International Journal of
Molecular Sciences

MDPI

Article

DNA Methylation Signatures of Multiple Sclerosis Occur Independently of Known Genetic Risk and Are Primarily Attributed to B Cells and Monocytes

Alexandre Xavier [1,†], Vicki E. Maltby [2,3,†], Ewoud Ewing [4], Maria Pia Campagna [5], Sean M. Burnard [1], Jesper N. Tegner [6,7,8,9], Mark Slee [10], Helmut Butzkueven [5,11], Ingrid Kockum [4], Lara Kular [4], Ausimmune/AusLong Investigators Group, Vilija G. Jokubaitis [5], Trevor Kilpatrick [12], Lars Alfredsson [4], Maja Jagodic [4], Anne-Louise Ponsonby [12,13], Bruce V. Taylor [14], Rodney J. Scott [1,15], Rodney A. Lea [2,16,‡] and Jeannette Lechner-Scott [2,3,*,‡]

1 School of Biomedical Sciences and Pharmacy, Hunter Medical Research Institute, University of Newcastle, New Lambton Heights, NSW 2305, Australia; alexandre.xavier@newcastle.edu.au (A.X.); sean.burnard@newcastle.edu.au (S.M.B.); rodney.scott@newcastle.edu.au (R.J.S.)
2 School of Medicine and Public Health, Hunter Medical Research Institute, University of Newcastle, New Lambton Heights, NSW 2305, Australia; vicki.e.maltby@newcastle.edu.au (V.E.M.); rodney.lea@qut.edu.au (R.A.L.)
3 Department of Neurology, John Hunter Hospital, New Lambton Heights, NSW 2305, Australia
4 Department of Clinical Neuroscience, Karolinska Institutet, Center for Molecular Medicine, Karolinska University Hospital, 17176 Stockholm, Sweden; ewoud.ewing@ki.se (E.E.); ingrid.kockum@ki.se (I.K.); lara.kular@ki.se (L.K.); lars.alfredsson@ki.se (L.A.); maja.jagodic@ki.se (M.J.)
5 Department of Neuroscience, Central Clinical School, Monash University, Melbourne, VIC 3004, Australia; maria.campagna@monash.edu (M.P.C.); helmut.butzkueven@monash.edu (H.B.); vilija.jokubaitis@monash.edu (V.G.J.)
6 Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia; jesper.tegner@kaust.edu.sa
7 Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia
8 Unit of Computational Medicine, Department of Medicine, Center for Molecular Medicine, Karolinska Institutet, Karolinska University Hospital, L8:05, 17176 Stockholm, Sweden
9 Science for Life Laboratory, Tomtebodavagen 23A, 17165 Solna, Sweden
10 College of Medicine and Public Health, Flinders University, Bedford Park, SA 5042, Australia; mark.slee@flinders.edu.au
11 MSBase Foundation, Melbourne, VIC 3004, Australia
12 Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Melbourne, VIC 3052, Australia; tkilpat@unimelb.edu.au (T.K.); annelouise.ponsonby@florey.edu.au (A.-L.P.)
13 National Centre for Epidemiology and Public Health, Australian National University, Canberra, ACT 2601, Australia
14 Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS 7000, Australia; bruce.taylor@utas.edu.au
15 Department of Molecular Genetics, Pathology North, John Hunter Hospital, New Lambton Heights, NSW 2305, Australia
16 Centre for Genomics and Personalised Health, School of Biomedical Science, Queensland University of Technology, Kelvin Grove, QLD 4059, Australia
* Correspondence: jeannette.lechnerscott@health.nsw.gov.au; Tel.: +61-02-4921-3540
† These authors share first authorship.
‡ These authors share last authorship.

check for updates

Citation: Xavier, A.; Maltby, V.E.; Ewing, E.; Campagna, M.P.; Burnard, S.M.; Tegner, J.N.; Slee, M.; Butzkueven, H.; Kockum, I.; Kular, L.; et al. DNA Methylation Signatures of Multiple Sclerosis Occur Independently of Known Genetic Risk and Are Primarily Attributed to B Cells and Monocytes. *Int. J. Mol. Sci.* **2023**, *24*, 12576. https://doi.org/10.3390/ijms241612576

Academic Editors: Valeria Gasperi and Shinichi Fukushige

Received: 22 May 2023
Revised: 20 July 2023
Accepted: 2 August 2023
Published: 8 August 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract: Epigenetic mechanisms can regulate how DNA is expressed independently of sequence and are known to be associated with various diseases. Among those epigenetic mechanisms, DNA methylation (DNAm) is influenced by genotype and the environment, making it an important molecular interface for studying disease etiology and progression. In this study, we examined the whole blood DNA methylation profiles of a large group of people with (pw) multiple sclerosis (MS) compared to those of controls. We reveal that methylation differences in pwMS occur independently of known genetic risk loci and show that they more strongly differentiate disease (AUC = 0.85, 95% CI 0.82–0.89, $p = 1.22 \times 10^{-29}$) than known genetic risk loci (AUC = 0.72, 95% CI: 0.66–0.76,

$p = 9.07 \times 10^{-17}$). We also show that methylation differences in MS occur predominantly in B cells and monocytes and indicate the involvement of cell-specific biological pathways. Overall, this study comprehensively characterizes the immune cell-specific epigenetic architecture of MS.

Keywords: multiple sclerosis; epigenetics; methylation; epigenome-wide association studies; genetic risk; cell deconvolution

1. Introduction

Multiple sclerosis (MS) is a chronic autoimmune and neurodegenerative disease of the central nervous system (CNS). It is caused by the combination of a complex genetic predisposition and environmental and lifestyle factors. Genome-wide association studies (GWAS) have identified over 200 single nucleotide polymorphisms (SNPs) associated with MS risk, the majority of which localize to genes relevant to immune function [1,2]. The strongest risk locus is found within the major histocompatibility complex (MHC) region of chromosome 6—the well characterized *HLA-DRB1*1501* MS risk allele [3–5].

Despite the progress made by GWAS, the exact pathogenesis of MS remains to be elucidated. There is good evidence for dysregulation of immune cell balance in the CNS, particularly with regards to immune phenotypes [6,7]. Studies suggest that altered ratios between adaptive and innate immune system cells lead to disease activity [6,7]. There are also a number of well-established environmental and lifestyle factors that contribute to disease risk, such as viral infection (specifically Epstein Barr Virus (EBV) and infectious mononucleosis [8–10]) and smoking [11], all of which interact with the genome. Other factors that have no direct genomic influence but are considered to contribute to disease include low ultraviolet radiation exposure, low vitamin D levels [12], and nulliparity [13]. The exact molecular mechanisms by which MS develops remain elusive, but epigenetic change is emerging as a contender that significantly influences disease risk.

One of the most well-studied epigenetic modifications is DNA methylation (DNAm), where increased levels of CpG methylation are generally associated with gene silencing and demethylation with gene activation. Epigenetic modulation via CpG methylation is unique in reflecting environmental exposure and lifestyle factors, making it a plausible modifier of MS onset and progression. Epigenome-wide association studies (EWAS) of DNA have identified changes in DNA methylation profiles associated with MS [14–16]. Although these studies have reported somewhat conflicting results, they all consistently identified differential methylation at *HLA-DRB1*1501*, which appears to partly mediate the relationship between the risk allele at this locus and MS [16]. The limited data from isolated cell type analyses suggest that each immune cell population has a unique epigenetic profile [17–20]. To date, isolated cell type EWAS studies have involved modest sample sizes, whereas larger whole blood EWAS studies have likely missed cell type-specific differential methylation due to masking of the signals from mixed cell populations. Statistical methods now exist to estimate the relative proportions of major blood cell types using methylation data derived from mixed cell DNA samples (reviewed by Campagna et al. [21]). Methods to identify cell-specific differential methylation (csDNAm) are derived from these algorithms (CellDMC).

In this study, we have conducted an EWAS of whole blood DNA from MS cases and controls and applied statistical deconvolution to reveal cell-specific methylation signatures associated with the disease. Importantly, the discovery group (from the Ausimmune study) included cases with early MS (principally first clinical diagnosis of CNS demyelination) who had a follow-up assessment of disease after 10 years [22]. This allowed us to identify causative epigenetic changes in specific cell subsets that occurred prior to, or early in, disease onset rather than as a consequence of long-standing pathology or therapeutic intervention. Validation analysis confirmed our results in two independent studies.

2. Results

The study design is described in Figure 1 and in the methods below.

