in other inflammatory diseases, i.e. *IL12B*, *TAGAP*, *PLEK*, and *ZMIZ1*.

Interpretation—We have performed a meta-analysis of GWAS in MS that more than doubles the size of previous gene discovery efforts and highlights three novel MS susceptibility loci. These and additional loci with suggestive evidence of association are excellent candidates for further investigations to refine and validate their role in the genetic architecture of MS.

[5]Department of Neurology, Scientific Institute San Raffaele, 20132 Milan, Italy

[34]University Hospital Bochum/Germany

[59]School of Medicine, Department of Neurology, Flinders University of SouthAustralia, Bedford Park, Adelaide, SA, AUS 5042.

INTRODUCTION

Multiple sclerosis (MS) is thought to emerge when genetically susceptible individuals encounter environmental triggers that initiate an inflammatory reaction against self-antigens in the central nervous system (CNS); these events result in recurring episodes of inflammatory demyelination and, in many cases, a progressive neurodegenerative process.[1] The genetic architecture underlying susceptibility to MS is complex, and recent efforts have revealed over a dozen susceptibility loci of modest effect[2–5] in addition to the major histocompatibility complex (MHC) that contains multiple independent susceptibility alleles in class I and class II loci. These discoveries were made possible by the emergence of genome-wide genotyping technologies and collaborative meta-analysis efforts to maximize statistical power.

Here, we perform a one-stage meta-analysis, of most of the genome-wide single nucleotide polymorphism (SNP) data generated in the field of MS, to conduct the largest gene discovery effort to date for this disease.

MATERIALS & METHODS

The present study comprises seven data sets of non-overlapping case and control subjects of European descent. Five strata (IMSGC-US, IMSGC-UK, GeneMSA-US, GeneMSA-NL, and GeneMSA-CH) were taken from a previously published meta-analysis.[3] Details on these data sets can be found elsewhere.[2–3] The sixth stratum (BWH/TT) is based on data from our previous study[3] enriched with additional 1453 MS cases and 2176 controls, with all samples genotyped on the Affymetrix GeneChip 6.0 platform. Finally, we have added another stratum (ANZ) from a recently described genome-wide study containing 1618 cases and 1988 controls.[5] Further information on these two strata and quality control applied can be found in the supplementary material. Table 1 summarizes the subject collections that have been assembled for this meta-analysis. All subjects met either 1) a diagnosis of MS per McDonald criteria[6] or 2) a diagnosis of clinically isolated demyelinating syndrome (CIS) in which individuals have had one episode of inflammatory demyelination and harbor two or more T2 hyperintense lesions in their brain or spinal cord. The majority of CIS subjects go on to have a second episode of inflammatory demyelination, which results in a diagnosis of MS. An earlier study did not find differences in the distribution of susceptibility alleles in CIS and MS subjects, suggesting that their genetic architecture is similar.[7]

We used EIGENSOFT to remove outliers in terms of genetic ancestry and to calculate the top ten eigenvectors of the genotype data within each stratum.[8] The seven data sets were genotyped using different genotyping platforms. To maximize genome-wide coverage, we used the imputation algorithm implemented in MACH to yield 2.5 million SNPs across the genome in all data sets.[9] Imputation based on linkage disequilibrium (LD) patterns observed in a representative European population sample in HapMap is a widely used approach to increase the power of GWAS and facilitate *in silico* meta-analysis.[10] After imputation, we excluded all SNPs with an imputation quality score less or equal to 0.10 or minor allele frequency (MAF) <0.01 per stratum. For each stratum, we tested the imputed dosages for association to case-control status using logistic regression, including the ten first eigenvectors as covariates to correct for population stratification. For each SNP, the dosage corresponds to the (imputed) number of the coded allele in a given individual and varies from 0 to 2 on a continuous scale, thus incorporating information about the imputation uncertainty. Under a per-allele model, we calculated the odds ratio (OR), its corresponding standard error (SE) and p-value. To evaluate the robustness of the observed distribution of the test statistic, we inspected the quantile- quantile (Q-Q) plot and calculated the genomic inflation factor (λ_{GC}).[11] To correct for any residual, unexplained inflation of the test

statistic, we corrected the SEs by multiplying them with the square root of the λ_{GC}.[10] Finally, we performed the same analyses adjusting for sex and report these sensitivity analyses in the Supplementary material.

Ensuring consistency in the strand orientation of the alleles across all strata, we meta-analyzed the ORs with the respective corrected SEs using inverse variance weighting under a fixed-effects model. We calculated the λ_{GC} of the genome-wide association results to evaluate the robustness of the meta-analysis. Furthermore, in a secondary sensitivity analysis, we used a random effects model to meta-analyze, thus allowing for between-study heterogeneity. We used Cochran's Q to test for the presence of statistical heterogeneity and I^2, with respective 95% confidence intervals, to quantify inconsistency of effects across the different strata.[12, 13]

We performed linkage disequilibrium (LD) pruning ($r^2 > 0.5$) among correlated SNPs to identify the most statistically significant SNP in regions of strong LD. For non-MHC loci, we performed conditional analysis using a forward stepwise logistic regression for the most statistically significant SNPs ($P < 10^{-6}$) within a 2-Mb distance from the best index SNP (with the lowest p-value) at a locus.

We pre-defined genome-wide significance for our meta-analysis at a P-value of $< 5\times10^{-8}$. At this type I error rate, and under a fixed effect model, we have more than 80% power to detect an OR of 1.15 for a risk allele with 0.4 minor allele frequency. The Cochran's Q test was considered to be statistically significant at $P < 0.10$. For analyses we used the PLINK v1.07[14] and R-2.11.

To leverage the rapidly growing list of susceptibility loci associated with inflammatory diseases, we tested all known SNP associations with Crohn's disease (CD), ulcerative colitis (UC), celiac disease (CE), type 1 diabetes (T1D), rheumatoid arthritis (RA), systemic lupus erythematosus (SLE) and psoriasis (PS) for a role in MS. To identify these *bona fide* associations, we searched the online NHGRI catalog (www.genome.gov/26525384) and PubMed for GWAs, meta-analyses of GWAS, or follow-up studies that reported a non-MHC SNP for these diseases with genome-wide significance ($P < 5\times10^{-8}$). For each of these SNPs we tested for replication of an effect using different p-value thresholds (0.05, 10^{-3}, and 10^{-4}), allowing for heterogeneity in the direction of effect. As a comparison, we also list the genome-wide non-MHC SNPs associated with type 2 diabetes (T2D),[15] height (HI),[16] lipid traits (LI),[17] and myocardial infraction (MI)[18–20] as negative control diseases, because we do not expect these diseases to have an etiologic relationship with MS.

For all SNPs that reached a $P < 10^{-6}$, we also performed a meta-analysis under a recessive and dominant model. We used the posterior probabilities for each of the three genotypes (AA, AB, BB) from the imputations to calculate the corresponding dominant and recessive dosage in each individual for each SNP. With these dosages, we calculated the per-stratum ORs and corrected SEs, and meta-analyzed these to obtain the overall ORs and the corresponding p-values.

