DNA methylation as a mediator of HLA-DRB1*15:01 and a protective variant in multiple sclerosis

Jenny Link
Lara Kular
Yun Liu
Sabrina Ruhrmann
Galina Zheleznyakova

See next page for additional authors

Follow this and additional works at: https://poetcommons.whittier.edu/bio

Part of the Biology Commons
Authors
Jenny Link, Lara Kular, Yun Liu, Sabrina Ruhrmann, Galina Zheleznyakova, Francesco Marabita, David Gomez-Cabrero, Tojo James, Ewoud Ewing, Magdalena Lindén, Bartosz Górnikiewicz, Shahin Aeinehband, Pernilla Stridh, Till F.M. Andlauer, Christiane Gasperi, Heinz Wiendl, Frauke Zipp, Ralph Gold, Björn Tackenberg, Frank Weber, Bernhard Hemmer, Konstantin Strauch, Stephanie Heilmann-Heimbach, Rajesh Rawal1, Ulf Schminke, Carsten O. Schmidt, Tim Schmidt, Andre Franke, Matthias Laudes, Alexander T. Dilthey, Elisabeth G. Celius, Helle B. Søndergaard, Jesper Tegnér, Hanne F. Harbo, Annette B. Oturai, Sigurgeir Olafsson, Hannes P. Eggertsson, Bjarni V. Halldorsson, Haukur Hjaltason, Elias Olafsson, Ingileif Jonsdottir, Kari Stefansson, Tomas Olsson, Fredrik Piehl, Tomas J. Ekström, Ingrid Kockum, Andrew P. Feinberg, and Maja Jagodic
DNA methylation as a mediator of \textit{HLA-DRB1*15:01} and a protective variant in multiple sclerosis

Lara Kular et al.\textsuperscript{#}

The human leukocyte antigen (HLA) haplotype \textit{DRB1*15:01} is the major risk factor for multiple sclerosis (MS). Here, we find that \textit{DRB1*15:01} is hypomethylated and predominantly expressed in monocytes among carriers of \textit{DRB1*15:01}. A differentially methylated region (DMR) encompassing \textit{HLA-DRB1} exon 2 is particularly affected and displays methylation-sensitive regulatory properties in vitro. Causal inference and Mendelian randomization provide evidence that \textit{HLA} variants mediate risk for MS via changes in the \textit{HLA-DRB1} DMR that modify \textit{HLA-DRB1} expression. Meta-analysis of 14,259 cases and 171,347 controls confirms that these variants confer risk from \textit{DRB1*15:01} and also identifies a protective variant (\textit{rs9267649}, \textit{p} < 3.32 \times 10^{-8}, \text{odds ratio} = 0.86) after conditioning for all MS-associated variants in the region. \textit{rs9267649} is associated with increased DNA methylation at the \textit{HLA-DRB1} DMR and reduced expression of \textit{HLA-DRB1}, suggesting a modulation of the \textit{DRB1*15:01} effect. Our integrative approach provides insights into the molecular mechanisms of MS susceptibility and suggests putative therapeutic strategies targeting a methylation-mediated regulation of the major risk gene.

\textsuperscript{#}A full list of authors and their affiliations appears at the end of the paper. Correspondence and requests for materials should be addressed to A.P.F. (email: afeinberg@jhu.edu) or to M.J. (email: Maja.Jagodic@ki.se)
Multiple sclerosis (MS), a leading cause of neurological disability in young adults, is a chronic inflammatory disease of the central nervous system (CNS) characterized by autoimmune destruction of myelin and subsequent loss of neurons. Although the exact cause of MS remains unknown, inheritance of the disease is consistent with one locus exerting a moderate effect and many loci with modest effects. The first genetic risk factor was established more than 40 years ago in the human leukocyte antigen (HLA) locus, which encodes molecules involved in key immune functions. The HLA genes are among the most polymorphic genes and several alleles are often inherited together in extended haplotypes due to extremely high linkage disequilibrium (LD) in this part of the genome. The extended haplotype of the HLA class II region (DRB1*15:01-DRB5*01-01-DQA1*01:02-DQB1*06:02), which has been further refined to DRB5*01:01-DRB1*15:01-01, confers the strongest risk for developing MS. HLA class II loci encode molecules involved in presentation of peptide antigens to T cells by antigen presenting cells (APCs) and DRB1*15:01 confers a 3-fold increased risk of developing MS. With the advent of genome-wide association studies (GWAS) more than 100 additional non-HLA loci have been identified predisposing for MS with modest effects. However, the MS risk loci identified to date explain only about half of the disease heritability and little is known about the underlying causal variants and their molecular mechanisms.

Recent genetic and epigenetic fine-mapping efforts suggest that a vast majority of causal candidate variants for autoimmune diseases are non-coding and likely play a role in regulating gene expression. Epigenetic mechanisms can regulate gene expression by modification of DNA in a manner that is heritable through cell divisions. The most studied epigenetic mechanism is the covalent addition of a methyl group to cytosines in the context of CpG dinucleotides, because of a known stable mechanism for the propagation of the methylation by DNA (Cytosine-5)-Methyltransferase 1. Alterations in DNA methylation have been reported in blood, CD4⁺, and CD8⁺ T cells as well as in pathology-free brain regions from MS patients. Recently, genetic variants in the loci encoding epigenetic machinery genes have been associated with MS suggesting a role for epigenetic mechanisms in disease pathogenesis. However, while studies have investigated genetic and epigenetic mechanisms independently, little focus has been on how they may interact at a locus-specific level and jointly affect susceptibility to MS. Indeed, a growing body of evidence suggests that genetic and epigenetic modifications can interact biologically. This paradigm has been instrumental in deciphering the contribution of DNA methylation to the genetic risk that predisposes to other complex diseases. Several studies suggest that DNA methylation in the HLA class II region could mediate genetic susceptibility to immune-mediated diseases such as rheumatoid arthritis (RA), type 1 diabetes, and food allergy.

![Flow diagram](image-url)

**Fig. 1** Study design and workflow diagram. MS: multiple sclerosis, HC: healthy controls, SNP: single nucleotide polymorphism, CIT: causal inference test, eQTL: expression quantitative trait loci, NINDC: non-inflammatory neurological disease controls, SCAND: Scandinavia, SWE: Sweden, DE: Germany, ICE: Iceland.
In this study, we investigate DNA methylation in MS patients in the context of genetic variation and gene expression with the aim to decipher biological consequences of inheritance of MS risk alleles and demonstrate that DNA methylation mediates risk of developing MS. Specifically, DNA methylation in the HLA-DRB1 gene mediates the effect of the strongest MS risk variant HLA-DRB1*15:01, and of a protective HLA variant (rs9267649) which has not been previously reported, on HLA-DRB1 expression and the risk of MS. Our results are summarized in Fig. 1.

**Results**

**DRB1*15:01** is hypomethylated and predominantly expressed. We conducted DNA methylation analysis in monocytes sorted from MS patients and matched controls (n = 36, Fig. 1, cohort 1, Supplementary Data 1) using Illumina Infinium HumanMethylation450 BeadChip arrays (450K arrays). Monocytes are important precursors of tissue APCs and they have been implicated in MS through their ability to present myelin antigens, produce pro-inflammatory mediators, and phagocytose myelin. We identified two differentially methylated regions (DMRs) that are associated with MS after adjustment for confounders (FWER < 0.05) (Supplementary Data 2). Both DMRs mapped to the HLA-DRB1 gene and comprised 19 consecutive CpGs, encompassing exon 2, which were hypomethylated in MS patients (Fig. 2a). Since the DRB1*15:01 allele of the HLA-DRB1 gene confers the strongest risk for developing MS, we analyzed methylation differences between DRB1*15:01 carriers and non-carriers. Homozygous DRB1*15:01 carriers displayed significantly lower DNA methylation levels at HLA-DRB1 compared to heterozygous carriers and non-carriers (Fig. 2b, Supplementary Data 2).