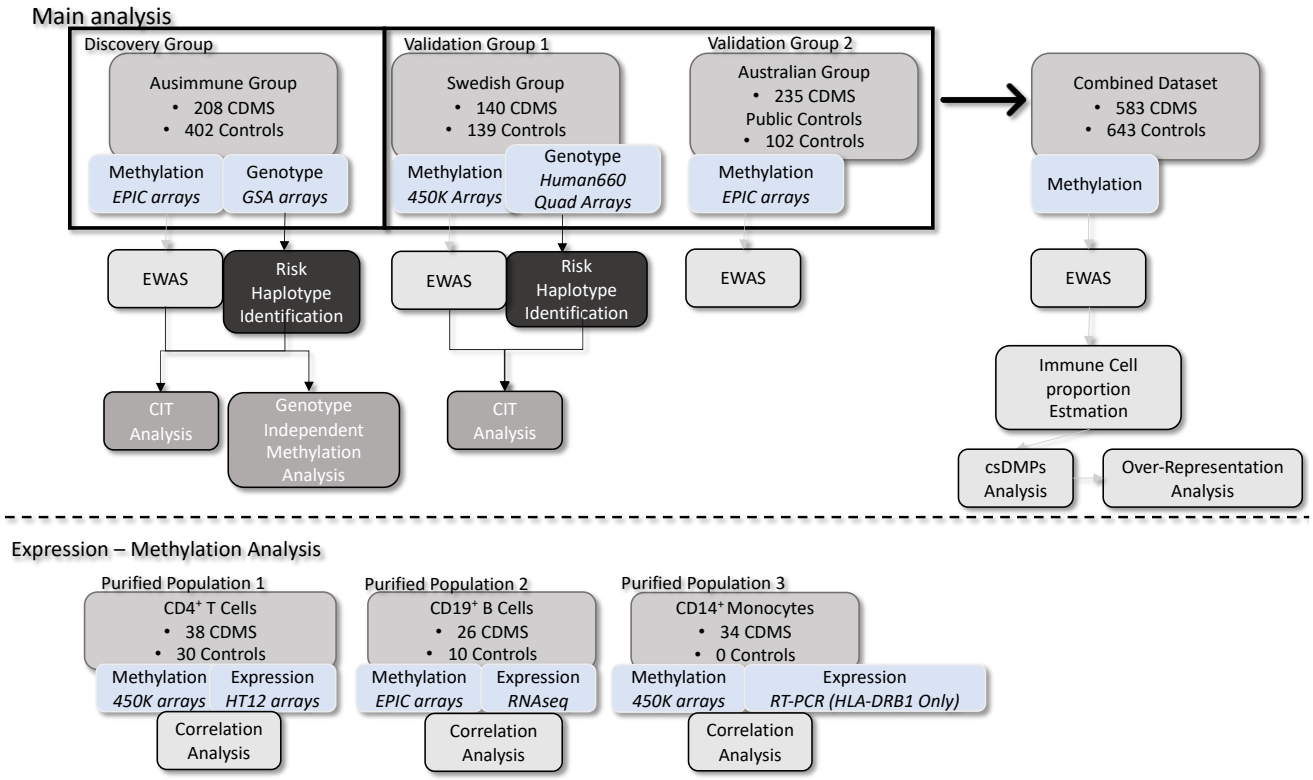


Figure 1. Study design and workflow.

2.1. DNA Methylation Analysis Implicates the HLA Locus

The strongest methylation differences were identified within the major histocompatibility complex (MHC) class II locus, within HLA genes. This replicated the findings of previous methylation studies in MS and highlighted again the importance of the MHC locus in MS. We first identified 13 replicated differentially methylated positions (DMPs) (see Supplementary Table S2), with 4 DMPs closely linked to interferon-related pathways (see Supplementary Table S3).

The combined dataset analysis identified 11,969 individual DMPs that met the genome-wide significance threshold of $p < 9.8 \times 10^{-8}$. Of these, 190 also had $\Delta\beta > 0.02$ and 81 were located at the MHC class II locus (see Figure 2a). Differentially methylated region (DMR) analysis identified 9 major DMRs across HLA-D genes (see Supplementary Table S4).

We then performed mediation analysis using the causal inference test (CIT) method [23] to test for causality between the *HLA-DRB1* risk haplotype (casual factor), DMR methylation (mediator), and clinical definite MS (CDMS) as the phenotype. CIT analysis allowed the establishment of a chain of causality between genotype, methylation, and MS. The results of the CIT analysis showed that the methylation differences identified between pwMS and controls were driven entirely by genotype in both the discovery and first validation groups ($p < 0.05$) (see Figure 2b) (the second validation group did not have genotyping information).

To determine if methylation was correlated with gene expression, we examined the relationship between methylation and gene expression using RNA extracted from isolated monocytes, $CD19^+$ B cells, and $CD4^+$ T cells from our previously published study [16], one unpublished study, and an unpublished dataset (Maltby et al., unpublished) (see methods for group details). We used Pearson's correlation to estimate the relationship between methylation and expression levels of *HLA-DRB1* in all three cell types (see Figure 2c) and found a strong negative correlation between gene expression and methylation at DMR-2

located in exon 2 of *HLA-DRB1*. These data indicated that hypomethylation at this locus was associated with increased gene expression in these immune cell subtypes. As expected, based on the DNA methylation levels, there was a strong correlation between *HLA-DRB1* expression and methylation levels, but this correlation was not as strong in $CD4^+$ T cells.