For the newly identified susceptibility loci, we sought to test the hypothesis that the identified SNPs can affect expression levels of nearby genes (within 1 Mb upstream and downstream of the SNP). We collected RNA expression data with an Affymetrix U133 v2.0 array from peripheral blood mononuclear cells (PBMCs) of 228 subjects with Relapsing Remitting (RR) MS or CIS. These data were collected between July 2002 and October 2007, as part of the Comprehensive Longitudinal Investigation of MS at the Brigham and Women's Hospital.[21,22] We regressed the observed gene expression on the SNP imputed dosages, adjusted for the treatment used. The probes that passed our quality check criteria (n=20,517) were used for the subsequent analyses. In an exploratory analysis, we performed

an eQTL analysis of all of probes for each newly identified loci and organized the tail of the distribution of the results, i.e. probes that reached a nominal significance threshold ($p < 0.05$), using the Ingenuity Pathway Analysis (IPA) software. Ingenuity maps probe IDs to its database and performs statistical computing to identify the most significant canonical pathways and networks overrepresented in a given gene list as compared with the whole list of genes in the Human Genome U133 Plus 2.0 array. The canonical pathway analysis tool identified the pathways from the IPA library of canonical pathways that were most significant to the dataset, based upon genes within the dataset that were associated with a canonical pathway in the Ingenuity Pathways Knowledge Base. In a similar way, the software leveraged the input gene expression data to provide networks. Specifically, molecules of interest, which interact with each other, and molecules in the Ingenuity Knowledge Base were identified as Network Eligible Molecules, which served as "seeds" for generating networks.

RESULTS

Overall, 5,545 cases and 12,153 controls passed QC and were included in the meta- analysis (Table 1), and 2,529,394 unique SNPs were analyzed in at least two strata. The genomic inflation factor (λ_{GC}) for the seven strata ranged from 1.026 to 1.061 (Table 1), suggesting that population stratification within the individual strata was limited. The genomic inflation factor of the genome-wide meta-analysis results was 1.051, indicating that the test statistic distribution is well calibrated and that the extent of residual bias (including unaccounted for stratification and technical artifact) is minimal.

Genome-wide significant SNPs under a per-allele genetic model ($P<5\times10^{-8}$)

Of 2,617 SNPs that reach genome-wide significance ($P < 5.0 \times 10^{-8}$), 2,583 SNPs are located within the MHC on chromosome 6p21, where index SNP rs3129889 (OR = 2.97, $P = 1.03\times10^{-206}$) and rs9260489 (OR = 1.21, $P = 1.16\times10^{-11}$) tag the *HLA-DRB1* and *HLA-B* associations, respectively.[3] Outside of the MHC, we observed SNP associations at seven loci with genome-wide significance (Figure 1, Table 2, Supplementary Table 2 for the sex-adjusted analyses). Three of these loci have not been described previously as being associated with MS or other inflammatory diseases. First, rs170934 at locus 3p24.1 ($P = 1.6 \times 10^{-8}$) demonstrated an OR of 1.17 for the minor T allele, with little evidence for statistical heterogeneity ($I^2 = 2\%$; $P = 4.7 \times 10^{-8}$ under a random-effects model). Figure 2A shows the regional association plot for this SNP, which is located in an intergenic area between *CMC1*, a gene with no known function, and *EOMES*, a T-box gene family member and a paralog of *TBX21/TBET*.

The second novel locus we have identified is tagged by rs2150702, a SNP in the second intron of the *MLANA* gene on chromosome 9p24.1, which is known as a melanoma antigen (Figure 2C, D). The minor G allele of this SNP increases risk (OR = 1.16) with no evidence for statistical heterogeneity across the seven strata ($P = 3.3 \times 10^{-8}$ under both fixed and random-effects models).

The third novel locus is tagged by rs6718520, which maps to an intergenic region on chromosome 2p21, with *THADA* as the nearest flanking genes at 132kbps distance (Figure 2E). The minor A allele of this SNP has an OR of 1.17 ($P = 3.4 \times 10^{-8}$) with modest evidence for statistical heterogeneity ($I^2 = 35\%$, $P = 2.6 \times 10^{-4}$ under a random-effects model). This heterogeneity may come from the fact that the quality of this SNP's imputation varies across the strata of the meta-analysis, that residual population substructure in some strata influences our analysis, or it may be true heterogeneity in the effect of the SNP in different groups of subjects. The stronger per-dataset effect was observed in a low imputation quality dataset (Gene MSA NL, Figure 2F), so we explored the influence of the

imputation's quality on the statistical heterogeneity and the overall effect size (Supplemental Table 3). Under the most conservative scenario of synthesizing only the high imputation quality studies (INFO>0.9) the p-value was 8.44×10^{-6}, although suffering a huge power loss due to the smaller sample size.

The other four loci with genome-wide significant SNPs correspond to already known MS susceptibility loci. At the *IL2RA* gene, we observed two index SNPs (rs2104286 and rs7089861), with independent effects (r^2=0.128 in HapMap-CEU) that correspond to the two previously described independent effects within the *IL2RA* locus.[23] Another SNP, rs1335532, tags the known association in the *CD58* locus (r^2=0.87 with rs2300747, the previously described best marker).[21] The fourth SNP, rs2293152, captures the recently described association of the *STAT3* locus with MS susceptibility.[24] Many of the subjects are shared between the present study and the original discoveries of these three loci, so the findings here do not constitute independent evidence for replication of these loci. In addition to the *CD58*, *IL2RA* and *STAT3* loci that reach genome-wide significance in the present meta-analysis, we have tabulated the association results for all MS susceptibility loci reported to date (Table 2). We observe substantial evidence of association for all known loci with three exceptions (*KIF1B*,[25] *CBLB*,[4] and *chr5p15.32*[26]) that do not reach nominal statistical significance (Table 2).

SNPs with suggestive evidence of association ($p < 10^{-6}$)

Ten additional SNPs present strongly suggestive evidence of association with MS (defined as $p < 10^{-6}$) (Table 2). Two of these are found the *IL12B* locus, which is known to be associated with Crohn's disease and psoriasis (Table 3, supplementary table 1).[27,28] Another 3 loci -- *TAGAP* (rs1738074), *ZMIZ1* (rs1250542), and *PLEK* (rs7592330) -- have been described previously as susceptibility loci for other inflammatory diseases (Table 3),[29,30] and *ZMIZ1* has been described previously as having suggestive evidence of association with MS.[3] Of the remaining loci, two contain genes that have long been described as being involved in the immune dysregulation seen in MS: *TBX21* and *CD86*. The rs8070463 SNP (OR = 0.87, p-value = 9.55×10^{-8}) lies within 42kb of *TBX21*, which is also known as *TBET* and is a paralog of the *EOMES* gene found in the novel 3p24.1 susceptibility locus described above. *TBX21* is of great interest in T cell function because it is a master regulatory gene necessary for the differentiation of pathogenic Th1 lymphocytes that play an important role in murine inflammatory demyelination.[31,32] Nonetheless, this locus contains another gene, *TBKBP1*, that is involved in *NFKB* signaling and could also be implicated in the effect of this locus; functional dissection will be required to differentiate the role of the different genes found in this locus. The second locus of immunological interest is tagged by rs2681424 (OR =1.16, p-value=2.33×10^{-7}) and contains *CD86*, a costimulatory molecule that is the receptor for the CD28 and CTLA4 molecules and is an important component of the machinery regulating the activation of T cells [33]. These two genes are found in loci associated with susceptibility to rheumatoid arthritis (*CD28*) and T1D as well as thyroiditis (*CTLA4*).[34]

Susceptibility allele overlap between MS and other inflammatory diseases

Motivated by recent progress in the identification of susceptibility loci of other inflammatory diseases, we sought to test the effect of all non-MHC SNPs described to date as *bona fide* inflammatory disease associated polymorphisms. Of these 145 SNPs, 48 are associated with MS risk at nominal significance, consistent with a shared genetic etiology of inflammatory diseases (Table 3, Supplementary Table 1, Figure 3A). Of all the listed inflammatory diseases, SNPs identified in ulcerative colitis and celiac disease seem to replicate better in MS (60% and 45% at a nominal level, respectively. Figure 3A). Under more stringent statistical significance cut-offs ($< 10^{-3}$ and $< 10^{-4}$) few SNPs replicated in

our MS dataset, with celiac disease being the most replicated (34% at $<10^{-3}$ and 15% at 10^{-4}, Figure 3A). It is interesting to note that, for celiac disease (and most of the other inflammatory diseases) a minority of loci appears to demonstrate association with MS, but in the opposite direction such that a celiac disease risk allele is protective for MS. While such opposite effects have been noted before, we gain an appreciation that they are widespread and that they highlight the complexity of the shared architecture among the inflammatory disease. One example of this complexity is the *TAGAP* locus in which rs1738074 reaches a p-value $< 10^{-6}$ in MS (Table 2); this locus has attained genome-wide significance in celiac disease[35] (opposite direction of effect relative to MS) and type 1 diabetes[30] (same direction of effect as MS) (Table 3, Figure 3B). Further, within the *TAGAP* locus, another SNP (rs212389, r^2=0.268 with rs1738074) that is associated with susceptibility to RA[34] has the same direction of effect in MS (Table 3, Figure 3B). Given the extended evidence of association of the *TAGAP* locus in many inflammatory diseases, it is a strong candidate MS susceptibility locus, although it has not reached a level of genome-wide significance in our study.