![Diagram](image-url)
A potential bias in estimating DNA methylation with 450K arrays can arise from the impact of single nucleotide polymorphisms (SNPs). This is of particular importance in the HLA region due to its high density of polymorphic sites and the sequence similarity between proximal HLA genes. To rule out the possibility that low DNA methylation in DRB1*15:01 carriers resulted from SNPs that abolish CpG sites or affect hybridization of the probes, we used three different approaches (Fig. 2c–e). We replicated 450K array findings using locus-specific pyrosequencing of bisulfite converted (BS)-DNA for selected CpGs for which robust assays could be designed (Fig. 2c, n = 49). We further assessed DNA methylation by cloning and sequencing single-strand BS-DNA from a larger segment encompassing exon 2 of HLA-DRB1. Sequencing results from five homozygous DRB1*15:01 carriers confirmed that the CpG sites are not disrupted by local SNPs and that these CpGs preferentially exist in an unmethylated state (Fig. 2d, only one carrier shown) in the DRB1*15:01 carriers, compared to carriers of other haplotypes.
e.g., DRB1*04 and DRB1*08. We also addressed allele-specific methylation at cg0632479 using methyl-sensitivity restriction enzyme-qPCR (MSRE-qPCR) on genomic DNA in a subset of DRB1*15:01 heterozygous individuals (haplotype-specific primers were obtained from Alerup et al.26, see Methods and Supplementary Data 3). The DRB1*15:01 allele exhibited a lower cg0632479 methylation compared to other tested alleles ($p = 7.9 \times 10^{-3}$ and $p = 2.2 \times 10^{-3}$ compared to DRB1*03 and DRB1*04, respectively). To minimize any sequence-bias in amplification, we used primer sets that encompassed (Lucia) and promoter-free (SEAP) reporter gene vectors, respectively. Constructs were partially or fully methylated using HhaI and SssI in monocytes compared to non-carriers ($p = 9.7 \times 10^{-3}$ activity if the insert was fully methylated (ANOVA with Turkey’s multiple comparison test, Fig. 3f, Supplementary Fig. 1f). This indicates that DNA methylation is a regulatory feature of the region encompassing exon 2 of DRB1*15:01 and that the hypomethylated DRB1*15:01 has a capacity to drive higher gene expression. Although the size of the monocyte cohort was insufficient for formal methylation mediation analysis, our data suggest that hypomethylation and predominant expression of DRB1*15:01 could be a mechanism by which DRB1*15:01 confers risk of MS.

**Genome-wide analysis of methylation mediation.** Our findings in monocytes suggest that DNA methylation may be an intermediate of genetic risk in MS. Therefore, we set out to identify the epigenetic marks that may mediate the genetic risk for MS by integrating genome-wide genetic and epigenetic analysis, similar to the original use of the Causal Inference Test (CIT) method in our study on RA19 (Fig. 1). This method is robust to issues such as pleiotropy and reverse confounding that are likely to occur in complex diseases. Using whole blood samples from an independent genotyped (500 K) MS case-control cohort ($n = 279$, cohort 2, Supplementary Data 1), analyzed with 450K arrays, we applied the CIT method19 with genotype as a causal factor, DNA methylation as a potential mediator and MS as the phenotypic outcome (G, M, and Y, respectively, Fig. 4a). As regulatory methylation changes generally encompass multiple CpGs, and to minimize potential measurement errors, we sought to identify DMRs27. We first identified seven DMRs that associated with MS after adjustment for confounders (FWER < 0.05), all of which mapped to the HLA class II region (HLA-DRB1, -DRB3, -DRB4, -DQA1, -DQB1 genes, Table 1). The methylation levels within these seven DMRs were correlated in MS cases and healthy controls (Supplementary Fig. 2), forming genetically controlled methylation clusters known as GeMs17. We found that these DMRs were under the genetic control of 202 unique SNPs, with 875 significant SNP-DMR pairs (Bonferroni-adjusted $p < 0.05$), also known as methylation Quantitative Trait Loci (meQTLs). Out of the 202 SNPs, 52 were significantly associated with MS status (adjusted $p < 0.05$, max$T$ permutation), and all were located in the HLA region. We finally performed CIT analysis to identify the genetic variants that are independent of case/control status after adjusting for DNA methylation, suggesting mediation (Fig. 4a). Among the 52 SNPs, 50 were significant after causal analysis.
using CIT \((p < 0.05)\) including all seven DMRs (Table 1, Fig. 4b) and thus represent potential methylation-mediated relationships between SNPs and MS disease risk. The type of disease or treatment status at sampling had no effect on the methylation levels at any of the seven DMRs \((p > 0.3\) for all comparisons). Results of the analysis are given in Supplementary Data 4 and one example (chr6: 32552039-32552350; rs3135338) is illustrated in Fig. 4c. We further replicated the methylation differences between MS cases and controls for six out of the seven DMRs using sorted CD14\(^+\) monocytes \((n = 36,\) cohort 1), CD19\(^+\) B cells \((n = 29)\), as well as CD4\(^+\) \((n = 33)\) and CD8\(^+\) \((n = 29)\) T cells, with the least pronounced differences being observed in CD8\(^+\) T cells (Fig. 4c, Supplementary Data 5). Collectively, our data suggest a functional link between DNA methylation at the \(H LA\) class II locus and the risk of developing MS.

Remarkably, four of the seven DMRs reside within the \(H LA-DRB1\) gene and the two largest (DMR3 and DMR4) encompass the same CpGs that we have identified as associated with \(H LA-\)

---

**Figure a**

DMRs that associate with MS

- M \(\rightarrow\) Y

7 DMRs (22 CpGs)

**Figure b**

**Figure c**

DMR4 (chr6:32,552,039-32,552,350) / rs3135338

Methylation vs. phenotype

- Case
- Control

Methylation vs. phenotype

- Genotype:
  - AA
  - Aa
  - aa

Genotype vs. phenotype

- CIT

Y~\(-\)G~\(-\)G+\(\beta\)M

---
DRB1 expression in monocytes. We addressed a potential functional impact of the identified methylation-mediated SNPs on the expression of HLA-DRB1 using RNA-seq data from PBMCs58 (cohort 3, n = 156, Supplementary Data 1). We found that the risk alleles of all genotyped SNPs (n = 34) were associated with high HLA-DRB1 expression in PBMCs and the risk alleles for 28 SNPs were also associated with low methylation at DMR3 or DMR4 in blood (Supplementary Data 6). This suggested a causal link between the identified SNPs, DNA methylation at HLA-DRB1 exon 2 and HLA-DRB1 expression. To address this, we employed a two-sample Mendelian randomization (MR) using the SNPs previously identified as meQTLS for DMR3 or DMR4 in the blood cohort and eQTLs for HLA-DRB1 in the PBMC cohort. We utilized the Egger’s regression slope (βMR)29 to estimate the causal effect (see Methods). The MR-Egger analysis revealed a significant causal relationship between DNA methylation at DMR3 (βMR ± SE = −1.92 ± 0.33, p = 6.75 × 10−9) or DMR4 (βMR ± SE = −1.80 ± 0.30, p = 3.04 × 10−9) and HLA-DRB1 gene expression (Fig. 5 and Supplementary Data 6). We also verified that the inferred causal direction is correct using the MR Steiger test for directionality30. Namely, for the variant with the strongest meQTL and eQTL effect (rs3132946), DNA methylation was found to cause changes in HLA-DRB1 expression (p = 2.87 × 10−3 and p = 8.65 × 10−4 for DMR3 and DMR4, respectively, Supplementary Fig. 3).

These findings provide additional insight into the molecular mechanisms of variants in the HLA region that likely mediate the risk for developing MS through changes in DNA methylation and expression of HLA-DRB1.