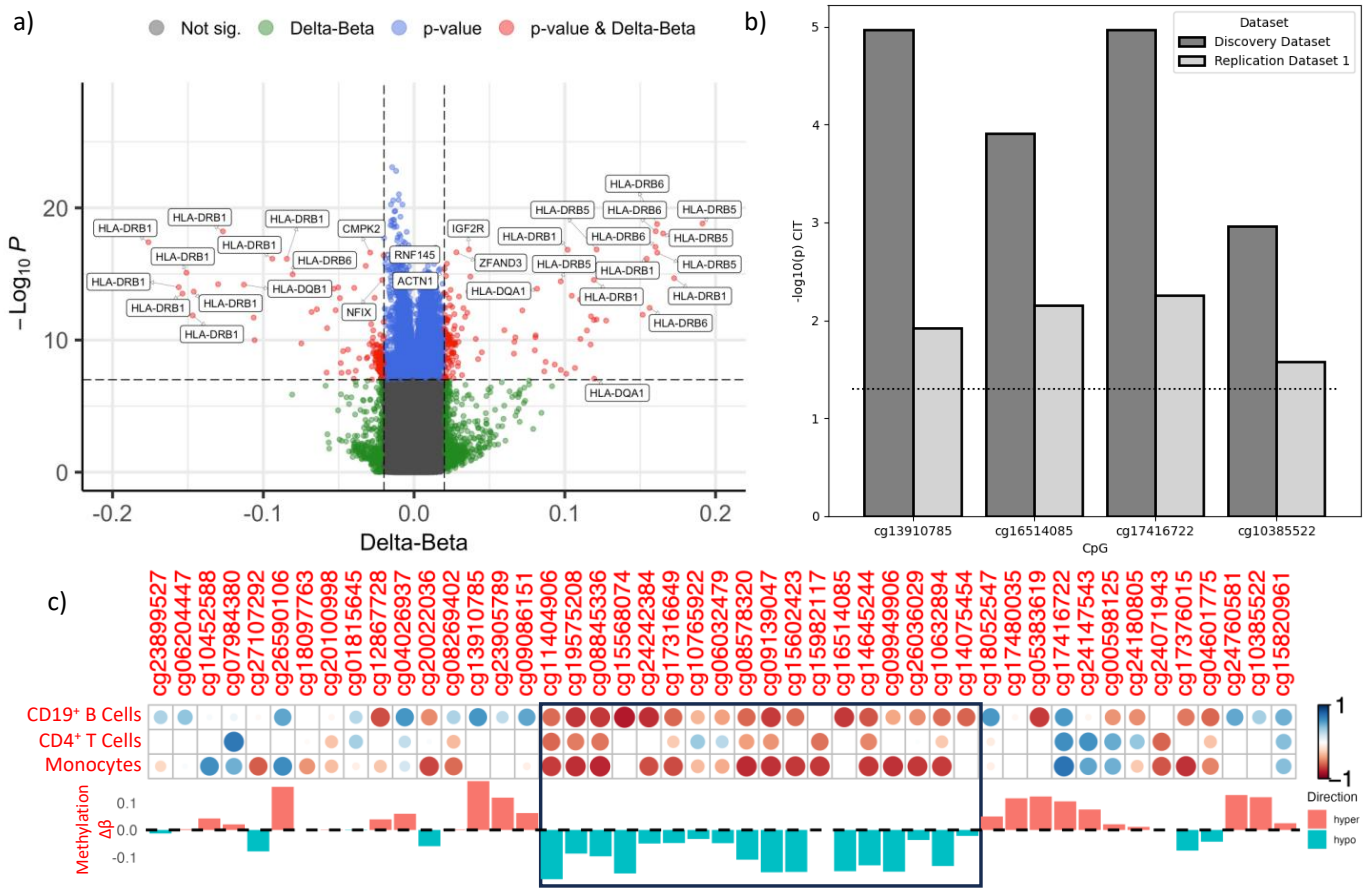


Figure 2. Methylation differences at the MHC class II locus are the largest between controls and pwMS. (**a**) Combined analysis of all three cohorts identified DMPs associated with MS. Volcano plot representing the 452,453 CpGs analyzed in the combined cohort DMP analysis. Red dots indicate DMPs that passed both the adjusted *p*-value ($-\log10(p) \leq 7$) and delta beta (± 0.02) thresholds. Blue dots represent DMPs that passed the *p*-value threshold only and green dots indicate DMPs that passed the delta beta threshold only. The locus to which each DMP is mapped is indicated for genes that passed both the *p*-value and delta beta thresholds. (**b**) Genotype at *HLA-DRB1* mediates methylation. Results of the CIT analysis measuring haplotype-related mediation of MS risk through methylation. p CIT: overall Omnibus CIT *p*-values. Dotted horizontal line represents significance threshold of $p \leq 0.05$. The CpGs used for *HLA-DRB1* are indexed CpGs for each of the 4 DMRs associated with MS. Dark gray = discovery dataset, light gray = validation dataset 1. (**c**) Correlation between *HLA-DRB1* expression and CpG methylation. Top panel: cell types are listed in rows, CpGs are in columns, except for column 1, which represents gene-of-interest expression in each cell type. The correlation coefficient (R) is represented through circle size and color. A larger circle area indicates a high absolute R value (maximum 1). Red indicates negative values; blue indicates positive values. The black rectangle represents the hypomethylated DMR 2 at HLA-DRB1 identified in whole blood and cell-specific analysis. Bottom panel: whole blood methylation differences ($\Delta\beta$) between pwMS and controls. Pink represent hypermethylation and blue represents hypomethylation.

2.2. Methylation Differences Occur Independently of Known Genetic Risk

Given the strong correlation between methylation and genotype at the MHC region (see Section 2.1 and Supplementary Figure S1), we wanted to determine if epigenetic

risk was independent of the known genetic risk of MS. To do this, we computed EWAS models controlling for the effect of both the *HLA-DRB1* risk haplotype and non-MHC SNPs identified in the 2019 International Multiple Sclerosis Genetics Consortium GWAS [2] using a polygenic risk score (PRS). This PRS was calculated using plink files with PRSice2 [24]. This analysis was only performed in the discovery group because (a) there were missing genetic data from the validation data sets and (b) the discovery group represented early-stage MS, which is an important attribute for risk assessment. From this analysis, we identified 1627 DMPs reaching statistical significance ($p < 9.8 \times 10^{-8}$), 123 of which had $\Delta\beta$ above 2% (see Supplementary File S1). None of the DMPs localized to the *HLA-DRB1* gene, which suggested that the epigenetic effects identified at this locus were linked to genetic variation. To assess if the identified DMPs were modulated by genotype, we tested each CpG for cis-acting methQTLs. An absence of metQTLs suggested that the methylation differences were most likely the result of lifestyle and/or environmental exposure rather than genetic influence. Only 104 out of 1627 (6.4%) produced a significant association with an SNP genotype, showing evidence of modulation by a nearby genetic locus (see Figure 3a). Similarly, we tested each CpG for association with each SNP that was used to construct the PRS. A total of 98 out of 1627 CpGs (6.02%) showed significant modulation by any PRS-associated SNP. The 1627 DMPs mapped to a total of 1092 unique genes. To explore potentially affected biological pathways, we performed over-representation analysis (ORA) on all genes containing DMPs. The list of 1092 unique genes identified after genotype correction revealed emerging pathways in MS such as the *NOTCH* signaling and axon guidance pathways (see Figure 3b).

Using these DMPs, we then constructed a methylation risk score that was corrected for all known genetic risk, both inside and outside of the MHC region (grcMethScore) for each subject. This grcMethScore represents the epigenetic risk of MS independent of the known genetic risk. We then compared the grcMethScore to the *HLA-DRB1* risk haplotype and PRS to assess how well each feature distinguished pwMS from controls. We found that the PRS and *HLA-DRB1* risk haplotype were not correlated (R = 0.067, 95% CI: −0.013–0.14, $p = 0.1$), indicating that the genetic risk conferred by *HLA-DRB1* was independent of other genetic risk factors. Next, we constructed a total known genetic risk score, which incorporated both the PRS and *HLA-DRB1* risk haplotype. The grcMethScore only had a very weak correlation with total genetic risk (see Figure 3c) (R = 0.23, 95% CI: 0.15–0.30, $p = 8.2 \times 10^{-9}$).

To assess how our three risk scores (PRS, HLA-DRB1*15:01, and grcMethScore) could act as classifiers for MS, we conducted area under the curve (AUC) receiver operating curve (ROC) analysis (see Figure 3d). We showed that the total known genetic risk score (HLA-DRB1 + PRS) provided moderate discriminatory accuracy (AUC = 0.72, 95% CI: 0.66–0.76). By comparison, the grcMethScore performed better at distinguishing MS from controls (AUC = 0.85, 95% CI: 0.82–0.89). Combining genotype with methylation score only marginally improved the accuracy of the grcMethScore (AUC = 0.87, 95% CI: 0.84–0.91), as did the addition of all other covariates, such as cell proportions, sex, and age (AUC = 0.89, 95% CI: 0.86–0.92).

2.3. Cell-Specific Differential Methylation Is Mainly Attributed to B Cells and Monocytes

To identify if specific immune cell types were responsible for the differential methylation identified in the whole blood samples, we employed a two-step approach using the combined dataset. First, we calculated cell proportion estimates in the combined dataset using the R package EpiDISH [24]. This revealed a statistically significant decrease in the proportions of natural killer (NK) cells ($mean_{case} = 0.021$, $mean_{control} = 0.031$, $p = 1.34 \times 10^{-8}$) and $CD8^+$ T cells ($mean_{case} = 0.095$, $mean_{control} = 0.106$, $p = 5.92 \times 10^{-5}$) and a statistically significant increase in the $CD4^+$ T cell proportions ($mean_{case} = 0.081$, $mean_{control} = 0.071$, $p = 1.58 \times 10^{-3}$) in MS cases compared to controls. There were no differences in the proportions of monocytes, B cells, and neutrophils between cases and controls (Supplementary Figure S2).