Analysis under dominant and recessive genetic models

For the SNP associations with $P < 10^{-6}$, we also explored dominant and recessive genetic models (Supplementary Table 2) to assess whether these models fit our data better than our default additive model.[36] In several cases, association is enhanced when considering a different model, which could guide the design of replication studies. For example, at the 2p21 locus, the recessive model for the minor rs6718520^A allele is slightly more significant (OR = 1.29, $P = 2.9 \times 10^{-8}$) than the additive model (OR = 1.17, $P = 3.4 \times 10^{-8}$). For two other SNPs, rs1335532 in *CD58* and rs2293152 in *STAT3*, the dominant model was more significant than the additive model. Finally, SNP rs9901869, which is found in an intergenic region on chromosome 17 near *NPEPPS* gene, reached genome-wide significance under the recessive model for the minor G allele (OR = 1.26, $P = 3.7 \times 10^{-8}$ under both fixed and random-effects models). Thus, considering non-additive models for certain loci appears to be warranted and informs the potential mechanism of a variant's effect.

Transcriptional and exploratory pathway analysis

To explore the functional consequences of our new MS susceptibility loci, we assessed whether each of the three index SNPs influenced the level of RNA expression from genes located in the vicinity of each SNP (*cis* expression quantitative trait locus (eQTL) analysis). We performed these analyses in a set of 228 subjects with demyelinating disease on which we have obtained a genome-wide RNA expression profile from PBMC. While a few suggestive *cis* associations (P<0.05) with RNA expression were noted in the *MLANA* (rs2150702) locus, none of the three newly identified susceptibility alleles had statistically significant *cis* effects after correcting for the number of hypotheses tested in these data (Supplementary Table 3). This is not unusual as only 625 out of 1598 SNPs associated with a human trait were found to have a cis eQTL effect in a recent survey of validated loci.[37]

In an exploratory analysis, we implemented a pathway analysis approach by highlighting those pathways or networks in which the expression of multiple genes is influenced by a given SNP in a modest manner (P<0.05). Of the three tested SNPs, rs2150702 in the *MLANA* locus has a number of pathways that are involved in immune system function. Specifically the best-scoring pathway is related to CD28 signaling, but T and B cell receptor signaling and the IL-2 signaling appear to be involved as well (Supplementary Table 4). We subsequently evaluated the networks generated for each of the new MS susceptibility loci. In the highest scoring network of the *EOMES* locus (Supplementary Figure 4A), TNFα signaling appears to play a central role, connecting the locus to our previously described *TNFRSF1A* susceptibility one. The network also contains a decrease in *HLA-A* gene

expression relative to the susceptibility allele, which is in agreement with the validation of at least one protective allele (*HLA-A*02*) in this gene.[38] The highest scoring network for the *MLANA* locus (Supplementary Figure 4B) contains *ERK* and connections with type 1 interferon responses previously implicated in MS with the *IRF8* locus. Reducing the activity of the ERK signaling pathway has been reported to ameliorate experimental inflammatory encephalomyelitis (EAE), a murine model systems that captures many features of MS. [39] The best network of the *THADA* locus includes *STAT3* (Supplementary Figure 4A), one of the other genes with genome-wide significance in our meta-analysis (Table 2). The expression of *STAT3* RNA is enhanced relative to the risk allele rs6718520^{A} which is consistent with reports that diminished expression of *STAT3* in humans blocks the development of Th17 cells that play an important role in mediating inflammatory demyelination in MS.[38] Further, *STAT3* is part of a broader network that includes key regulators of cellular signaling such as *ERK*.

DISCUSSION

We have performed a meta-analysis of GWAS in multiple sclerosis and have highlighted three novel loci that reach genome-wide significance. The *EOMES* and *MLANA* loci showed no evidence of statistical heterogeneity, while the third locus near *THADA* showed some degree of heterogeneity among the strata of data. In addition, we point out several loci as having suggestive evidence of being associated with MS, such as the locus containing the *TAGAP* gene (previously associated with celiac disease, type 1 diabetes, and rheumatoid arthritis) and a locus on chromosome 17 that includes the transcription factor *TBX21* which plays an important role in the immunopathogenesis of murine models of MS.[32]

These results extend the list of loci associated with MS from earlier genome scans[2–5] and confirm theoretical predictions that increasing sample sizes will lead to additional discoveries given the magnitude of effect seen for non-MHC MS susceptibility loci. However, >10,000 subjects are needed to be fully powered to identify such common susceptibility alleles of modest effect.[41] One element of our strategy that enhances the likelihood for gene discovery is imputation of genotypes at SNPs not sampled by genotyping arrays using reference maps such as HapMap that catalogue the correlation structure among SNPs in human populations. Imputation allows the integrated analysis of datasets generated on different platforms and extends the analysis from roughly 750,000 genotyped SNPs at the end of quality control pipelines to the 2.5 million SNPs that we considered in this meta-analysis. Critically, confidence in an imputed genotype varies depending on how well it correlates with genotyped SNPs, and, if any uncertainty is present, it is incorporated into the test statistic, down-weighting the importance of the association results at that SNP. Imputation can be a very valuable tool: for example, when a genotyped SNP is an imperfect surrogate marker for a causal variant, the genotyped SNP may not display very strong evidence of association, and such a result may be lost in a genome-wide analysis. By estimating the results of the association test with a causal variant that is not genotyped, imputation can provide more robust evidence of association with a locus than a genotyped marker that is poorly correlated with the causal variant. Thus, imputed SNPs enhance our ability to detect and localize the effect of a new susceptibility locus. [41, 42] However, such results, as with any result from genome scans, needs to be independently replicated before being considered a validated locus.

Several limitations are present in our study. First, not all samples were genotyped with the same microarray platform, which can incorporate heterogeneity due to differences in coverage or genotype quality despite genome-wide imputation of >2 million SNPs on the HapMap reference dataset. Second, residual population substructure (even after excluding outliers and incorporating principal components into our analysis as covariates) may have

inflated the test statistic. When implementing a very conservative correction for the modest level of genomic inflation seen in each stratum, we find that the *EOMES* locus remains significant at a genome-wide level, but that the other two new loci fluctuate over this threshold of significance. Thus, the *EOMES* locus is our most robust result, and we look forward to efforts to validate the role of these loci that we describe in additional collections of MS subjects.