Methylation mediation of DRB1*15:01 and a protective variant. Considering the complex structure of the HLA locus, with LD extending over large distances, and our findings of DRB1*15:01-specific hypomethylation in monocytes, we investigated whether the identified methylation-mediated SNPs confer risk of MS independently of the DRB1*15:01 haplotype (Fig. 1). We addressed this question in a large Scandinavian case-control cohort (n = 8172 cases and 13,263 controls, cohort 4, Supplementary Data 1) by testing association of the SNPs (45/50 with genotype data) with MS after adjusting for DRB1*15:01 (and associated terms) and all other established risk variants in the HLA locus3 (Fig. 6a, Supplementary Data 7). The majority of methylation-mediated SNPs (41/45) showed limited evidence of association with MS (p ≥ 1 × 10−5) after conditioning on DRB1*15:01 (Fig. 6a), suggesting that they confer risk from DRB1*15:01. However, three methylation-mediated SNPs displayed suggestive association with MS (p < 1 × 10−5) after adjusting for all known MS-associated variants in the HLA locus5 (Fig. 6a, Supplementary Data 7), suggesting potential novel variants. We followed up these associations by performing a meta-analysis using the Scandinavian cohort and three additional MS case-control cohorts (Supplementary Data 1) from Sweden (n = 1418 cases and 1058 controls), Germany (n = 3934 cases and 8455 controls)9, and Iceland (n = 735 cases and 148,571 controls), adjusting for all known MS-associated variants in the HLA locus5 (Supplementary Data 7). rs9267649 displayed a genomewide significant association with MS (p = 3.32 × 10−8, OR = 0.86) with a similar protective effect in all cohorts (Fig. 6b). Three additional methylation-mediated SNPs (rs2227956, rs2395182, and rs9271640) exhibited suggestive association with MS (p < 1 × 10−7) with rs2395182 and rs9271640 being in LD (r2 = 0.72, Supplementary Fig. 4). Among these suggestive SNPs, only rs2227956 (p = 7.08 × 10−8, OR = 0.86) was in high LD with rs9267649 (r2 = 0.96, Supplementary Fig. 4), thus representing the same association. Interestingly, the protective alleles at both SNPs were associated with higher blood methylation levels at DMR3 (chr6: 32551749-32551949) in exon 2 of the HLA-DRB1 gene and a lower HLA-DRB1 gene expression in PBMCs, particularly in DRB1*15:01 heterozygous individuals (Fig. 6c, d, Supplementary Fig. 5). This suggests that the protective variant could be counteracting the effect of the major DRB1*15:01 allele at the molecular level.

Collectively, our findings strongly suggest that DNA methylation mediates the risk of DRB1*15:01 on MS. Moreover, we identify an MS-associated variant which was not detectable by conventional genetic studies and protects against MS potentially via modulating DNA methylation at HLA-DRB1.

Discussion
Over the past decade, series of GWAS and custom-designed arrays have steadily added more regions to the list of MS-associated loci27–10. However, identification and interpretation of the causal variants remains difficult and their mechanisms are still largely unknown. Considering that the majority of susceptibility loci reside in regulatory regions of the genome and the important role of DNA methylation in gene regulation, we sought to investigate DNA methylation in MS patients in the context of the underlying genetic variation. Our data strongly suggest that DNA methylation in the HLA class II locus, especially encompassing exon 2 of the HLA-DRB1 gene, mediates the effect of DRB1*15:01 and of a protective HLA variant, which has not been previously reported, on HLA-DRB1 expression and the risk of MS. These findings provide new insights into the molecular mechanisms of MS susceptibility and suggest alternative therapeutic strategies based on modulating HLA-DRB1 levels.

Although the HLA region has been known as the strongest genetic risk factor for MS for over 40 years, the exact causal gene (s) and the mechanisms by which they affect MS susceptibility are still elusive. It is generally accepted that disease-associated variants in HLA-DRB1 primarily influence the structure of the peptide-binding groove encoded by exon 2. Altered amino acid residues of the HLA class II beta chain expressed on the APCs could lead to a changed T cell repertoire that causes autoimmune responses in the CNS31–34. However, specific antigens in MS have...
yet to be defined\textsuperscript{35}, which hinders the elucidation of the mechanisms underlying susceptibility conferred by the major risk haplotype DRB1*15:01, as well as development of antigen-specific therapies. On the other hand, it appears that the structural theory alone does not fully explain the association with MS, and recent studies have suggested that association of DRB1*15:01 with MS might also be related to gene expression\textsuperscript{28,36–38}. Consistent with this, we found that in monocytes of DRB1*15:01 carriers, HLA-DRB1 is unmethylated and expressed at a higher level compared to other haplotypes. We then utilized causal inference strategies to test the hypothesis that DNA methylation mediates the effect on HLA-DRB1 expression and, in turn, on MS susceptibility.

We first applied CIT in a case-control GWAS cohort and found statistical evidence of DNA methylation at HLA-DRB1 mediating the risk for MS from several SNPs located in the extended HLA locus. The majority of these SNPs conferred the effect from DRB1*15:01, as demonstrated by the lack of association after conditioning on DRB1*15:01. The impact of the DRB1*15:01 genotype on HLA-DRB1 methylation was similar in MS patients and healthy controls (Supplementary Fig. 5). This strong genetic effect can explain the observed methylation differences between MS patients and controls in all studied cell types, i.e., CD14\textsuperscript+ monocytes, CD19\textsuperscript+ B cells, CD4\textsuperscript+ T cells, and CD8\textsuperscript+ T cells. Similar to our findings, HLA-DRB1 methylation differences between MS patients and controls have been reported in CD4\textsuperscript+ T cells and suggested to be partially dependent on DRB1*15:01\textsuperscript{12}. Previous meQTLs studies have reported strong genetic regulation of methylation at CpGs shared by the DMRs identified in our study in white blood cells\textsuperscript{21}, pancreatic islets\textsuperscript{20}, and brain samples\textsuperscript{39}. Since these studies have not focused on DRB1*15:01, we have investigated the SNPs underlying the reported meQTLs and we found that many are in high LD with DRB1*15:01 and show the same direction of effect as in our study. It is thus not surprising that HLA-DRB1 methylation differences between cases and controls exist in multiple cell types, although the functional consequences of this genotype-driven methylation likely differ between distinct cell types and depend on the contribution of HLA class II molecules to a particular cell type-specific function.

Most of the CpGs identified by CIT mapped to the same DMRs in exon 2 of HLA-DRB1 found in monocytes, supporting a functional link between genetic variation, methylation, expression and risk of MS. Since DNA methylation changes can actively impact gene expression or be a consequence of transcriptional activity in the locus\textsuperscript{40}, we further addressed the potential causal effect of DNA methylation in exon 2 of HLA-DRB1 on expression both experimentally and analytically. Using an in vitro reporter system, we demonstrated that the exon 2 sequence of DRB1*15:01 exerts regulatory properties on gene expression in a DNA methylation-dependent manner. This suggests that hypomethylation of exon 2 could mediate the effect of DRB1*15:01 on HLA-DRB1 gene expression. This is also supported by the significant causal relationship between methylation at DMRs in exon 2 and HLA-DRB1 expression in PBMCs, obtained using two-sample MR. The directionality of the causal relationship was further confirmed using a method that combines MR with a Steiger test (MR-Steiger)\textsuperscript{30}. Altogether, our data strongly support that DRB1*15:01-dependent and DNA methylation-mediated levels of expression, together with the structural characteristics of the DRB1*15:01 molecule, contribute to MS risk. Indeed, both the quantity and the quality of the peptide-HLA complexes define immune responses\textsuperscript{41–43} and levels of the HLA genes in humanized HLA and TCR transgenic mice influence the severity of experimental MS-like disease\textsuperscript{44}. Our findings highlight the potential of using alternative or complementing strategies to antigen-specific therapies in MS, which would reduce the expression of HLA-DRB1. This strategy might be relevant beyond MS, given the associations between genetic variants and DNA methylation in the HLA class II region in several immune-mediated diseases\textsuperscript{19–21}.