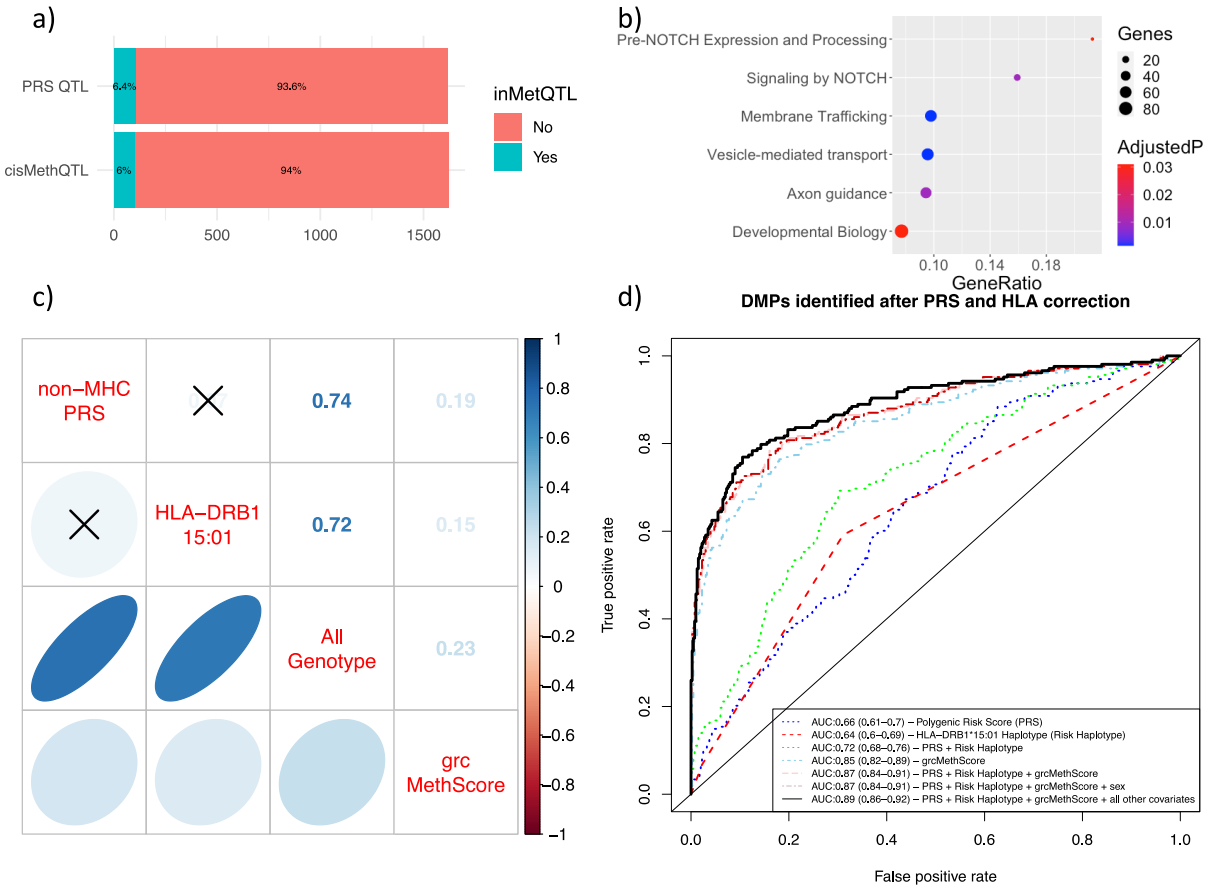


Figure 3. Methylation analysis correcting for known MS risk loci. (**a**) Proportion of DMPs mediated by polymorphisms. Either close polymorphisms (cis methQTLs within ± 20 kb) or SNPs that were used to construct the PRS (PRS QTL). (**b**) ORA on all DMPs identified in the genotype-corrected EWAS. Colors represent FDR *p*-values, dot size represents the number of genes, and gene ratio represents the number of genes in our list divided by the total number of genes in the given pathway. (**c**) Correlation matrix between various scores. Crossed cells represent non-significant values. PRS = PRS inferred from 201 non-MHC SNPs. DR15 haplotype = risk haplotype conferring 3× risk of MS, HLA-DRB1*15:01. All genotype = risk score incorporating PRS and HLA-DRB1*1 5:01. Genotype risk-corrected. MethScore = methylation score corrected for genotype at known MS risk loci. (**d**) ROC curves related to each score discriminating between MS and non-MS controls. AUC = area under the curve.

Second, we used these cell type proportions, in conjunction with CpG methylation (M) values, to calculate cell type-specific DMPs (csDMPs). Here we employed an adaptation of the functions in CellDMC [25]. Although it accounts for all cell types simultaneously, the base CellDMC regression model can become overburdened when incorporating many cell types (i.e., many terms), thus reducing the power and preventing the identification of cell-specific effects. To overcome this, we constructed a "per cell type" linear model (see methods Section 4).

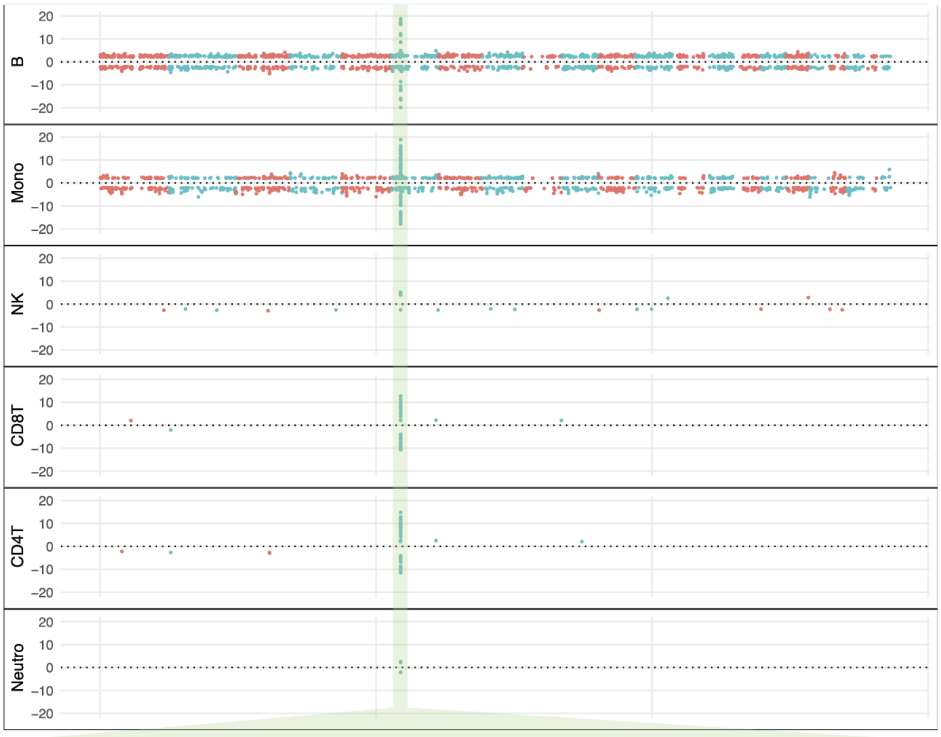
Genome-wide analysis of cell-specific methylation patterns revealed an over-representation of csDMPs in both B cells and monocytes. From the highest number of csDMPs to the lowest were: B cells > monocytes > $CD4^+$ T cells > $CD8^+$ T cells > NK cells, and then neutrophils (Figure 4a). Interestingly, csDMPs were predominantly hypermethylated in B cells (over 60%) while monocytes had a more evenly distributed set of csDMPs (Figure 4a). Consistent with the whole blood analysis, there were marked differential methylation effects at *HLA*, but notably, this effect exhibited obvious variation across cell subtypes (Figure 4b). More detailed analysis of the cell-specific modeling

revealed that the strong differential signal observed at the *HLA* region in whole blood originated primarily from monocytes and B cells but was also present to a lesser extent in T lymphocytes (Figure 4). It was notably absent or nearly absent in both NK cells and neutrophils (Supplementary File S2).

a)

	NK	B	Mono	Neutro	CD8T	CD4T
Total	24	1740	1212	4	38	50
Hypo	16	661	658	2	17	20
Hyper	8	1079	554	2	21	30

b)



c)

Gene	DMR	Index CpG	NK Cells	B Cells	Monocytes	Neutrophils	CD8 T Cells	CD4 T Cells
HLA-DRB1	1	cg13910785	0	21	16	0	10	11
	2	cg16514085	0	-22	-18	-2	-10	-12
	3	cg17416722	0	16	14	0	8	9
	4	cg10385522	0	0	11	0	0	9
HLA-DRB5	5	cg01341801	0	27	21	3	13	15
HLA-DRB6	6	cg19774683	0	19	15	0	9	10
	7	cg06559318	0	17	13	0	7	10
HLA-DQA1	8	cg02919082	0	8	6	0	4	4
HLA-DQB1	9	cg05341252	0	-24	-17	-2	-11	-11

Figure 4. Genome-wide cell-specific DNA methylation profile. (**a**) Total number of cell type-specific DMPs (csDMPs) and breakdown by hyper- and hypomethylated. (**b**) Plot representing genome-wide effect size for all csDMPs, by cell type. Each dot represents a significant csDMP. The *Y*-axis represents the model estimate for effect size and the *X*-axis represents genome coordinates; chromosomes are in chronological order and represented by alternating color. The highlighted region represents the MHC class II regions (**c**) MHC class II region cell-specific DNA methylation. Shown is the effect size in each immune cell type for the index CpG of the nine major DMRs identified in whole blood. Values are estimates extracted from our cell-specific models. Blue/negative values represent negative estimates or hypomethylation at those CpGs, while red/positives values represent positive estimates or hypermethylation. A darker color indicates a higher effect size.

Comparison of the four cell types with the highest number of csDMPs that met both the significance and effect size cut-offs (monocytes, B cells, $CD4^+$ T cells, and $CD8^+$ T cells) revealed that both B cells and monocytes had largely different sets of csDMPs. There were 1165 CpGs unique to B cells, 771 CpGs unique to monocytes, and 347 csDMPs common to both. When we investigated all cell types, we found that CD4+ T cells, CD8+ T cells, monocytes, and B cells shared 28 common csDMPs, with the same directionality. These were located on 7 genes: *HLA-DQB1*, *CMPK2*, *IGF2R*, *HLA-DRB1*, *HLA-DRB6*, *HLA-DRB5*, and *HLA-DQA1* (Supplementary Figure S3 and Supplementary Table S5).