In conclusion, we report three new loci involved in the etiology of MS. Several additional loci are involved in inflammatory disease or immunologic function. Replication in large samples is now required to validate these loci as *bona fide* susceptibility loci in MS. Once validated, future fine-mapping studies across these loci are needed to provide a comprehensive picture of which variants have causal relationships with MS risk.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

P.L.D. is a Harry Weaver Neuroscience Scholar Award Recipient of the National MS Society (NMSS). This project was also supported in part by R01 NS067305 and RC2 NS070340. We thank the subjects with MS who have generously donated DNA samples for theses studies. We thank individuals with MS in Australia and New Zealand for supporting this research. We are grateful to J. Wright and C. Remediakis from Multiple Sclerosis Research Australia (MSRA) for expediting this research. This work was supported by MSRA, John T. Reid Charitable Trusts, Trish MS Research Foundation and the Australian Research Council, under the Linkage Projects Scheme (LP0776744). We thank the International MS Genetics Consortium for the use of genotype data. We thank the Myocardial Infarction Genetics Consortium (MIGen) study for the use of their genotype data as control data in our study. The MIGen study was funded by the U.S. National Institutes of Health and National Heart, Lung, and Blood Institute's STAMPEED genomics research program and a grant from the National Center for Research Resources. We acknowledge use of genotype 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.

Bibliography

1. Hauser SL, Oksenberg JR. The neurobiology of multiple sclerosis: genes, inflammation, and neurodegeneration. Neuron. 2006; 52:61–76. [PubMed: 17015227]
2. Hafler DA, Compston A, et al. International Multiple Sclerosis Genetics Consortium. Risk alleles for multiple sclerosis identified by a genomewide study. N Engl J Med. 2007; 357:851–62. [PubMed: 17660530]
3. De Jager PL, Jia X, Wang J, et al. Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. Nat Genet. 2009; 41:776–782. [PubMed: 19525953]
4. Sanna S, Pitzalis M, Zoledziewska M, et al. Variants within the immunoregulatory CBLB gene are associated with multiple sclerosis. Nat Genet. 2010; 42:495–7. [PubMed: 20453840]
5. The Australia and New Zealand Multiple Sclerosis Genetics Consortium (ANZgene). Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. Nat Genet. 2009; 41:824–828. [PubMed: 19525955]
6. McDonald WI, Compston A, Edan G, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. Ann Neurol. 2001; 50:121–127. [PubMed: 11456302]
7. De Jager PL, Chibnik LB, Cui J, et al. Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. Lancet Neurol. 2009; 8:1111–1119. [PubMed: 19879194]
8. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–909. [PubMed: 16862161]

9. Li Y, Abecasis GR. Rapid haplotype reconstruction and missing genotype inference. American Journal of Human Genetics. 2006; S79:2290.
10. de Bakker PI, Ferreira MA, Jia X, et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet. 2008; 17:R122–8. [PubMed: 18852200]
11. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999; 55:997–1004. [PubMed: 11315092]
12. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002; 21:1539–1558. [PubMed: 12111919]
13. Ioannidis JP, Patsopoulos NA, Evangelou E. Uncertainty in heterogeneity estimates in meta-analyses. BMJ. 2007; 335:914–916. [PubMed: 17974687]
14. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]
15. Voight BF, Scott LJ, Steinthorsdottir V, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010; 42:579–589. [PubMed: 20581827]
16. Lango Allen H, Estrada K, Lettre G, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. 2010; 467:832–8. [PubMed: 20881960]
17. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010; 466:707–713. [PubMed: 20686565]
18. Kathiresan S, Voight BF, Purcell S, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet. 2009; 41:334–341. [PubMed: 19198609]
19. Erdmann J, Grosshennig A, Braund PS, et al. New susceptibility locus for coronary artery disease on chromosome 3q22.3. Nat Genet. 2009; 41:280–282. [PubMed: 19198612]
20. Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. N Engl J Med. 2007; 357:443–453. [PubMed: 17634449]
21. Gauthier SA, Mandel M, Guttmann CR, et al. Predicting short-term disability in multiple sclerosis. Neurology. 2007; 68:2059–2065. [PubMed: 17562826]
22. De Jager PL, Baecher-Allan C, Maier LM, et al. The role of the CD58 locus in multiple sclerosis. Proc Natl Acad Sci U S A. 2009; 106:5264–5269. [PubMed: 19237575]
23. Maier LM, Lowe CE, Cooper J, et al. IL2RA genetic heterogeneity in multiple sclerosis and type 1 diabetes susceptibility and soluble interleukin-2 receptor production. PLoS Genet. 2009; 5:e1000322. [PubMed: 19119414]
24. Jakkula E, Leppa V, Sulonen AM, et al. Genome-wide association study in a high-risk isolate for multiple sclerosis reveals associated variants in STAT3 gene. Am J Hum Genet. 2010; 86:285–291. [PubMed: 20159113]
25. Aulchenko YS, Hoppenbrouwers IA, Ramagopalan SV, et al. Genetic variation in the KIF1B locus influences susceptibility to multiple sclerosis. Nat Genet. 2008; 40:1402–1403. [PubMed: 18997785]
26. International Multiple Sclerosis Genetics Consortium (IMSGC). Comprehensive follow-up of the first genome-wide association study of multiple sclerosis identifies KIF21B and TMEM39A as susceptibility loci. Hum Mol Genet. 2010; 19:953–962. [PubMed: 20007504]
27. Barrett JC, Hansoul S, Nicolae DL, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet. 2008; 40:955–962. [PubMed: 18587394]
28. Nair RP, Duffin KC, Helms C, et al. Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. Nat Genet. 2009; 41:199–204. [PubMed: 19169254]
29. Hunt KA, Zhernakova A, Turner G, et al. Newly identified genetic risk variants for celiac disease related to the immune response. Nat Genet. 2008; 40:395–402. [PubMed: 18311140]
30. Smyth DJ, Plagnol V, Walker NM, et al. Shared and distinct genetic variants in type 1 diabetes and celiac disease. N Engl J Med. 2008; 359:2767–2777. [PubMed: 19073967]
31. Lovett-Racke AE, Rocchini AE, Choy J, et al. Silencing T-bet defines a critical role in the differentiation of autoreactive T lymphocytes. Immunity. 2004; 21:719–731. [PubMed: 15539157]
32. Placek K, Coffre M, Maiella S, et al. Genetic and epigenetic networks controlling T helper 1 cell differentiation. Immunology. 2009; 127:155–162. [PubMed: 19476511]

33. Nurieva RI, Liu X, Dong C. Yin-Yang of costimulation: crucial controls of immune tolerance and function. Immunol Rev. 2009; 229:88–100. [PubMed: 19426216]
34. Stahl EA, Raychaudhuri S, Remmers EF, et al. Genome-wide association study meta- analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet. 2010; 42:508–14. [PubMed: 20453842]
35. Dubois PC, Trynka G, Franke L, et al. Multiple common variants for celiac disease influencing immune gene expression. Nat Genet. 2010; 42:295–302. [PubMed: 20190752]
36. Pereira TV, Patsopoulos NA, Salanti G, et al. Discovery properties of genome-wide association signals from cumulatively combined data sets. Am J Epidemiol. 2009; 170:1197–1206. [PubMed: 19808636]
37. Nicolae DL, Gamazon E, Zhang W, et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. 2010; 6:e1000888. [PubMed: 20369019]
38. Brynedal B, Duvefelt K, Jonasdottir G, et al. HLA-A confers an HLA-DRB1 independent influence on the risk of multiple sclerosis. PLoS One. 2007; 2:e664. [PubMed: 17653284]
39. Brereton CF, Sutton CE, Lalor SJ, et al. Inhibition of ERK MAPK suppresses IL-23- and IL-1-driven IL-17 production and attenuates autoimmune disease. J Immunol. 2009; 183:1715–1723. [PubMed: 19570828]
40. de Beaucoudrey L, Puel A, Filipe-Santos O, et al. Mutations in STAT3 and IL12RB1 impair the development of human IL-17-producing T cells. J Exp Med. 2008; 205:1543–1550. [PubMed: 18591412]
41. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet. 2010; 11:499–511. [PubMed: 20517342]
42. Li Y, Willer C, Sanna S, et al. Genotype imputation. Annu Rev Genomics Hum Genet. 2009; 10:387–406. [PubMed: 19715440]