The CpGs in exon 2 of HLA-DRB1 that mediate the effect on MS risk belong to an intragenic CpG island that shows enrichment for binding sites of the ‘architectural’ regulatory protein of the genome CCCTC-binding factor (CTCF) (Table 1). A DNA

### Table 1 DMRs that mediate genetic risk in multiple sclerosis

| Probe     | DMR | Location                      | Nber SNPs | FWER\textsuperscript{a} | Value\textsuperscript{a} | Gene         | Description | TF\textsuperscript{b}                          | Chromatin state\textsuperscript{b} |
|-----------|-----|-------------------------------|-----------|--------------------------|--------------------------|--------------|-------------|----------------------------------------------|-------------------------------------|
| cg26981746| DMR 1| chr6:32490012-32490043         | 47        | 0.013                    | 0.12                     | HLA-DRB5     | Intron 1/2  | EZH2, ZNF263                                | Weak transcribed                     |
| cg12015991| DMR 2| chr6:32549849-32549935         | 41        | 0                        | 0.21                     | HLA-DRB1     | Intron 2/3  | POLR2A                                     | Transcriptional elongation            |
| cg13910785| DMR 3| chr6:32562864-32551949         | 41        | 0                        | −0.14                    | HLA-DRB1     | Intron 2/3-Exon 2 | POLR2A, ZNF263, SIN3A, CTFC | Weak/poised promoter, weak/poised enhancer |
| cg23905789| DMR 4| chr6:32552039-32552350         | 42        | 0                        | −0.16                    | HLA-DRB1     | Exon 2-Intron 1/2 | POLR2A, ZNF263, SIN3A, CTFC | Weak/poised promoter, weak/poised enhancer |

DMR: differentially methylated region, chr: chromosome, Nber: number, SNP: single nucleotide polymorphism, FWER: family wise error rate, TSS: transcription starting site, TF: transcription factor

\textsuperscript{a}Methylation vs. phenotype

\textsuperscript{b}From ENCODE
methylation-dependent role of intergenic CTCF has previously been suggested in the gene-specific regulation of the HLA class II locus. Furthermore, an insulator function of intergenic CTCF has been reported when bound to intron 2 of DRB1*04. In line with this, non-canonical intergenic CTCF occupancy has been shown to regulate intragenic chromatin boundaries, ultimately affecting gene-specific transcription and alternative splicing. However, whether differential methylation at this locus affects binding of regulatory proteins and fine-tunes regulation of specific genes within the HLA locus during disease development is yet to be explored.

In addition to the methylation-mediated SNPs that conferred the effect from DRB1*15:01, we found potential associations with MS that were independent of any known MS variant in the HLA region. A previously unreported protective variant was discovered through significant and highly suggestive association of two SNPs in high LD ($r^2 = 0.96$), rs9267649 and rs2227956. Two additional suggestive SNPs in high LD with each other ($r^2 = 0.72$), rs2395182 and rs9271640, were associated with the risk of developing MS. The two latter SNPs do not necessarily represent a variant that is independent of the significant rs9267649 protective variant, as there is correlation between the SNPs ($r^2 = 0.23$ and $r^2 = 0.22$ for rs2395182 and rs9271640, respectively). In contrast to the potent hypomethylation of DRB1*15:01, the protective variant was associated with increased DNA methylation of DMR3 in exon 2 of HLA-DRB1 and with lower HLA-DRB1 expression in PBMCs. Due to the strong influence of DRB1*15:01 on methylation and expression levels and its correlation with rs9267649 ($r^2 = 0.40$), the significant functional impact of the protective variant could only be observed in individuals stratified by DRB1*15:01 genotype, and it was particularly evident in heterozygotes. This opposing effect suggests a putative interaction with DRB1*15:01 and its potential to alter, possibly antagonize, the effect of the major risk allele at the molecular level. Interactions within the HLA region have previously been suggested in immune-mediated diseases. The protective variant might influence HLA-DRB1 transcription through long-range interactions between regulatory regions, as previously suggested to occur within the HLA class II locus. However, given far-extending LD in the region, it is also plausible that the associated SNP tags a true causal variant elsewhere in the HLA locus, warranting future efforts to refine this association.

Our findings highlight the importance of integrating genetic and epigenetic data to explore molecular mechanisms underlying causal variants and to identify variants that might escape detection by conventional genetic studies. Given the robust genetic association of the HLA region with susceptibility to immune-mediated diseases, our findings, together with other studies,
suggest a role for DNA methylation in the pathogenesis of MS and autoimmune diseases in general. This in turn opens new avenues for development of therapeutic strategies aiming at controlling immune reactions by modulating HLA protein levels.

**Methods**

**Cohorts.** Description of the cohorts is shown in the Supplementary Data 1. Briefly, all samples used for DNA methylation and expression studies in MS, i.e., peripheral blood cells and sorted cells from blood, were collected in Sweden between 2005 and 2011. No formal sample size calculation was conducted, all available samples that passed quality control have been included in all analyses. Cohort 1 consisted of monocytes isolated from 62 MS patients and 20 healthy controls and was used for for pyrosequencing validation. All samples with available RNA of sufficient quality (n = 38) were used for qPCR-based gene expression analysis. All samples that were heterozygous and where DNA/RNA of a sufficient quantity and quality was available, were used for allele-specific methylation and expression analyses. Details of individuals used in specific analyses are provided in the Supplementary Data 1. Cohort 2 used for CIT analysis on peripheral blood cells included 140 MS patients with relapsing-remitting (RRMS, n = 121), primary progressive (PPMS, n = 4), or secondary progressive (SPMS, n = 15) disease, and 139 healthy controls. In total, 65% (91/140) of MS patients were treated at the time of sampling but the majority, 85% (77/91), received drugs that have a moderate impact on disease activity (e.g., 63 received interferon beta preparations and 12 glatiramer acetate). Cohort 3 used for expression QTL analysis comprised PBMCs from 156 patients, including 21 clinically isolated syndrome (CIS), 105 MS patients, and 30 non-inflammatory neurological disease controls (NINDC), such as neuropathy, paresthesia, sensory symptoms, vertigo, tension headache. Cohort 4 comprised four non-overlapping case-control cohorts used for conditional association analyses. The Scandinavian (SCAND) cohort consisted of 8,172 MS cases and 13,263 controls from the three Swedish national studies of MS: EIMS, GEMS, and IMSE. In addition, cases were included from a local biobank of MS cases as well as a cohort of Swedish blood donors and Swedish controls from the OLVIA.
SASBCA(6), and Twingem(28) studies. Danish cases and Norwegian cases and controls from in the DANJ TCCCG GWAs were also included in the SCAND (SweDan and SchWeizerische) cohort comprised of 3951 cases and 4181 controls (n = 1058) from Germany. The German (DE) cohort comprised MS cases (n = 3934) recruited from multiple sites in Germany(29) and matched controls (n = 8455) from several population-based cohorts across Germany namely KORA(30), HIRLIN(31), SHIP(32), DOGS(33), and FoCAS(34), as previously described(2). The Icelandic (IC) cohort comprised of 735 MS cases diagnosed between 1990 and 2005 and followed up at Landspitali, the National University Hospital of Iceland(35) and 148,571 population-based controls available through on-going projects at deCODE Genetics. The study was approved by the National Bioethics Committee of Iceland (VSN, 2010-0017) and conducted in agreement with the conditions issued by the Icelandic Data Protection Authority (DPA). All subjects who donated DNA samples signed informed consent. Personal identifiers of the patient data and biological samples were encrypted by a third party system approved and monitored by the DPA. A total of 333 healthy individuals with genotype and 450K array data have been obtained from the EIRA (the Epidemiological Investigation of Rheumatoid Arthritis) cohort from Sweden, GSE42861(36). All study participants had given their informed consent according to ethical board approval. Specifically, the study complies with ethical regulations and was approved by the Regional Ethical Board at Karolinska Institutet (Solna, Sweden), Copenhagen and Friedrichsberg, (Denmark), and Milan (Italy); the Ethikkommission der Stadt und des Landes Rheinland-Pfalz with the approval ID 837.019.10(7028)), and University of Munich, Munich), University of Münster, University Medical Center of Berg, University of Marburg, and University of Leipzig.