Our validation groups had older mean ages and longer disease duration than the discovery group. To examine if the predominant methylation signals coming from B cells and monocytes occurred early in disease pathology or was a result of long-standing disease, we performed the same analysis described above in only the discovery group. As in the combined analysis, we found the majority of DMPs appeared to be from B cells and monocytes, to a lesser extent from T lymphocytes, with absent or nearly absent signals from both NK cells and neutrophils (see Supplementary Figure S4).

2.4. Over-Representation Analysis Shows Enrichment of DMPs in Immunological Pathways

In the combined whole blood group, we identified 190 DMPs with $\Delta\beta > 2\%$. These DMPs mapped to 92 unique genes (Supplementary File S3). Over-representation analysis (ORA) using this gene list identified 28 pathways. The ORA showed enrichment in immunological pathways, including cytokine- and interferon-related pathways as well as antigen presentation through MHC class I and II (Supplementary Figure S5).

Our cell-specific analysis identified 1740 csDMPs in B cells and 1212 csDMPs in monocytes (Figure 2a). We created two gene lists, one for B cells and one for monocytes, using genes associated with at least one csDMP from that cell type. We then used the two gene lists as well as the gene list from the whole blood DMP analysis to perform over-representation analysis. All the pathways identified in B cells were also present in both whole blood and monocytes. However, the gene list from monocytes was associated with pathways not found in B cells. For example, the cytokine signaling and IFN alpha/beta pathways identified in whole blood were only found in monocytes and not in B cells, suggesting cell-specific changes (Figure 5).

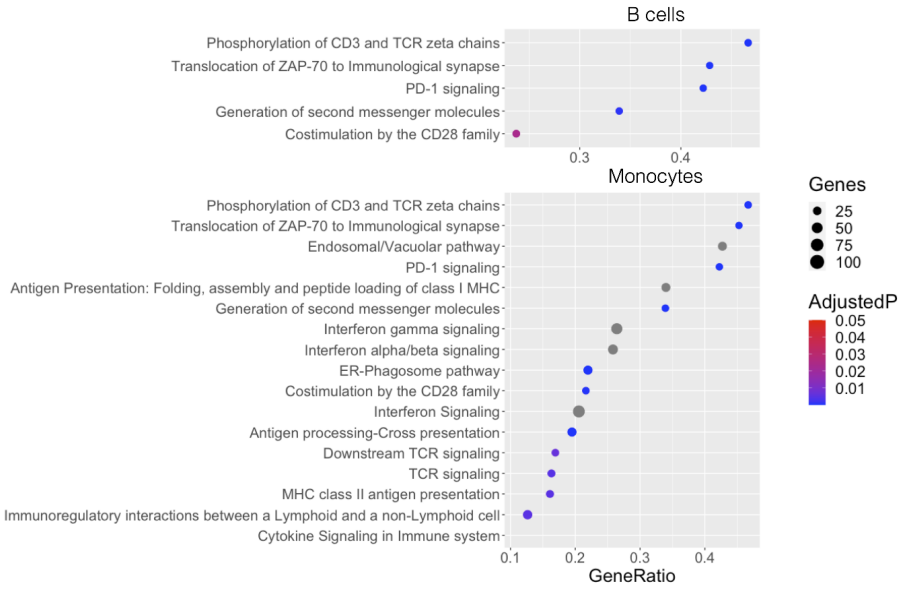


Figure 5. ORA of genes containing cell type-specific DMPs (csDMPs). Lollipop plots represent the ORA results using gene lists of the csDMPs attributed to B cells and monocytes. The size of the marker represents the number of genes. The *X*-axis represents the gene ratio or the number of genes differentially methylated divided by the number of all genes in the given pathway. Levels of significance are represented by colors.

3. Discussion

Despite the significant progress in developing treatments for MS, the pathogenic molecular events that occur prior to, and early in, the disease are not understood. The combined effects of genetics and environmental risk factors suggest that epigenetic mechanisms are likely to be involved.

The key results of this study, recapitulated in Figure 6, indicated that methylation is a better discriminator of MS status than genetics. The moderate discriminatory power (risk) accuracy emerging from both the PRS and *HLA-DRB1* haplotype underlines the limited role played by genetics in MS risk [2]. Correcting for known strong MS risk loci, we identified robust CpGs that were largely not modulated by genotype. We provided evidence that methylation alone is a stronger discriminator of MS than previously described known genetic risk factors. A chain of causation still needs to be established, identifying if these changes in methylation are causative or a consequence of MS. Since these findings occurred in an early-stage MS group (most at the first demyelinating event (FDE)), we can rule out that these effects were a consequence of chronic disease. However, whether these effects occur at birth, during the prodromal period, or at disease onset will be difficult to determine.

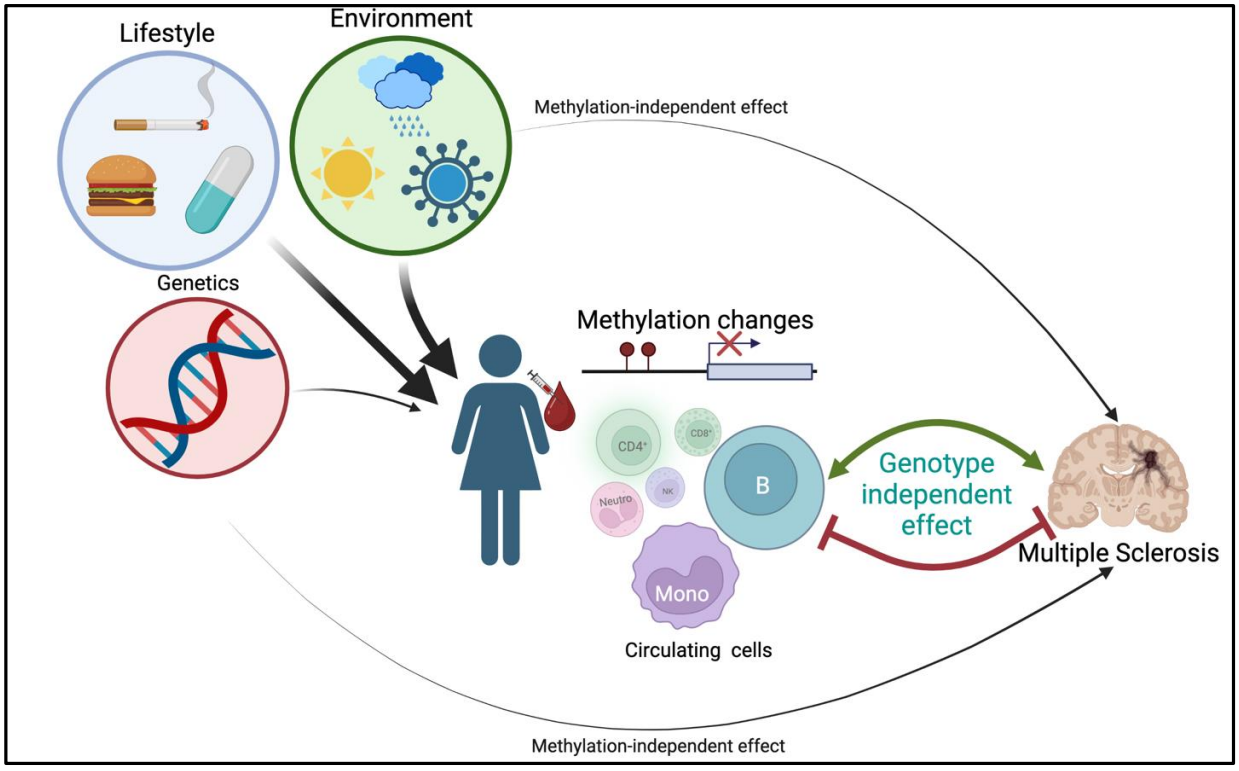


Figure 6. Diagram showing the influence of lifestyle, the environment, and genetics on methylation changes in specific immune cell types in MS.

Since DNA methylation is substantially influenced by environmental factors, the results highlight the importance of environmental exposure in MS risk. Further studies need to systematically evaluate the effect of environmental exposure on methylation change and MS risk, such as seen in vitamin D [26], smoking [27], or EBV infection. Interestingly, controlling for the influence of genotype also unveiled non-immune-related pathways disrupted in pwMS, such as the NOTCH signaling or axon guidance pathways, both of which emerged as key pathways in MS [28,29]. Axon guidance molecules are known to modulate axon regeneration in disease and are involved in different inflammation-related pathways [30]. NOTCH signaling regulates oligodendrocyte progenitor cell differentiation and downstream myelin production [31].