Author Index

Nikolaos A. Patsopoulos[1–4],, Federica Esposito[4–6], Joachim Reischl[7,8], Stephan Lehr[7,8], David Bauer[7,8], Jürgen Heubach[7,8], Rupert Sandbrink[7,8,9], Christoph Pohl[7,8,10], Gilles Edan[11,12,13], Ludwig Kappos[11,12,14,15,16], David Miller[11,12,17], Javier Montalbán[11,12,18], Chris H. Polman[11,12,16,19], Mark S. Freedman[11,12,20], Hans-Peter Hartung[11,12,15,21,22], Barry G.W. Arnason[11,15,23], Giancarlo Comi[4-6,11,15], Stuart Cook[11,15,24], Massimo Filippi[4,11,15,22,25], Douglas S. Goodin[11,15,26,27], Douglas

[1]Program in Translational NeuroPsychiatric Genomics, Neurosciences Institute, Departments of Neurology & Psychiatry, Brigham & Women's Hospital, Boston, MA 02115, USA.
[4]International Multiple Sclerosis Genetics Consortium (IMSGC)
[6]Institute of Experimental Neurology, Scientific Institute San Raffaele, 20132 Milan, Italy
[7]The Bayer Schering Pharma Genomics (BSP-Gen) working group
[8]Bayer Schering Pharma AG, 13342, Berlin, Germany.
[9]Department of Neurology, Heinrich Heine University Düsseldorf, Germany
[10]Department of Neurology, University Hosptial Bonn, Germany
[11]steering committees of studies evaluating IFNb-1b and a CCR1-antagonist
[12]The BEtaferon®/BEtaseron® in Newly Emerging Multiple Sclerosis For Initial Treatment (BENEFIT) study
[13]Service Neurologie, C.H.U. Rennes Hôpital Pontchaillou/France.
[14]Neurologische Klinik und poliklinik, Universitaetsspital Basel/Switzerland.
[15]The Betaferon® Efficacy Yielding Outcomes of a New Dose (BEYOND) study
[16]The GeneMSA Consortium
[17]Institute of Neurology, The National Hospital, Queen Square, London/GB
[18]Unitat de Neuroimmunología Clínica Hospitals Vall dHebrón, Madrid/Spain
[19]Dept. Neurology Free University Hospital, Amsterdam/The Netherlands
[20]General Campus - Division of Neurology, The Ottawa Hospital, Ottawa/Canada
[21]Neurologische Klinik der Heinrich-Heine-Universität, Düsseldorf/Germany
[22]The CCR1-antagonist study
[23]Department of Neurology, Surgery Brain Research Institutes, Chicago/USA
[24]Department of Neurology and Neuroscience, UMDNJ-New Jersey Medical School, Newark, New Jersey/USA
[25]Neuroimaging Research Unit, University Hospital San Raffaele, Milano/Italy
[26]Department of Neurology, University of California San Francisco, San Francisco/USA

Jeffery[11,15,28], Paul O'Connor[11,15,29], George C. Ebers[11,27,30], Dawn Langdon[11,27,31], Anthony T. Reder[11,27,32], Anthony Traboulsee[11,27,33], Frauke Zipp[11,22,34], Jan Schimrigk[11,22,35], Jan Hillert[4,11,22,36], Melanie Bahlo[36,37], David R Booth[36,38], Simon Broadley[36,39,40], Matthew A Brown[36,41,42], Brian L Browning[36,43], Sharon R Browning[36,43], Helmut Butzkueven[36,44,45], William M Carroll[36,46,47], Caron Chapman[48], Simon J Foote[49], Lyn Griffiths[50], Allan G Kermode[36,46,47], ,Trevor J Kilpatrick[36,44,51,52], Jeanette Lechner-Scott[36,53,43], Mark Marriott[36,52], Deborah Mason[36,55], Pablo Moscato[36,54,56], Robert N Heard[36,38], Michael P Pender[36,57,58], Victoria M Perreau[36,51], Devindri Perera[36,49], Justin P Rubio[44], Rodney J Scott[36,53,54,56], Mark Slee[36,58], Jim Stankovich[36,44], Graeme J Stewart[36,38], Bruce V Taylor[36,44], Niall Tubridy[36,60], Ernest Willoughby[36,61], James Wiley[36,62], Paul Matthews[16,63], Filippo M. Boneschi[4-6], Alastair Compston[4,64], Jonathan Haines[4,65], Stephen L Hauser[4,66], Jacob McCauley[4,67], Adrian Ivinson[4,68], Jorge R Oksenberg[4,66], Margaret Pericak-Vance [4,67], Stephen J Sawcer[4,64], Philip L. De Jager[#*1–4], David A. Hafler[*3,4,68], Paul I.W. de Bakker[2–4,69,70]

[27]The 16 year long term follow-up of the pivotal study on IFNB-1b in RRMS
[28]Wake Forest University, Department of Neurology, Winston-Salem/USA
[29]Division of Neurology, St. Michael's Hospital, Toronto/Canada
[30]University Dept of Clinical Neurology, Oxford/UK
[31]Psychology Department, Royal Holloway, University of London/UK
[32]The University of Chicago, Department of Neurology, Chicago/USA
[33]UBC Hospital, Vancouver/Canada
[34]Cecilie-Vogt-Klinik, Charite – University Hospital Berlin, Germany
[35]Karolinska Institutet at Huddinge University Hospital, Huddinge/Sweden
[36]ANZGene Consortium
[37]The Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, AUS, 3050.
[38]The Westmead Millenium Institute, Westmead, NSW, AUS 2145.
[39]School of Medicine, Griffith University, QLD, AUS, 4222.
[40]Department of Neurology, Gold Coast Hospital, QLD, AUS 4215.
[41]Diamantina Institute of Cancer, Immunology and Metabolic Medicine, Princess Alexandra Hospital, University of Queensland, Brisbane, QLD, AUS 4102.
[42]Botnar Research Centre, Nuffield Department of Orthopaedic Surgery, University of Oxford, Oxford, OX3 7BN, UK.
[43]Department of Statistics, The University of Auckland, Auckland, NZ.
[44]The Howard Florey Institute, University of Melbourne, VIC, AUS 3010.
[45]Department of Medicine, University of Melbourne, VIC, AUS 3010.
[46]Sir Charles Gairdner Hospital, Nedlands, WA, AUS 6009.
[47]Australian Neuromuscular Research Institute, Nedlands WA, AUS 6009.
[48]Barwon Health, Geelong, VIC, AUS 3220.
[49]Menzies Research Institute, University of Tasmania, Hobart TAS 7001.
[50]Genomics Research Centre, Griffith University, QLD, AUS 4222.
[51]Centre for Neuroscience, University of Melbourne, VIC, AUS 3010.
[52]Royal Melbourne Hospital, Parkville, VIC, AUS 3050.
[53]John Hunter Hospital, Hunter New England Health Service, Newcastle NSW, AUS 2310.
[55]Canterbury District Health Board, Christchurch, NZ.
[54]Hunter Medical Research Institute, Newcastle, NSW, AUS 2308.
[56]Centre for Bioinformatics, Biomarker Discovery and Information-based Medicine, University of Newcastle, NSW, AUS 2308.
[57]School of Medicine, University of Queensland, QLD, AUS 4029.
[58]Department of Neurology, Royal Brisbane and Women's Hospital, QLD, AUS 4029.
[60]Department of Neurology, St. Vincent's University Hospital, Dublin, Republic of Ireland.
[61]Auckland District Health board, Auckland, NZ.
[62]Department of Medicine, Nepean Hospital, Penrith, NSW, AUS 2751.
[63]GSK Clinical Imaging Centre, Hammersmith Hospital and Department of Clinical Neurosciences, Imperial College, London.
[64]University of Cambridge, Department of Clinical Neuroscience, Addenbrooke's Hospital, Hills Road, Cambridge, CB2 2QQ, UK
[65]Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN 37232-0700, USA
[66]Department of Neurology, University of California San Francisco, San Francisco, CA, 94143, USA
[67]John P. Hussman Institute for Human Genomics, University of Miami, Miller School of Medicine, Miami, FL 33136, USA
[68]Department of Neurology, Yale University School of Medicine, New Haven, CT 06520, USA.
[#]Corresponding author: Philip L. De Jager, MD PhD, Program in NeuroPsychiatric Genomics, Department of Neurology, Brigham & Women's Hospital, 77 Avenue Louis Pasteur, NRB168, Boston, MA 02115, Tel: 617 525 4529, Fax: 617 525 5722, pdejager@rics.bwh.harvard.edu
[*]contributed equally with PIW de Bakker
[3]Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