Sample preparation. PBMCs were obtained from 50 ml of blood and extracted using standard Ficol gradient procedures directly after collection. Monocytes were isolated using CD14(+) positive selection on MACs microbeads magnetic separation (Miltenyi), according to manufacturer’s instructions (>95% purity). Sorting of the CD4(+) and CD8(+) T cell and CD19(+) B cell populations was performed from the negative fraction obtained after sorting of monocytes by adding fluorochrome conjugated antibodies against human CD4 (clone SK3, APC-conjugated, Becton Dickinson), CD8 (clone SK1, FITC-conjugated, Becton Dickinson), CD19 (clone UCHT1, PE-conjugated, BD Bioscience), and CD14 (clone S125C1, APC-Cy7-conjugated, Becton Dickinson) using high speed MoFlo(®) cell sorter with >99% purity (Beckman Coulter, Inc). Extraction of genomic DNA and RNA from sorted cells was performed using Gen Elute Mammalian Genomic DNA Miniprep kit (Sigma-Aldrich) and RNeasy Mini kit (Qiagen), respectively. The amount and quality of DNA/RNA was accessed using Nanodrop ND-1000 Spectrophotometer (Nanodrop Technologies Inc). Processing of samples for 450K arrays was done at BEA, a company in Sweden and University Mannheim by the ethically approved procedure of the medical association of Rheinland-Pfalz with the approval ID 837.019.10(7028), Ruhr-Universität Bochum (reg. nr. 4319 12), University Medical Center Hamburg- Eppendorf (approved by the ethics committee of the Ärztekammer Hamburg), Charité—Universitätsmedizin Berlin, University of Rostock, University of Heidelberg, University of Marburg, and University of Leipzig.

Expression analyses of HLA-DRB1 in monocytes. The full list of primer sequences used in this study is shown in Supplementary Data 3. Expression of HLA-DRB1 was quantified using primers targeting different segments ( exon 1, exon 4–6, and exon 6) of the transcript. Allele-specific expression analysis of HLA-DRB1 was performed using primer sets shown in Table 1 and specific for each group of alleles targeting multiple haplotype-specific SNPs, adapted from Olerup et al(37) and confirmed using in silico alignment to IPD-IMGT/HLA database. All individual allele expressing monocytes were identified by applying a single SNP-based approach targeting g9270503 using two allele-specific forward primers specifically, partly methylated (for DRB1*) or non-methylated (for other HLA-DRB1 alleles (95–144), and one common reverse primer. Real-time PCR was performed on a BioRad CFX384/C1000 Real-Time Detection System with a three-step PCR protocol using SYBR green fluorophore: 95°C 30 s, followed by 40 cycles of 95°C 10 s, 60°C (65°C for the allele-specific SNP-based amplification); 30 s and 40°C 20 s. Quantification was not tested using the standard curve method, with amplification of target mRNA and endogenous control Actin mRNA. Assumption of equal variance between groups was tested using Brown-Forsythe test. Differences in HLA-DRB1 levels between DRB1*15:01 homozygotes, heterozygotes and non-carriers, and after 5-Aza-2‘-deoxycytidine treatment were tested using ANOVA (Tukey) multi comparative tests, respectively. DRB1*15:01 expression was compared to expression of the other allele in DRB1*15:01 heterozygotes using t-test. Correlation between HLA-DRB1 levels and methylation at HLA-DRB1-associated CpGs generated with 450K arrays was performed using the Spearman test. All statistical analyses were performed in GraphPad Prism 6 and 7 (GraphPad Software).

Methylation analyses of HLA-DRB1 in monocytes. For pyrosequencing and cloning-sequencing, 50 ng of genomic DNA from each sample was bisulfite-converted using an EZ DNA methylation Kit (ZYMO research). The kit of CpGs exhibiting robust pyrosequencing assays (cg09949906, cg11993350, cg20471943) were selected for pyrosequencing validation. Primers were designed using PyroMark Design software (Qiagen) (Supplementary Data 3). A first run of pre-PCR was necessary to amplify the region including cg11993350, cg20471943 ("PCR out", Supplementary Data 3). One microliter of BS- DNA (~10 ng) was applied as a template in the PCRs performed with the PyroMark PCR kit (Qiagen) using 5’-biotinylated reverse primers. The entire PCR product, 4 pmol of the respective sequencing primer, and streptavidin sepharose high-performance beads (GE Healthcare), were used for pyrosequencing on the PSQ 96 system and PyroMark Gold 96 reagent kit (Qiagen). The PyroMark CpG software 1.0.11 (Qiagen) served for data analysis. Assumption of equal variance between groups was tested using Brown-Forsythe test. Differences in methylation levels between DRB1*15:01 homozygotes, heterozygotes and non-carriers were tested using Kruskal-Wallis test with Dunn’s multiple comparison test in GraphPad Prism 6.

A larger fragment including exon 2 of HLA-DRB1 was selected to further validate methylation differences among groups using PyroMark Gold HSA reagent kit (Qiagen). As controls for this study, we selected a CpG island in this region that bisulfite conversion results in a fragmentation of the DNA, amplification of the final product (~675 bp comprising 43 CpGs, 16 of them being annotated in the 450K array) required a 3-step nested-PCR protocol using SupraTherm Taq (GeneCraft, Germany) and the primers listed in Supplementary Data 3. The last PCR primer was methylated and methylatable motifs to allow subsequent cloning. The first 2 PCRs (PCR_A, Supplementary Data 3) were conducted according to specific cycle parameters(38) and all PCR products were diluted 1:100 prior to next PCR. After gel excision and extraction of the target band using a gel extraction kit (Qiagen), the exon 2 of HLA-DRB1 was cloned into the pGEM-T Easy vector using DH5α cells (One Shot MAX Efficiency®). Positive clones were verified by PCR and by restriction enzyme digestion and the plasmids were sent for sequencing (Eurofins MWG Operon, Germany). Sequence alignment was performed with Vector NTI software (InforMax).

Individual CpG allele-specific DNA methylation assessment was performed using methyl-sensitive restriction enzyme-qPCR (MSE-qPCR). The CpG methylation-sensitive digestion of genomic DNA was carried out with the EpIJet DNA Methylation Analysis Kit based on MspI/HpaII digestion (ThermoScientific) using 30 ng of genomic DNA. Briefly, HpaII cuts only unmethylated CGCG motif and MspI cuts both unmethylated and methylated CGCG equally. The allele-specific methylation levels were quantified using the aforementioned allele-specific primers-qPCR and the 2–△Ct method and expressed as the ratio between HpaII-digested DNA (target) and input/non-digested DNA (reference) used for enzymatic reaction. DRB1*15:01 methylation was compared to methylation of the other allele DRB1*15:01:1 heterozygotes using the Mann–Whitney test in GraphPad Prism 6.

Investigation of the regulatory properties of the identified DMR was conducted using in vitro DNA methylation reporter assays. A fragment, including exon 2 of DRB1*15:01 and the regions upstream and downstream of exon 2 (1133 bp), was amplified using specific primers (Supplementary Data 3), containing Avell or Spy cI sites. PCR-amplified products were cloned into pCpG-free promoter vector (Inviogen) containing a Luciferase reporter gene and into a pCpG-free basic vector (Inviogen) containing a murine secreted embryonic alkaline phosphatase (mSEAP) reporter. In order to address the impact of DNA methylation on the regulatory properties of the identified region, the 5’-flanking region and the regions upstream and downstream of exon 2 (1133 bp), was amplified using specific primers (Supplementary Data 3), containing Avell or Spy cI sites. PCR-amplified products were cloned into pCpG-free promoter vector (Inviogen) containing a Luciferase reporter gene and into a pCpG-free basic vector (Inviogen) containing a murine secreted embryonic alkaline phosphatase (mSEAP) reporter gene. In order to address the impact of DNA methylation on the regulatory properties of the identified region, the 5’-flanking region and the regions upstream and downstream of exon 2 (1133 bp), was amplified using specific primers (Supplementary Data 3), containing Avell or Spy cI sites. PCR-amplified products were cloned into pCpG-free promoter vector (Inviogen) containing a Luciferase reporter gene and into a pCpG-free basic vector (Inviogen) containing a murine secreted embryonic alkaline phosphatase (mSEAP) reporter gene.