Numerous studies have now been published investigating the genome-wide DNA methylation changes in whole blood and isolated cell types; however, the results are inconsistent and limited by heterogeneous cell populations, disease heterogeneity, use of a wide variety of disease modifying therapies (DMTs), or small sample size [14,16,18,19,32]. Using a large dataset of MS cases early in their diagnosis with age/sex/location-matched controls, we confirmed previous findings that DNA methylation acts as mediator of the *HLA-DRB1*1501* allele [16] and provided evidence that this occurs early in disease pathology rather than as a result of treatment or long-standing disease. Furthermore, we revealed that the differential methylation was primarily associated with B cells and monocytes, with a smaller contribution from T lymphocytes ($CD4^+$ and $CD8^+$). This is consistent with our previous observations in these four isolated cell types [18]. We also correlated methylation change with gene expression at the *HLA-DRB1* locus, thereby substantiating its key role in MS pathogenesis, which we replicated in two independent patient cohorts. However, outside of the MHC region, significant methylation changes appeared to be unrelated to meaningful genotype changes thought to modulate MS risk (identified through the 2019 IMSGC GWAS [2]).

A major strength of our study was the discovery group, with data collected early in diagnosis and the group followed prospectively over a ten-year period. This allowed us to examine the methylation profiles of pwMS at baseline collection, who, at their 10-year follow-up visit, had fulfilled the criteria for MS. We demonstrated that B cells and monocytes accounted for the largest contribution of the whole blood DNA methylation signal in MS and that this was not simply a result of long-standing disease. This is consistent with previous work on isolated cell subtypes from pwMS [17,18,33] and with current knowledge of B cells playing an important role in the pathogenesis of disease, highlighted by the efficacy of treatments targeting the CD20 surface antigen predominantly expressed on B cells [34]. There has been the suggestion that these therapies specifically target the memory B cell compartment [35], so future studies investigating the methylation of memory B cells will be of interest. In contrast to previous work, the csDMPs identified in B cells were predominantly hypermethylated.

Monocytes, which also have antigen-presenting properties, likely play an important role in early-stage disease by mediating both pro- and anti-inflammatory responses [36–38]. Furthermore, our ORA and GAT analyses revealed monocyte differential methylation signals included in unique pathways, suggesting they may play an early role in MS pathogenesis. Increased proportions of $CD4^+$ T cells and a deficit of $CD8^+$ T cells in MS cases compared to controls have been previously documented [39–41]. The increased proportions of $CD4^+$ T cells are likely driven by Th17 helper cells, whereas the deficit of $CD8^+$ T cells is probably due to reduced effector cell numbers [39–41]. Our cell proportion estimates are consistent with these past findings. The proportion of total NK cells has not previously been evaluated in MS; however, we have previously demonstrated that subsets of NK cells are affected by disease-modifying therapies and an increase in the abundance of NK_{bright} cells is associated with stable disease [42]. To overcome issues of heterogeneous cell populations and small sample size, we used an adapted CellDMC method to identify which cell types contributed the most to the differential methylation signal. We were able to demonstrate that the main source of differential methylation in whole blood came from B cells and monocytes, particularly in the *HLA-DR* region. This is not surprising, given that HLA-DR is expressed primarily on antigen-presenting cells such as B lymphocytes and monocytes [43]. We also found strong differential methylation signals from both $CD4^+$ and $CD8^+$ T cells. HLA-DR is normally absent on resting T lymphocytes; however, it has been proposed as a marker of immune system activation [44]. In systemic lupus erythematosus, $CD8^+$ T and $CD4^+$ T cells have increased HLA-DR expression during active disease, leading to speculation that expression of HLA-DR represents an inflammatory response. It is plausible that this is also happening in the T lymphocytes of people with MS, but further studies are required to confirm this.

In addition to the 13 DMPs located at *HLA* genes, we also identified two robust signals that localized to interferon-related genes. This suggests that these DMPs may be partially

influenced by treatment. One small study found global methylation changes in monocytes after treatment with interferon beta [45]. The authors of this study, however, were unable to identify changes at specific CpGs [45]. A study of monozygotic twin pairs with MS found DMPs associated with IFN beta treatment; however, we did not find any overlap with this study [46]. A larger sample size and prospective study design to confirm these findings is required. The results of our csDMP analysis was compared to two previously reported studies, Ewing et al. [18] and Ma et al. [33], which revealed 154 shared genes containing csDMPs in B cells and two shared genes in monocytes in all three studies (Supplementary Figure S6), suggesting these may be very robust DMPs for these cell types. No shared DMPs were identified in either the $CD4^+$ or $CD8^+$ T cell populations. Differences would be expected between studies due to the use of different assays (Ewing et al. used the 450 K platform and Ma et al. used bisulfite sequencing), different group statistical powers, and different group characteristics. The overlap likely represents replicated MS-associated DMPs.

One limitation of this study is that the combined dataset was heterogenous in terms of clinical characteristics. Although our discovery group was primarily treatment naïve and early in their disease course, the validation groups contained participants on mixed treatments and with long-standing disease. Thus, we cannot draw conclusions regarding how treatment use or disease progression affects methylation change. Furthermore, despite the large size of the combined dataset, the statistical deconvolution methods were still only likely to identify major effect DMPs. Smaller effect CpGs probably exist, but much larger studies are needed to identify them. Another limitation is that the expression data were from separate participant samples; therefore, we cannot assign causation between expression and methylation, we can only make correlations. Future studies with larger groups of treatment-naïve populations, followed prospectively over time, are necessary to confirm our DNA methylation and expression data and confirm the results presented herein. Finally, we acknowledge that ORA is an exploratory analysis to identify potentially affected biological pathways and help guide further studies, but the results need to be confirmed at a molecular level.

4. Materials and Methods

See Figure 1 for a study workflow diagram.

Clinical characteristics have been listed in Supplementary Table S1.

4.1. Whole Blood Methylation

The discovery group was comprised of whole blood samples collected at baseline from The Australian Multi-Centre Study of Environment and Immune Function (Ausimmune study) [22]. From this group, there were 208 MS cases and 402 matched non-MS controls. Most of this group entered the Ausimmune study after their first clinical diagnosis of CNS demyelination (FCD), were primarily treatment naïve at the time of sampling and had mainly the relapse-onset phenotype. We used whole blood samples from the baseline collection in cases who, at the 10-year follow-up, had converted to MS based on the 2017 McDonald criteria [47]. Controls were matched for age, sex, and region [22].

The first validation group was from Sweden and consisted of 140 MS cases and 139 controls. This dataset consisted of previously published data from prevalent MS cases who were primarily on treatment at the time of blood sampling, mostly had the relapse-onset phenotype, and had been profiled using the Illumina 450 K array [16].

The second validation group consisted of 235 MS cases sourced from four Australian MS specialist clinics in Melbourne, Newcastle, Hobart, and Adelaide. The cases in this group exclusively had the relapse-remitting phenotype, but had longer disease duration (median 16 years), and were on a variety of treatments at sample collection [48]. This case group did not have controls, so they were matched with data from 102 non-MS controls from a publicly available dataset.

All DNA samples from the discovery groups and case samples from the second validation group were tested in-house for genome-wide methylation using the Illumina Infinium Human Methylation EPIC BeadChip Kit (EPIC array) (Illumina Inc., San Diego, CA, USA). The publicly available dataset matched to the second validation group was also profiled using the EPIC array.

To identify DNA methylation changes associated with MS, we performed an EWAS on bisulfite-treated genomic DNA isolated from whole blood (see Figure 1 and methods Section 4). A total of 583 people with MS (pwMS) and 643 controls from three independent groups were analyzed in this study. Further details of each group are shown in Supplementary Table S1 and in the methods section.

Discovery Group: details of the Ausimmune Study design have been previously published [22]. Briefly, between 2003 and 2006, participants were recruited across four different centers in Brisbane, Newcastle, and surrounding areas; Geelong; the Western Districts of Victoria; and the Island of Tasmania. Cases were aged 18–59 years and entered the study around the time of their first demyelinating event (FDE). They were prospectively followed over 10 years. The Australian electoral roll was used to select controls from the general population at baseline. Non-MS controls were matched based on age ($\pm$2 years), gender, and region, with a ratio of up to 4 controls for each case.

Epidemiological information was collected by face-to-face interview, self-completed questionnaires, and neurological examination. At the same time, biological samples were collected and stored. Detailed analysis of the information collected has been previously published [22]. For this study, we had access to whole blood from 208 MS cases and 402 matched controls. All 208 MS cases had converted to CDMS by the 10-year follow-up.