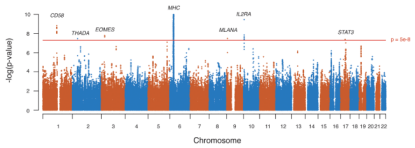

Figure 1. Manhattan plot for the meta-analysis genome-wide – log(p-values) (fixed effects)
X axis displays the 22 autosomal chromosomes and Y axis the –log(p-values) per SNP. The red line represents the genome-wide significance level (5×10^{-8})

[2]Division of Genetics, Department of Medicine, Brigham & Women's Hospital, Harvard Medical School, Boston, MA 02115, USA.
[69]Department of Medical Genetics, Division of Biomedical Genetics, University Medical Center, Utrecht, The Netherlands
[70]Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, The Netherlands

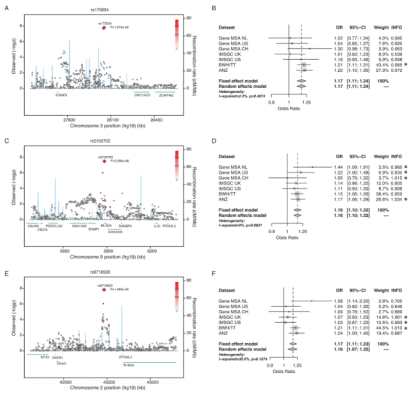

Figure 2. Regional association plots for the newly identified genome-wide significant loci and respective forest plots

(A, B) rs170934 in *EOMES,* **(C, D)** rs2150702 in *MLANA*, and **(E, F)** rs6718520 near *THADA*. Regional plots: The X axis plots 1 million basepairs around the most statistically significant (index) SNP, which is highlighted by a large red diamond. r^2 of a given SNP with the index SNP is illustrated with the intensity of the red color. The blue line represents the recombination rate. Each square represents one SNP. Forest Plots: The per-datasets' weights are from the fixed effects meta-analysis. The p-value is for the Cohran's Q test for statistical heterogeneity. INFO score is an imputation quality metric, corresponding to the ratio of observed vs. expected allele frequency. Values greater of 0.8 indicate high imputation quality. Genotyped SNPs have a value of 1. SNPs that were genotyped in a given dataset are marked with an asterisk.

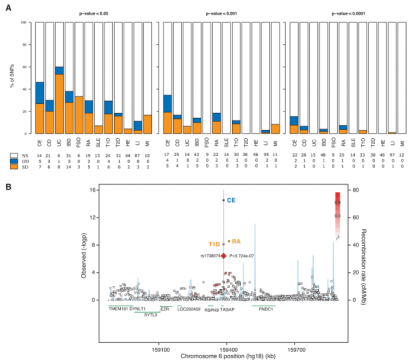

Figure 3. Overlap of the genetic architecture of MS with that of other inflammatory diseases

(A) Percentage of non-MHC genome-wide significant ($P<5\times10^{-8}$) SNPs of inflammatory diseases that are non-statistically significant (NS), or significant in the same direction (SD) or the opposite direction (OD) in the current MS meta-analysis. CE: celiac disease, CD: Crohn's disease, UC: ulcerative colitis, IBD: inflammatory bowel diseases (CD+UC), PSO: psoriasis, RA: rheumatoid arthritis, SLE: systemic lupus erythematosus, T1D: type 1 diabetes, T2D: type 2 diabetes, HE: height, LI: lipids, MI: myocardial infraction. NS: non-statistically significant, OD: opposite direction of effects, SD: same direction of effects. **(B)** Regional association plot for the *TAGAP* gene. All SNPs report –lop(p-values) from the MS meta-analysis, besides the 3 ones indicated by the respective disease names. The p-values reported for these 3 SNPs come from the respective original publications. The T1D and CE SNP is rs1738074, whereas the RA one is rs212389.

Table 1

Characteristics of the meta-analyzed datasets

Dataset	IMSGC UK*	IMSGC US*	Gene MSA CH*	Gene MSA NL*	Gene MSA US*	ANZ†	BWH/TT§
Number of individuals (cases/controls)	453/2950	342/1679	253/208	230/232	486/431	1618/1988	2313/4857
Clinical characteristics (cases only)							
Female:Male Ratio	3.0:1:0	3.2:1.0	2.8:1.0	2.9:1.0	3.1:1.0	2.6:1	2.6:1
Mean Disease Duration, years (range)	11 (0–40)	16 (<1–36)	12 (<1–58)	13 (<1–39)	15 (1–59)	NA	14 (2–53)/5(0–33)‡
Mean age at onset, years (range)	27 (10–48)	29 (11–50)	33 (8–59)	33 (13–71)	33 (1–70)	34 (7–67)	33 (9–60)/30 (8–54) ‡
Genotypic and analysis characteristics							
Genotyping platform	Affy 500K	Affy 500K	Illumina 550	Illumina 550	Illumina 550	Cases: Illumina Hap370CNV Controls: Illumina Infinium	Affy 6.0
Analyzed individuals (cases/controls)[a]	449/2928	341/1679	251/208	225/228	477/425	1616/1987	2186/4698
Genomic inflation factor (lambda)	1.029	1.034	1.04	1.026	1.029	1.061	1.054

*Analyzed in our previous meta-analysis [2]

†Analyzed in the ANZgene Study[4]

§Analyzed in part in our previous meta-analysis.[3] The dataset used in De Jager et al. has 860 subjects with MS from BWH and their 1720 matched healthy controls.

‡Values are for BWH cases and TT cases separately.

NA: not available

[a]Analyzed cases and controls per dataset after Quality Control.

Table 2

Meta-analysis results for SNPs with p-value $<10^{-6}$ and known genes/loci.