Supplementary Methods. HLA-DRB1 risk haplotypes were curated in Roswell Park Memorial Institute (RPMI) medium supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin, 2 mM L-glutamine, and 1 mM sodium pyruvate at 37°C in 5% CO2. Cells were stimulated with phorbol 12-myristate 13-acetate (PMA, Sigma) and exposed to different doses (0.5 μM, 5 μM) of 5-Aza-2‘-deoxycytidine (Sigma) for 3 days prior to harvesting and subsequent RNA/DNA extraction using AllPrep DNA/RNA Micro Kit (Qiagen).
SEAP and Renilla luciferase activity were measured after 48 h using QUANTI-Luc (Invivogen), the Phospha-Light System (Applied Biosystems) and the Dual-Glo Luciferase Assay System (Promega), respectively, according to manufacturer’s instructions, on the GloMax 96 Microplate Luminometer (Promega). Direct or inverted orientation of the sequence yielded similar results. Results are expressed as relative activity (Lucia or SEAP normalized against Renilla) and represent the mean value ± SEM. Gene expression in a representative experiment replicated at least 2–3 times. Assumption of equal variance between groups was tested using Brown-Forsythe test. Statistical analyses were performed using ANOVA with Turkey’s multiple comparison test in GraphPad Prism 7.

**Genome-wide DNA methylation analysis in monocytes.** We estimated the methylation level in sorted cells as the ratio between the methylated signal and the sum of methylated and unmethylated signals. We filtered probes that (i) contain known common SNPs outside HLA, (ii) are located on the X and Y chromosomes and (iii) with detection p-values larger than 0.01. The sorted cell types were confirmed using the estimateCellCounts function from the minfi package. To preprocess raw beta values before differential methylation analysis, we used the 3-step pipeline considered as optimal in Marabita et al.\(^7\): quantile normalization and Beta Mixture Quantile dilution\(^8\). Normalization with ComBat from the SVA package was used to correct for slide effects as identified using PCA\(^9\). Batch corrected data and Limma (Linear Models for Microarray Data)\(^10\) were used to compute differential methylation. The linear model included age and disease stage (RRMS and SPMS) as covariates. DMR analysis was carried out using the Bumphunter package and included the same model of age, sex, and disease stage (RRMS and SPMS) as covariates. All HLA-DRB1-associated CpGs were subsequently tested for association with the DRB1*15:01 genotype using additive, dominant and recessive models and correcting for sex, age, and status (HC, RRMS, or SPMS).

**Methylation mediation analysis in peripheral blood.** The methylation data from 450K arrays was preprocessed using the Illumina default procedure implemented in the Bioconductor minfi package\(^11\). The probe level raw data for each sample were normalized using the control probe scaling procedure and converted to methylation values on the 0–1 scale (\(M(M + U + 100\), where \(M\) and \(U\) represent the methylated and unmethylated signal intensities, respectively). For DMR analysis (collapsed CpGs analysis), we averaged measurements from CpG islands, shores, and shelves into one value for each sample using the ccpCollapse function in minfi package\(^11\). These methylation measurements were then used for genotype-dependent DMR identification and CIT analysis, as we did previously in the single CpG analyses\(^12\). All analyses were performed in R 2.14 and Bioconductor 2.9.

**Cell counts for the six major cell types in blood (Granulocytes, B cells, CD4\(^+\) T cells, CD8\(^+\) T cells, monocytes, and NK cells) for each individual were estimated using the estimateCellCounts function in minfi package\(^11\), which obtain sample-specific estimates of cell proportions based on reference information on cell-specific methylation signatures\(^7\).

To identify the DMRs associated with the MS phenotype, we used the bumphunter function in minfi package\(^11\) with adjustment for confounders: age, sex, self-reported smoking status (ever smokers vs. never smokers), diagnosis date, and the first two principal components of estimated differential cell counts. Regions, that have a family wise error rate (FWER) less than 0.05 with 1000 re-samplings and contain at least 2 probes, were identified as MS-associated DMRs. To evaluate the effect of DMR on type of disease and the presence of type 1 autoimmune diseases, the methylation measurements were then used for genotype-dependent DMR identification and CIT analysis, as we did previously in the single CpG analyses\(^12\). All analyses were performed in R 2.14 and Bioconductor 2.9.

**Expression analysis from RNA-seq in PBMCs.** RNA was extracted from PBMCs from individuals that had been genotyped on the ImmunoChip and diagnosed with either MS, CIS or NINDs. Library preparation and sequencing of samples that passed the criteria of sufficient amount, concentration, and quality (>3.0 of RNA. A total number of 17 samples were later discarded after quality control of sequencing data. A final number of 156 samples were included, after matching with genotypic data from the same individuals (cohort 3, Supplementary Data 1). cDNA libraries for sequencing were prepared using Illumina TruSeq kit (Illumina, San Diego, USA) and sequenced on an Illumina HiSeq 2000 machine. We generated data in fastq format using Illumina 1.8 quality scores. We obtained paired-end reads with a length of 100 bp from all 156 samples on an average sequence depth of 36 million reads per sample. The reads were mapped to the H. sapiens reference genome (NCBI 38.2 GRCh37) using STAR aligner. Conditional Quantile Normalization (CQN) method was used to normalize the count datasets and to account for the GC content bias. The residuals obtained after correcting for batch-effect and disease type using limFit and eBases from limma package\(^13\) on QCN values is used to correlate MS risk genotypes obtained from Illumina immunochip\(^4\) and Human660-Quad chip\(^5\), the effect of the genetic risk of MS using the CIT, we extracted the exposure summary statistics (effect size, standard error, sample size, p-value) from the linear regression analysis on cohort 2 (blood) using an additive minor allele dosage model. We obtained the outcome summary statistics from the HLA-DRB1 eQTLs considered in cohort 1 (PBMC), using an additive risk allele dosage model. The MR analysis was performed using TwoSampleMR\(^14\) and MendelianRandomization R libraries. We firstly clumped the meQTL SNPs by removing the SNPs in strong LD (\(R^2 > 0.8\) in 10,000 kb window) and then we retrieved the corresponding eQTL SNPs and associated summary statistics. We harmonized the effects of the SNP on the outcome and the exposure by ensuring that they refer to the same allele, correcting the strand for non-palindromic SNPs, and dropping all palindromic SNPs from the analysis. We then retrieved an LD correlation matrix from the MR-base database (−500 Europeans in 1000 genomes data)\(^15\) and performed MR Egger regression accounting for the correlation between variants. This method is valid under the assumption that the association of the genetic variants with the exposure is independent of the direct effect of the genetic variants on the outcome, and it allows us to obtain a consistent causal effect estimate as the slope from the Egger regression, as explained in Bowden et al.\(^16\). Noteworthy, there was no evidence of a directional pleiotropic effect (\(p = 0.37\) and 0.41 for DM3 and DM4, respectively) of the CIT (Cochran’s Q statistic, \(p = 0.34\) for DM3 and 0.84 for DM4, respectively). As the method considers that all the variants must be orientated such as the associations with the exposure have the same (positive) sign, we set all the associations with the exposures to be positive and the associations with the outcomes were re-oriented. The MR Steiger test for directionalality was considered for assessing the correct direction of the effect. The p value for the MRSteiger test (GisQTL and eQTL effect) is (3.132946). Briefly, this method performs the Steiger test to orient the direction of causality and explores a range of potential values of measurement error in the exposure and the outcome, to assess how reliable the inference of the causal direction is and gives a reliability ratio (\(R^2\)) for the variable. Supplementary Figure 3, means that it is \(R\) times more likely that the inferred direction of causality is correct (i.e., DNA methylation causes gene expression) compared to the opposite direction (gene expression causes methylation).

**Regression analyses.** We fit linear regression model with the methylation at each DMR, generated by CIT analysis in blood (cohort 2), as the outcome and the following predictors: SNP (rs9267649 or rs2227956), age, sex, self-reported smoking status (ever smokers vs. never smokers), batch, status (MS, HC), and the first two principle components of estimated differential cell counts in DRB1*15:01 homozygotes, heterozygotes, and non-carriers separately. Analysis was performed on 279 individuals from cohort 2 and additional 333 healthy individuals from the EIRA cohort. We fit linear regression model with the HLA-DRB1 expression value as the outcome and the following predictors: SNP: rs9267649 or rs2227956, batch and status (MS, CIS, NINDC) in DRB1*15:01 homozygotes, heterozygotes, and non-carriers separately. All analyses were done in Rcmdr.