The first validation group was from a previously published case–control study (group 2) and the 450 K array data were available from the Gene Expression Omnibus (GEO) database under accession numbers GSE43976 and GSE106648. The dataset was comprised of whole blood DNA from 140 MS cases and 139 non-MS controls. Samples were collected in Sweden between 2005 and 2011, the cases were mostly on low-efficacy treatment (65%) and mostly had the relapsing-remitting phenotype.

The second validation group was comprised of 235 MS cases obtained from the specimen banks of 4 Australian specialist centers in Melbourne, Newcastle, Hobart, and Adelaide and 102 publicly available control idat files obtained through the EWAS data hub (https://ngdc.cncb.ac.cn/ewas/datahub/repository, accessed on 8 July 2022) using the following criteria: age between 23 and 73 years, sex = female, platform = 850 K, tissue = whole blood, and sample type = control. The second validation group was publicly available in the GEO database (GSE106648) and was previously used in a whole blood case–control EWAS [48].

4.2. DNA Extraction and Quality Control

For the discovery group and second validation dataset, DNA was extracted from whole blood using the QIAamp DNA Blood Mini Kit™ (Qiagen, Venlo, The Netherlands,). Extracted DNA was quantified using the Qbit (Invitrogen, Waltham, MA, USA) and DNA integrity was assessed using the Genomic DNA ScreenTape assay with the TapeStation system (Agilent, Santa Clara, CA, USA), using the DNA integrity number (DIN) as a metric. All samples had a DIN $\geq$ 7, indicating minimal genomic DNA degradation.

4.3. Methylation and Genotyping Arrays

To determine the SNP profiles, the Global Screening Array v2.1 (Illumina, USA) (hereafter referred to as GSA) was used according to the manufacturer's guidelines. For genome-wide methylation analysis, 500 ng of genomic DNA was bisulfite-converted using the EZ-DNA Methylation™ Kit (Zymo Research, Irvine, CA, USA,) according to the manufacturer's guidelines. The converted DNA was hybridized to the Illumina Methylation EPIC BeadChip array (hereafter referred to as EPIC array). To avoid batch effects, samples were randomized on both the GSA and EPIC arrays using the R package OSAT [49]. Both

the GSA and EPIC arrays were read using the iScan system (Illumina Inc., San Diego, CA, USA) to produce raw idat files.

4.4. Methylation Analysis

For the discovery and both validation groups, analysis was performed using the R package ChAMP [50,51]. To summarize, idat files were loaded and filtered to remove low performing probes, probes mapping to multiples loci, and low performing samples. Probes next to known polymorphisms and on the XY chromosome were kept. Beta values were then normalized using the BMIQ method [52]. Batch effects at both the array and chip levels were corrected using the Combat algorithm [53].

The final discovery dataset contained 742,961 CpGs and 610 individuals (208 cases and 402 controls). The first validation dataset contained 478,687 CpGs across 279 samples. The second validation dataset contained 831,858 probes across 337 samples.

The final model used to identify differentially methylated positions (DMPs) was constructed using logistic regression, whereby phenotype (case or control) was considered the outcome variable and each CpG beta value and cell type proportion (natural killer cells, monocytes, B cells, $CD8^+$ T cells, $CD4^+$ T cells, and neutrophils) were considered predictors. For CpG_i:

$$glm\left(Outcome_{\frac{Case}{Control}} \sim CpG_i + Age + Sex + CellFraction + Region\right)$$

The mean difference in methylation between the case and control groups was used to assess the effect size (i.e., delta beta or $\Delta\beta$).

When performing the genotype risk-corrected EWAS (correcting for known genotype risk loci (HLA-DRB1 RISK HAPLOTYPE) and aggregated polygenic risk score), we also incorporated genotype information in the model to identify genotype risk-independent DMPs:

$$glm\left(Outcome_{\frac{Case}{Control}} \sim CpG_i + Age + Sex + CellFraction + Region + HLA\ Haplotype + PRS\right)$$

Significant DMRs were determined using multiple Fisher's exact tests, Stouffer's test, and the harmonic mean of the individual component FDRs, at $p \leq 0.05$.

4.5. Sensitivity Analysis

Sensitivity analysis was performed on a range of covariates to assess their relative individual importance on CpG beta value. Numeric continuous variables were assessed using linear models such as $lm(CpG_i \sim Covariate)$ and categorical variables were assessed using ANOVA models such as $AOV(CpG_i \sim Covariate)$.

4.6. Genotype Analysis

Raw idat files were loaded in GenomeStudio (Illumina Inc., San Diego, CA, USA) for filtering. Recommendations from Guo et al., 2014 [54] were followed (except the ethnicity-specific steps) and the analysis-ready dataset was exported to plink [55] format. In short, QC was performed to filter haploid probes, then the probes were filtered based on their GenTrain score (score ≥ 0.7), cluster separation, SNP call rate ($\geq$98%), and sample call frequency ($\geq$95%). After exporting to plink format, samples were tested for gender mismatch, relatedness, and Hardy–Weinberg equilibrium. Plink was also used to generate association statistics, using Fisher's exact test. The HLA-DRB1*15:01 haplotype was detected using the rs3135388 tag SNP [56].

For the CIT analysis, we used Haploview to identify the distribution of haplotypes. For each gene and group, we identified the haplotype that was most significant in an X^2 test as a risk haplotype.

4.7. Scores and Receiver Operator Characteristic Curves

We constructed a genotype risk-corrected methylation score (grcMethScore) only incorporating DMPs identified in the genotype risk-corrected EWAS. The grcMethScore was constructed using the following formula:

$$\text{MethScore}_{sample} = \sum \beta value_{\ at\ CpGi} \times \Delta\beta_{CpGi}$$

where $\beta value_{\ at\ CpGi}$ represents the beta value of this particular sample at a CpG of interest and $\Delta\beta_{CpGi}$ is a weight, using the value of delta beta identified in the genotype risk-independent EWAS (described in the methylation analysis Section 4.4).

For genotyping, we used two scores: the risk haplotype HLA-DRB1 score (either 0, 1, or 2) and a polygenic risk score (PRS) generated using PRSice [24] with all SNPs identified through the 2019 IMSGC GWAS [2].

ROC curves and the area under the curve (AUC) were computed using the R package ROCR [57] using the scores generated with logistic regression.

$$glm\left(Outcome_{\frac{Case}{Control}} \sim Predictor_1 + \cdots + Predictor_n\right)$$

where outcome is the MS-Control outcome and *Predictors* are grcMethScore, PRS, cell proportion, sex, etc.

4.8. Immune Cell Deconvolution Analysis

Cell proportion estimates were obtained using the R package EpiDISH [58]. Cell proportions, in conjunction with M-values, were used to calculate cell type-specific DMPs, using a method inspired by the CellDMC function [25]. CellDMC has two major limitations. First, the results are biased based on initial cell proportions when using beta values. Second, CellDMC modeling is overburdened when incorporating numerous cell types, preventing the identification of effect in all or multiple cell types.

To compensate for this, we used a two-step approach using the combined dataset for maximum statistical power. First, we calculated cell proportion estimates using the R package EpiDISH. Second, we used the cell type proportions in conjunction with CpG methylation (M) values to calculate cell type-specific DMPs (csDMPs). Here we employed an adaptation of the functions in CellDMC [25]. Although it accounts for all cell types simultaneously, the base CellDMC regression model can become overburdened when incorporating many cell types (i.e., many terms), thus reducing the power and preventing the identification of cell-specific effects in multiple cell types.

Instead, for each cell type, we constructed a linear model as follows:

$$lm(CpG_i \sim cellFrac + cellFrac : Phenotype + Covariates)$$

where CpG_i represent the M-value at CpG_i, *cellFrac* is the cell type fraction (ranging from 0 to 1), and *cellFrac: Phenotype* represents the interaction term between cell fraction and phenotype. CpGs with an interaction cellFrac:Phenotype estimate above 2 or under -2, as well as a p-value under 9.8×10^{-8}, were kept for further downstream analysis.

4.9. Expression and DNA Methylation Analysis

Monocytes: Expression data were comprised of data from an independent, cross-sectional case–control study. Expression data from monocytes from previously published work were acquired from the authors under a data transfer agreement and DNA methylation data obtained using the 450 K platform were freely available from GEO [59] (GSE43976). Monocytes were isolated from PBMCs by magnetic separation (Miltenyi, Bergisch Gladbach, Germany) according to the manufacturer's instructions and expression of HLA-DRB1 was quantified by real-time PCR, as previously described [16].