SNP	Position	Minor Allele	Major Allele	CHR	MAF[$]	Locus (±200Kbp)	OR	p-value	p-value for statistical heterogeneity	I^2 (95%CI)
Never reported as genome-wide significant										
rs170934	28054089	T	C	3	0.48	*EOMES*	1.17	1.57E-08	0.407	2% (0–72%)
rs2150702	5883861	G	A	9	0.49	*MLANA*	1.16	3.28E-08	0.682	0% (0–71%)
rs6718520	43179074	A	G	2	0.48	*THADA*	1.17	3.42E-08	0.157	35% (0–73%)
Published genes/loci										
rs10492972	10275699	C	T	1	0.32	*KIF1B*	1.04	0.243	0.605	0% (0–71%)
rs2300747	116905738	G	A	1	0.09	*CD58*	0.73	6.46E-09	0.715	0% (0–71%)
rs2760524	190797171	A	G	1	0.17	*RGS1*	0.88	3.00E-04	0.013	63% (16–84%)
rs12122721	199251103	A	G	1	0.28	*KIF21B*	0.91	2.43E-03	0.247	24% (0–66%)
rs9846534[a]	107064440	C	T	3	0.19	*CBLB*	0.97	0.434	0.327	13% (0–75%)
rs1132200	120633526	T	C	3	0.15	*TMEM39A*	0.90	3.67E-03	0.442	0% (0–58%)
rs4680534	161181639	C	T	3	0.37	*IL12A*	1.12	3.28E-05	0.957	0% (0–71%)
rs1393122	4778148	G	A	5	0.16	*Intergenic*	0.96	0.287	0.949	0% (0–71%)
rs6897932	35910332	T	C	5	0.26	*IL7RA*	0.89	2.26E-04	0.304	16% (0–60%)
rs9260489	30028311	T	G	6	0.45	*HLA-B*	1.21	1.16E-11	0.221	27% (0–68%)
rs3129889	32521523	G	A	6	0.20	*HLA-DRB1*	2.97	1.03E-206	0.506	0% (0–71%)
rs12722489 [b]	6142018	T	C	10	0.15	*IL2RA*	0.81	3.66E-08	0.818	0% (0–71%)
rs7089861 [b]	6150332	G	C	10	0.27	*IL2RA*	0.84	3.84E-08	0.113	42% (0–73%)
rs17824933	60517188	G	C	11	0.26	*CD6*	1.14	3.38E-05	4.70E-03	68% (29–86%)
rs1800693	6310270	C	T	12	0.42	*TNFRSF1A*	1.14	1.41E-05	0.349	10% (0–74%)
rs703842	56449006	G	A	12	0.31	*METTL1*	0.88	1.72E-05	0.877	0% (0–71%)
rs1790100	122222678	G	T	12	0.23	*MPHOSPH9*	1.11	6.61E-04	0.486	0% (0–71%)
rs17445836	84575164	A	G	16	0.22	*IRF8*	0.91	5.30E-03	0.211	28% (0–69%)
rs12708716	11087374	G	A	16	0.35	*CLEC16A*	0.90	1.08E-04	0.509	0% (0–71%)
rs744166[c]	37767727	G	A	17	0.43	*STAT3*	1.13	6.35E-06	0.327	13% (0–75%)
rs2293152[c]	37735055	G	C	17	0.38	*STAT3*	0.82	4.09E-08	0.486	0% (0–71%)

SNP	Position	Minor Allele	Major Allele	CHR	MAF[$]	Locus (±200Kbp)	OR	p-value	p-value for statistical heterogeneity	I^2 (95%CI)
rs763361	65682622	T	C	18	0.48	*CD226*	1.06	0.045	0.994	0% (0–71%)
rs6074022[d]	44173603	C	T	20	0.27	*CD40*	1.15	4.91E-06	0.170	34% (0–72%)
Suggestive (5×10^{-8} <p<1×10^{-6})										
rs2546890 [e]	158692478	G	A	5	0.48	*IL12B*	0.86	7.95E-08	0.509	0%
rs8070463	43123835	C	T	17	0.50	*KPNB1/TBKBP1/TBX21*	0.87	9.55E-08	0.586	0% (0–71%)
rs10411936	16409375	A	G	19	0.30	*EPS15L1*	1.16	2.04E-07	0.881	0% (0–71%)
rs2681424	123252212	T	C	3	0.40	*ILDR1/CD86*	1.16	2.33E-07	0.996	0% (0–71%)
rs7592330	68500287	G	A	2	0.44	*PLEK/FBXO48/C2orf13*	0.87	2.42E-07	0.983	0% (0–71%)
rs1738074	159385965	T	C	6	0.42	*TAGAP*	0.87	3.724E-07	0.238	25% (0–67%)
rs1250542	80704676	A	G	10	0.37	*ZMIZ1*	1.15	3.97E-07	0.370	8% (0–73%)
rs7191700	11314304	T	C	16	0.33	*TNP2/PRM3/PRM2/PRM1/C16orf75*	0.87	6.40E-07	0.705	0% (0–71%)
rs10866713[e]	158851472	A	G	5	0.22	*IL12B*	1.17	6.57E-07	0.658	0% (0–71%)
rs9596270	49740441	C	T	13	0.07	*Intergenic*	0.74	7.00E-07	0.347	11% (0–74%)

[$] Weighted minor allele frequency across all datasets.

[a] The reported SNP in the original publication is rs9657904. These two are in perfect LD (r^2=1).

[b] rs12722489 and rs7089861 are in week linkage disequilibrium (r^2=0.128).

[c] rs744166 was reported in the original publication. rs744166 and rs2293152 are in week LD (r^2=0. 0.128). These represent one effect.

[d] rs6074022 did not reached genome wide significance in the original publication (p=1.30E-07).

[e] rs2546890 and rs10866713 are in week LD (r^2= 0.01). These represent two independent effects.

Table 3

Meta-analysis results for previously published known ($p<5\times10^{-8}$) non-MHC SNPs in inflammatory diseases that are nominally (<0.05) significant in the current meta-analysis.