**Conditional association analyses and meta-analysis.** Genotyping for the different samples in the SCAND cohort was carried out using different Illumina genotyping platforms and quality controls were used for each cohort separately. The genotyping and basic quality control for the Norwegian and Danish cohort (n=1362) as well as 648 of the Swedish MS cases and controls has been carried out by IMMSGC\(^12\). 1479 of the Swedish cases and controls were genotyped using the Illumina OmniExpress bead chip by deCODE. The TwinGene cohort (n=6748) was genotyped with Illumina Omni-Quad 2.5 bead chip. The *HLA* (HLA-DRB1*01) information was obtained from the Medical Genetics Department, St Olav’s University Hospital, and one only individual in each twin pair was included in subsequent analysis. Genotypes from ancestry informative markers from all the autosomes (n=3062) were combined and used for identifying related individuals (n=1335) using the genome command in PLINK v1.9 in a combined analysis of all the cohorts\(^4\). Population outliers identified using the SmartPCA program with standard settings (n=199) were removed\(^4\). Four principal component vectors were significant in the remaining individuals and were used for controlling population stratification.

Genotypes in the HLA region were used to impute HLA genotypes using HaplotypePhaser\(^4\). Since genotyping for all DNA methylation-mediated CIT SNPs were not available, we replaced the missing SNPs with SNPs in high LD in the 1,000 Genomes study in the European population [http://browser.1000genomes.org/index.html]. The association of CIT markers conditional on established HLA MS risk alleles\(^5\) and four principal component vectors was performed using the GLM command in R version 3.3.3. The established MS risk HLA model used contained the following terms: DRB1*15:01 presence, DRB1*15:01 homozygotes, DRB1*15:03 presence, DRB1*03:01 presence, DRB1*03:01 homozygotes, DRB1*08:01 presence, DRB1*15:01 presence, DRB1*15:01 homozygotes, DRB1*15:03 presence, DRB1*03:01 presence, DRB1*03:01 homozygotes, DRB1*08:01 presence, DRB1*15:01 presence, DRB1*15:01 homozygotes, DRB1*15:03 presence, DRB1*03:01 presence, DRB1*03:01 homozygotes, DRB1*08:01 presence,
DQB1*03:02 presence, DQA1*01:01 presence, DQB1*03:01 presence, A*02:01 presence, A*02:01 homozygotes, R*44:02 presence, B*38:01 presence, B*55:01 presence, two SNPs in the HLA region (rs2299092 and rs9277565, in both cases presence of minor allele) as well as interaction terms between presence of DRB1*15:01 and DQA1*01:01 presence and presence of DQB1*03:02 and DQB1*03:01. In analyses conditioned on DRB1*15:01, DRB1*15:01 presence, DRB1*15:01 homozygotes and interaction term between presence of DRB1*15:01 and DQA1*01:01 were included. The genotyping on SWE cohort was carried out using MS replication chip, a custom-made Illumina chip with 19,000 markers in the extended homozygotes and interaction term between presence of minor allele) as well as interaction terms between presence of DRB1*15:01 and DQA1*01:01 presence and presence of DQB1*03:02 and DQB1*03:01. In analyses conditioned on DRB1*15:01, DRB1*15:01 presence, DRB1*15:01 homozygotes and interaction term between presence of DRB1*15:01 and DQA1*01:01 were included.

The genotyping on SWE cohort was carried out using MS replication chip, a custom-made Illumina chip with 19,000 markers in the extended homozygotes and interaction term between presence of minor allele) as well as interaction terms between presence of DRB1*15:01 and DQA1*01:01 presence and presence of DQB1*03:02 and DQB1*03:01. In analyses conditioned on DRB1*15:01, DRB1*15:01 presence, DRB1*15:01 homozygotes and interaction term between presence of DRB1*15:01 and DQA1*01:01 were included. The genotyping on SWE cohort was carried out using MS replication chip, a custom-made Illumina chip with 19,000 markers in the extended homozygotes and interaction term between presence of minor allele) as well as interaction terms between presence of DRB1*15:01 and DQA1*01:01 presence and presence of DQB1*03:02 and DQB1*03:01. In analyses conditioned on DRB1*15:01, DRB1*15:01 presence, DRB1*15:01 homozygotes and interaction term between presence of DRB1*15:01 and DQA1*01:01 were included.
37. Irizar, H. et al. HLA-DRB1*15:01 and multiple sclerosis: a female association? *Mult. Scler.* 18, 569–577 (2012).

38. Aas, T. L. et al. Genome-wide differences of gene expression associated with HLA-DRB1 genotype in multiple sclerosis: a pilot study. *J. Neurommunol.* 257, 90–96 (2013).

39. Do, C. et al. Mechanisms and disease associations of haplotype-dependent allele-specific DNA methylation. *Am. J. Hum. Genet.* 93, 934–955 (2016).

40. Schubert, D. Function and information content of DNA methylation. *Nature 517*, 321–326 (2015).

41. Baumgart, M., Moos, V., Schubhauer, D. & Muller, B. Differential expression of major histocompatiblility complex II genes on murine macrophages associated with T cell cytokine profile and protective/suppressive effects. *Proc. Natl Acad. Sci. USA* 95, 6936–6940 (1998).

42. Hinterberger, M. M. Autonomous role of modulatory thymic epithelial cells in central CD4(+) T cell tolerance. *Nat. Immunol.* 11, S12–S19 (2010).

43. Gottschalk, R. A. et al. Distinct influences of peptide-MHC quality and quantity on in vivo T-cell responses. *Proc. Natl Acad. Sci. USA* 109, 881–886 (2012).

44. Eilmerich, S. et al. High incidence of spontaneous disease in an HLA-DR15 and TCR transgenic multiple sclerosis model. *J. Immunol.* 174, 1938–1946 (2005).

45. Majumder, P. & Boss, J. M. DNA methylation dysregulates and silences the HLA-DQA locus by altering chromatin architecture. *Genes Immun.* 12, 291–299 (2011).

46. Molineaux, S. et al. The insulator protein CTCF represses transcription on the HLA-DQ locus by altering chromatin architecture. *J. Immunol.* 189, 5745 (2012).

47. Arnold, R. et al. The insulator protein CTCF represses transcription on the HLA-DQ locus by altering chromatin architecture. *J. Immunol.* 189, 5745 (2012).

48. Gomes, N. P. & Espinosa, J. M. Gene-specific repression of the p53 target gene PIGS by CTCF-CCTC- Cohesion binding. *Genes Dev.* 24, 1022–1034 (2010).

49. Sheng, J. et al. Transcription of angiogenin and ribonuclease 4 is regulated by DNA polymerase III elements and a CCCTC-binding factor (CTCF)-dependent intragenic intron loop. *J. Biol. Chem.* 289, 12520–12534 (2014).

50. Marina, R. J. et al. TET-catalyzed oxidation of intragenic S-methylcytosine regulates CTCF-dependent alternative splicing. *EMBO J.* 35, 335–335 (2016).

51. Shukla, S. et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature 479*, 74–79 (2011).

52. Grange, L. et al. Filter-free exhaustive odds ratio-based genome-wide association analysis identifies a single nucleotide polymorphism associated with T cell cytokine profiles and is required for the regulation of long-distance chromatin interactions. *J. Exp. Med.* 205, 785–798 (2008).

53. Patterson, K., Molloy, L., Qu, W. & Clark, S. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 13, 8 (2012).

54. Millstein, J., Zhang, B., Zhu, J. & Schadt, E. Disentangling molecular relationships with a causal inference test. *BMC Genet.* 10, 23 (2009).

55. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 85, 97–108 (2009).

56. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909 (2006).

57. Diltay, A. et al. Multi-population classical HLA type imputation. *PLoS Comput. Biol.* 9, e1002877 (2013).

58. Zheng, X. et al. HIBAG—HLA genotype imputation with attribute tagging. *Pharmacogenomics* 14, 192–200 (2013).

59. Li, Y. et al. Genetic variation of ESR1 and its co-activator PPARGC1B is synergistic in augmenting the risk of estrogen receptor-positive breast cancer. *Breast Cancer Res.* 16, 317 (2014).

60. Volzke, H. et al. Cohort profile of the study of health in Pomerania. *Int. J. Epidemiol.* 40, 294–307 (2011).

61. Holle, R., Happich, M., Lowel, H., Wichmann, H. E. & Group, M. K. S. KORA Gesundheitsforschung Gesundheitsschutz (KORA): the study of health in Pomerania. *Int. J. Epidemiol.* 212, 816–823 (2003).

62. Scherr, M. et al. Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf Recall Study. *Study. Risk factors, evaluation of coronary calcium and lifestyle. Am. Heart J.* 164, 212–218 (2007).

63. Volzke, H. et al. Cohort profile of the study of health in Pomerania. *Int. J. Epidemiol.* 40, 294–307 (2011).

64. Berger, K. [DHS: The Dortmund health study]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 55, 816–821 (2012).

65. Muller, N. et al. IL-6 blockade by monoclonal antibodies inhibits angiogenin expression and lipoprotein (a) synthesis in humans. *J. Lipid Res.* 56, 1034–1042 (2015).

66. Patterson, K., Molloy, L., Qu, W. & Clark, S. DNA methylation array design: bisulfite modification and analysis. *J. Vis. Exp.* 56, e3170 (2011).
Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-04732-5.

Competing interests: Sigurgeir Olafsson, Hannes P. Eggertsson, Bjarni V. Halldorsson, Ingi Jonsdottir and Kari Stefansson are employees of deCODE genetics/Amgen Inc at the time work related to this study was carried out. H.F. Harbo has received honoraria for advice and lecturing from Biogen, Genzyme, Merck, Novartis; Sanofi Aventis and Teva. She has received modest unrestricted research grant for research from Novartis. B. Tackenberg received personal speaker honoraria and consultancy fees as a speaker and advisor from Bayer Healthcare, Biogen, CSL Behring, GRIFOLS, Merck Serono, Novartis, Octapharma, Roche, Sanofi Genzyme, TEVA and UCB Pharma. His University received unrestricted research grants from Biogen-idec, Novartis, TEVA, Bayer Healthcare, CSL Behring, GRIFOLS, Octapharma, SanofiGenzyme and UCB Pharma. B. Hemmer has served on scientific advisory boards for F. Hoffmann-LaRoche Ltd, Novartis, Bayer AG, and Genentech; he has served as DSMC member for AllergieCare and TGtherapeutics; he or his institution has received speaker honoraria from Biogen Idec, Teva NeuroScience, Merck Serono, Medimmune, Novartis, Desitin, and F. Hoffmann-La Roche Ltd; Hiis institution has received research support from Chugai Pharmaceuticals and Biogen; holds part of two patents; one for the detection of antibodies against KIR4.1 in a subpopulation of MS patients and one for genetic determinants of neutralizing antibodies to interferon β. The remaining authors declare no competing interests.

Lara Kular1, Yun Liu2,3, Sabrina Ruhrmann1, Galina Zheleznyakova1, Francesco Marabita1, David Gomez-Cabrerod, Tojo James1, Ewoud Ewing1, Magdalena Lindén1, Bartosz Górnikiewicz1, Shahin Aeinehband1, Pernilla Stridh1, Jenny Link1, Till F.M. Andlauerd, Christiane Gasperi6,7, Heinz Wiendl7,8, Frauke Zipp7,7, Ralf Gold10, Björn Tackenberg7,11, Frank Weber5,7, Bernhard Hemmer6,7,12, Konstantin Strauch13, Stefanie Heilmann-Heimbach14, Rajesh Rawal15, Ulf Schminke5, Carsten O. Schmidt17, Tim Kacprowski18, Andre Franke19, Matthias Laudes20, Alexander T. Dilthey21,22, Elisabeth G. Celius23,24, Helle B. Søndergaard25, Jesper Tegnér4,26, Hanne F. Harbo23,27, Annette B. Oturai25, Sigurgeir Olafsson28, Hannes P. Eggertsson28, Bjarni V. Halldorsson28,29, Haikur Hjaltason30,31, Elias Olafsson30,31, Ingileif Jonsdottir28,31, Kari Stefansson28,31, Tomas Olsson1, Fredrik Piehl1, Tomas J. Ekström1, Ingrid Kockum1, Andrew P. Feinberg2 & Maja Jagodic1

© The Author(s) 2018

1Department of Clinical Neuroscience, Center for Molecular Medicine, Karolinska Institutet, 171 77 Stockholm, Sweden. 2Center for Epigenetics, and Departments of Medicine, Biomedical Engineering and Mental Health, Johns Hopkins University, Baltimore, MD 21205, USA. 3Key Laboratory of Metabolism and Molecular Medicine, Ministry of Education; Department of Biochemistry and Molecular Biology, Fudan University Shanghai Medical College, 200032 Shanghai, China. 4Unit of Computational Medicine, Department of Medicine, Solna, Center for Molecular Medicine, Karolinska Institutet, 171 77 Stockholm, Sweden. 5Max Planck Institute of Psychiatry, 80804 Munich, Germany. 6Department of Neurology, Klinikum rechts der Isar, School of Medicine, Technische Universität München, 81675 Munich, Germany. 7German Competence Network Multiple Sclerosis (KKNMS), Klinikum Rechts der Isar, Technische Universität München, 81675 Munich, Germany. 8Department of Neurology with Institute of Translational Neurology, University of Münster, 48149 Münster, Germany. 9Department of Neurology, University Medicine Mainz, Johannes Gutenberg University Mainz, 55122 Mainz, Germany. 10Department of Neurology, St. Josef-Hospital, Ruhr-University Bochum, 44801 Bochum, Germany. 11Neuroimmunology Center, Marburg University, 35037 Marburg, Germany. 12Munich Cluster for Systems Neurology (SyNergy), 81377 Munich, Germany. 13Institute of Genetic Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany and Institute of Medical Informatics, Biometry, and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, 80539 Munich, Germany. 14Institute of Human Genetics, University Hospital Bonn and Division of Genomics, Life & Brain Research Centre, University of Bonn School of Medicine, 53113 Bonn, Germany. 15Institute of Epidemiology and Social Medicine, University of Münster, 48149 Münster, Germany. 16Department of Neurology, University Medicine Greifswald, 17489 Greifswald, Germany. 17Institute for Community Medicine, University Medicine Greifswald, 17489 Greifswald, Germany. 18Interfaculty Institute for Genetics and Functional Genomics, Ernst Moritz Arndt University Greifswald and University Medicine Greifswald, 17489 Greifswald, Germany. 19Institute of Clinical Molecular Biology, Kiel University, 24105 Kiel, Germany. 20Department I of Internal Medicine, Kiel University, 24105 Kiel, Germany. 21Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. 22Institute of Medical Microbiology and Hospital Hygiene Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany. 23Department of Neurology, Oslo University Hospital, 0372 Oslo, Norway. 24Institute of Health and Society, Faculty of Medicine, University of Oslo, 0450 Oslo, Norway. 25Danish Multiple Sclerosis Center, Department of Neurology, Rigshospitalet, University of Copenhagen, 2100 Copenhagen, Denmark. 26Biological and Environmental Sciences and Engineering Division, Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia. 27Institute of Clinical Medicine, University of Oslo, 0450 Oslo, Norway. 28deCODE genetics/Amgen Inc, 101 Reykjavik, Iceland. 29School of Science and Engineering, Reykjavik University, 101 Reykjavik, Iceland. 30Department of Neurology, Landspitali, The National University of Iceland, 101 Reykjavik, Iceland. 31Faculty of Medicine, School of Health Sciences, University of Iceland, 101 Reykjavik, Iceland. 32Department of Immunology, Landspitali, The National University Hospital of Iceland, 101 Reykjavik, Iceland. These authors contributed equally: Lara Kular, Yun Liu, Ingrid Kockum, Andrew P. Feinberg, Maja Jagodic.