					Inflammatory diseases			Multiple Sclerosis						
									Fixed effects		Random effects			
SNP	Position	Allele	CHR	Gene/locus*	Disease	OR	p-value	Allele (other allele)	OR	p-value	OR	p-value	p-value for statistical heterogeneity	I^2 (95%CI)	FREQ
rs3748816	2516596	G	1	TNFRSF14, MMEL1	CE	0.89	3.28E-09	G	0.89	2.81E-05	0.89	2.81E-05	0.977	0% (0–71%)	0.33
rs10903122	25176153	A	1	RUNX3	CE	0.89	1.73E-10	A	1.08	3.27E-03	1.07	0.127	0.038	55% (0–81%)	0.50
rs2201841	67466780	A	1	IL23R	UC	1.16	1.30E-13	A	0.93	0.013	0.93	0.019	0.37	8% (0–73%)	0.69
-	-	G	-	-	PS	1.35	3E-08	G	1.07	0.013	1.07	0.019	0.37	8% (0–73%)	0.31
rs2816316	190803426	C	1	RGS1	CE	0.8	2.20E-17	C	0.89	9.03E-04	0.85	0.013	0.019	60% (9–83%)	0.17
rs296547	199158750	A	1	-	CE	0.89	4.11E-09	T	0.91	7.48E-04	0.91	0.011	0.135	39% (0–74%)	0.35
rs11584383	199202479	T	1	-	CD	1.18	1.43E-11	T	1.11	4.63E-04	1.11	6.12E-03	0.201	30% (0–70%)	0.78
rs13017599	61017825	A	2	REL	RA	1.21	2E-12	A	1.12	6.66E-05	1.13	4.69E-04	0.191	31% (0–71%)	0.37
rs13003464	61040323	G	2	REL	CE	1.15	3.71E-13	G	1.07	0.015	1.07	0.015	0.970	0% (0–71%)	0.39
-	-	-	-	-	UC	1.2	7.4E-09	G	1.07	0.015	1.07	0.015	0.970	0% (0–71%)	0.39
rs934734	65449080	G	2	SPRED2	RA	1.13	5.30E-10	G	1.10	4.5E-04	1.06	0.2776	3E-03	69% (33–86%)	0.49
rs17035378	68452449	G	2	PLEK	CE	0.88	7.79E-09	C (T)	1.13	3.52E-05	1.13	3.79E-04	0.266	21% (0–65%)	0.29
rs3828309	233845139	G	2	ATG16L1	CD	1.28	2.36E-32	G	1.06	0.018	1.06	0.018	0.981	0% (0–71%)	0.53
rs4957048	636432	C	5	CEP72, TPPP	UC	1.3	1.2E-09	G (A)	1.07	0.035	0.93	0.035	0.756	0% (0–71%)	0.79
rs4613763	40428475	C	5	PTGER4	CD	1.32	6.82E-27	C	1.18	1.09E-05	1.18	1.09E-05	0.901	0% (0–71%)	0.13
rs6859219	55474327	A	5	ANKRD55, IL6ST	RA	0.85	9.60E-12	A	0.90	0.014	0.90	0.057	0.119	41% (0–75%)	0.10
rs2082412	158650357	G	5	IL12B	PS	1.56	2E-28	G	1.08	0.018	1.08	0.018	0.467	0% (0–71%)	0.80
rs6887695	158755213	C	5	-	PS	0.7	4.08E-10	C	0.91	2.02E-03	0.91	2.02E-03	0.906	0% (0–71%)	0.21
rs10806425	90983323	A	6	BACH2	CE	1.13	3.89E-10	A	1.08	3.90E-03	1.08	3.90E-03	0.954	0% (0–71%)	0.40
rs11755527	91014942	G	6	BACH2	T1D	1.13	5E-12	G	1.08	2.69E-03	1.08	3.33E-03	0.404	2% (0–72%)	0.46
rs802734	128320481	G	6	PTPRK	CE	1.17	2.62E-14	G	0.90	2.01E-04	0.90	2.01E-04	0.883	0% (0–71%)	0.30
rs10499194	138044320	T	6	TNFAIP3	RA	0.75	1.00E-09	T	1.11	3.58E-04	1.06	0.358	2.3E-03	71% (36–87%)	0.29
rs1738074	159385955	A	6	TAGAP	CE	1.16	2.94E-15	T (C)	0.87	3.72E-07	0.87	1.69E-04	0.238	25% (0–67%)	0.42
-	-	T	-	-	T1D	0.92	7.59E-09	T (C)	0.87	3.72E-07	0.87	1.69E-04	0.238	25% (0–67%)	0.42
rs212389	159409769	G	6	TAGAP	RA	0.87	2.70E-09	G	0.90	2.17E-04	0.90	5.15E-03	0.229	26% (0–68%)	0.38
rs3093023	167454270	A	6	CCR6	RA	1.11	1.50E-11	A	0.94	0.021	0.94	0.021	0.903	0% (0–71%)	0.43
rs10758669	4971592	C	9	JAK2	CD	1.12	3.46E-09	C	1.11	4.21E-04	1.11	4.21E-04	0.546	0% (0–71%)	0.36

					Inflammatory diseases			Multiple Sclerosis						
								Fixed effects		Random effects				
SNP	Position	Allele	CHR	Gene/locus*	Disease	OR	p-value	Allele (other allele)	OR	p-value	OR	p-value	p-value for statistical heterogeneity	I^2 (95%CI)	FREQ
rs706778	6138945	T	10	IL2RA	RA	1.11	1.40E-11	T	1.09	1.78E-03	1.09	1.78E-03	0.586	0% (0–71%)	0.42
rs17582416	35327646	G	10	-	CD	1.16	1.79E-09	G	1.06	0.028	1.06	0.028	0.601	0% (0–71%)	0.35
rs10995271	64108482	C	10	ZNF365	CD	1.25	4.46E-20	C	0.94	0.019	0.93	0.065	0.101	43% (0–76%)	0.38
rs1250552	80728023	G	10	ZMIZ1	Celiac	0.89	9.09E-10	G	1.10	5.43E-04	1.10	5.43E-04	0.765	0% (0–71%)	0.46
rs1250550	80730313	T	10	ZMIZ1	IBD (early onset)	0.86	6E-09	A (C)	1.12	1.62E-04	1.12	1.62E-04	0.749	0% (0–71%)	0.34
rs4763879	9801421	A	12	CD69	T1D	1.09	1.9E-11	A	0.93	0.012	0.93	0.012	0.929	0% (0–71%)	0.35
rs653178	110492129	G	12	SH2B3	CE	1.2	7.15E-21	C (T)	1.10	3.82E-04	1.10	3.82E-04	0.809	0% (0–71%)	0.50
rs17696736	110971191	G	12	C12orf30,SH2B3,LNK,TRAFD1,PTPN1	T1D	1.22	2E-16	G	1.09	9.37E-04	1.09	9.37E-04	0.515	0% (0–71%)	0.45
rs4900384	97568694	G	14	-	T1D	1.09	3.7E-09	G	0.94	0.039	0.94	0.039	0.817	0% (0–71%)	0.29
rs3825932	77022491	C	15	CTSH	T1D	1.16	3E-15	C	1.08	0.011	1.08	0.011	0.525	0% (0–71%)	0.67
rs12708716	11087364	A	16	CLEC16A	T1D	1.23	3E-18	A	1.11	1.08E-04	1.11	1.08E-04	0.509	0% (0–71%)	0.65
rs12928822	11311384	A	16	SOCS1	CE	0.86	3.12E-08	T (C)	0.87	8.66E-05	0.87	8.66E-05	0.513	0% (0–71%)	0.16
rs11860650	31324197	T	16	ITGAM	SLE	1.43	1.90E-20	T	1.13	2.72E-03	1.13	2.72E-03	0.535	0% (0–71%)	0.13
rs1728785	67148721	G	16	CDH1	UC	1.17	3E-08	C (A)	1.08	0.021	1.08	0.021	0.893	0% (0–71%)	0.78
rs2872507	35294299	A	17	ORMDL3	CD	1.12	5.00E-09	A	1.10	5.53E-04	1.10	5.53E-04	0.464	0% (0–71%)	0.46
rs2305480	35315712	A	17	ORMLD3 region	UC	1.14	3.0E-08	A	1.10	5.91E-04	1.09	9.46E-04	0.404	3% (0–72%)	0.45
rs2290400	35319756	G	17	ORMDL3	T1D	0.87	5.5E-13	C (T)	1.10	5.54E-04	1.08	0.017	0.208	29% (0–69%)	0.49
rs744166	37767717	A	17	STAT3	CD	1.18	6.82E-12	A	0.89	6.35E-06	0.89	6.14E-05	0.327	13% (0–75%)	0.57
rs763361	65682612	A	18	CD226	T1D	1.16	1E-08	T (C)	1.06	0.045	1.06	0.045	0.994	0% (0–71%)	0.48
rs425105	51900311	A	19	-	T1D	0.86	2.7E-11	T (C)	1.09	0.025	1.08	0.137	0.150	36% (0–73%)	0.84
rs4810485	44181364	T	20	CD40	RA	0.85	8.20E-09	T	1.15	4.43E-06	1.16	2.47E-04	0.180	32% (0–71%)	0.27
rs2836878	39387394	A	21	PSMG1	UC	0.73	1.4E-08	A	0.90	1.47E-03	0.90	1.47E-03	0.973	0% (0–71%)	0.28
rs762421	44439979	G	21	ICOSLG	CD	1.13	1.41E-09	G	0.94	0.029	0.94	0.029	0.597	0% (0–71%)	0.39
rs5771069	48777597	G	22	IL17REL	UC	1.17	4E-08	G	1.08	0.025	1.08	0.025	0.507	0% (0–90%)	0.54

*Gene/locus as reported in the original publications.

CD: Crohn's disease, UC: ulcerative colitis, CE: celiac disease, T1D: type 1 diabetes, RA: rheumatoid arthritis, SLE: systemic lupus erythematosus, PS: psoriasis, and IBD: inflammatory bowel disease. All SNPs had data from 7 strata, except rs5771069 that had data from only 2.