$CD4^+$ T cells: Our group previously published the results of DNA methylation analysis using the 450 K array. From this study, we also had matched RNA available from 18 relapse-

onset MS cases (unpublished results). The MS cases were all female and either treatment naïve or had been off MS-specific treatment for at least 6 months. They were matched with 33 age- and sex-matched non-MS controls.

The RNA from $CD4^+$ T cells was extracted using the RNAeasy mini kit (Qiagen, Germantown, MD, USA) following the manufacturer's protocols. The resulting RNA quality was assessed using the Agilent Bioanalyzer and met a quality cut-off of an RIN number of 7 or higher. Samples meeting QC requirements were hybridized to the Illumina HT12 expression array card (service provided by Diamantina Institute, Australia).

Expression analysis was carried using the R package beadarray [60]. In short, idat files were loaded using the readIdatFiles function, data were normalized using the normaliseIllumina function with the quantile method. Finally, the limmaDE function was used to determine differential gene expression and we used the illuminaHumanv4.db package to annotate the results.

B cells: The B cell dataset was composed of a total of 73 samples, including 23 controls (13 non-inflammatory neurologic disease controls and 10 healthy individuals) as well as 50 MS cases (39 RRMS and 11 SPMS). Peripheral blood mononuclear cells (PBMCs) were isolated directly after collection using standard Ficoll (GE Healthcare) and sodium citrate-containing preparation tubes (Becton Dickinson) procedures, respectively. CD19+ B cells were isolated using positive selection by magnetic separation on MACS MicroBeads (Miltenyi), according to the manufacturer's instructions (>95% purity). Cells were prepared within an hour after sampling using an AutoMACS (Milteny Biotec, Bergisch, Germany) according to the manufacturer's standard protocol and stored at −70 °C. DNA and RNA were extracted simultaneously using the Qiagen Allprep DNA/RNA kit (Qiagen, Venlo, The Netherlands) according to the manufacturer's standard protocol and stored at −70 °C. The amount and quality of DNA were assessed using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies Inc.). Samples with sufficient DNA amounts were used in further analysis. Processing of the samples for the Infinium HumanMethylationEPIC array (Illumina), including bisulfite conversion, was performed at the National Genomics Institutet (NGI) at the Science for Life Lab (Uppsala, Sweden).

RNA-Seq libraries were generated from 500 ng of total RNA using the Illumina TruSeq mRNA Stranded Library Preparation Kit (cat. no. RS-122-2103) according to the manufacturer's protocol. Library quality was determined using the Agilent High Sensitivity DNA Kit (cat. no. 5067-4626) and a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA). Libraries were sequenced on the Illumina HiSeq 2500 as per the manufacturer's instructions, and ∼20 M 75 bp paired-end reads were generated per sample. Read quality was assessed before and after trimming using FastQC v0.11.4. Illumina adapters, and low-quality nucleotides were trimmed using Cutadapt v1.9.1. Reads were aligned to the human genome (GRCh38), and read count per gene was determined using STAR.

All cell types:

Methylation profiles for every cell type were analyzed individually using the R packages Minfi [61] and ChAMP [50,51] following the pipeline according to Marabita et al. [62]. Briefly, type 1 and type 2 probes were normalized using quantile normalization and BMIQ. Sex was confirmed using the GetSex function in the Minfi package and cell type identity was confirmed using the cell type deconvolution method in the Minfi package based on the Houseman algorithm [63]. Age was confirmed using the AgeP function in the WateRmelon package [64]. Furthermore, sample identification was confirmed by running a 44 SNP panel at NGI, and genotypes were confirmed by overlapping SNP calls with data from our genetic studies. Except for data related to pwMS using rituximab and DMF because of the pairwise analysis, the following probes were filtered out: (i) probes not passing the detection cut-off p-value of 0.01, (ii) probes with known SNPs, and (iii) X and Y chromosome probes. Batch effects were identified using principal component analysis (PCA) and corrected using the ComBat function in the SVA package [53]. The loading of methylation profiles was performed in this manner for each dataset used in this study. DMPs were determined by

linear modeling using the limma package [65] in a model that included age, sex, and B cell purity from the Houseman algorithm as covariates.

4.10. Correlation Analysis (DNA Methylation and Gene Expression)

Correlation analysis was performed using the R package corrplot. All three cell types were analyzed separately. In every case, a measure of expression (from RT-PCR, array, or sequencing) was correlated with a beta value for every sample. The Pearson R coefficient is reported on the plots.

4.11. metQTL Analysis

We tested for the influence of genotype on methylation in two instances. Each time, we performed an ANOVA test:

$$AOV(CpG \sim SNP_{AA-AB-BB})$$

where CpG represents the beta value at a specific CpG and $SNP_{AA-AB-BB}$ represents the genotype group.

We tested both the influence of the *HLA-DRB1*15:01* risk haplotype on genome-wide methylation and all SNPs within ±18 kBp (36 kBp window) for each DMP identified in the genotype risk-independent EWAS.

4.12. Over-Representation Analysis

ORA was performed using the online Reactome platform [66] to analyze various gene lists with the inclusion of IntAct [67] interactors in the analysis. Pathways were kept if the FDR value of the individual component was under 0.05. Figures were generated using the R package ggplot2 [68].

5. Conclusions

This study provides compelling evidence that DNA methylation, independently of genotype, is strongly modified in pwMS and is implicated in the NOTCH and axon guidance pathways. The results of this study advance previous work implicating B cells and monocytes in MS pathogenesis by revealing that differential methylation is present at disease onset. The effect of differential methylation of B cells and monocytes is linked to gene expression differences in these cell types. While therapies affecting B cells have been shown to be extremely effective in MS, monocytes may also represent a future therapeutic target. Finally, we show that methylation, after correcting for known genetic risk, outperforms genetics when discriminating between MS and non-MS individuals (see Figure 3). This provides compelling evidence that environmental exposure and lifestyle play a larger role in MS onset than genetics.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms241612576/s1.

Author Contributions: R.J.S., R.A.L. and J.L.-S. contributed to the conception of the work. R.J.S., R.A.L. and J.L.-S. contributed to the design of the work. A.X., J.N.T., M.S., H.B., I.K., L.K., V.G.J., T.K., M.J., L.A., B.V.T., R.J.S., A.-L.P., R.A.L. and J.L.-S. contributed to the acquisition of data. A.X., V.E.M., E.E., M.P.C., S.M.B. and R.A.L. contributed to the analysis of data. A.X., V.E.M., R.A.L. and J.L.-S. contributed to the interpretation of data. A.X., V.E.M., R.J.S., R.A.L. and J.L.-S. contributed to drafting the manuscript. V.E.M., H.B., V.G.J., T.K., M.J., B.V.T., R.J.S., A.-L.P., R.A.L. and J.L.-S. contributed to substantial revision of the manuscript. The Ausimmune/AusLong investigator group were responsible for patient recruitment, assessment and subsequent follow-ups. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the following funding bodies: NHMRC (AusLong) grant APP1127819, NMSS (AusImmune) grant RG-1803-30499, and MSRA (severity study) grant 18-0424.

Institutional Review Board Statement: The Ausimmune Study was approved by nine regional Human Research Ethics Committees (study number 09/04/15/5.04 and HREC/09/HNE/139).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets generated and analyzed during the current study are not publicly available due to the General Data Protection Regulation but are available from the corresponding author upon reasonable request.

Acknowledgments: The authors acknowledge support from the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. The Ausimmune/AusLong investigators group: Robyn Lucas (National Centre for Epidemiology and Population Health, Canberra), Keith Dear (Duke Kunshan University, Kunshan, China), Ingrid van der Mei, Leigh Blizzard, Steve Simpson-Yap, Simon Broadley (Griffith University, Gold Coast, Australia), David Williams, Cameron Shaw and Caron Chapman (Barwon Health, Geelong, Australia), Alan Coulthard (University of Queensland, Brisbane, Australia), and Patricia Valery (QIMR Berghofer Medical Research Institute, Brisbane, Australia), as well as Terry Doyer and Michael Pender. The authors would also like to acknowledge Peri Noori and David Gomez-Cabrero for their contribution to the work on isolated B cells. Finally, the authors would like to thank all the pwMS and control subjects who participated in this research.

Conflicts of Interest: H.B. has received institutional (Monash University) funding from Biogen, F. Hoffmann-La Roche Ltd., Merck, Alexion, CSL, and Novartis; has carried out contracted research for Novartis, Merck, F. Hoffmann-La Roche Ltd., and Biogen; has taken part in speakers' bureaus for Biogen, Genzyme, UCB, Novartis, F. Hoffmann-La Roche Ltd., and Merck; has received personal compensation from Oxford Health Policy Forum for the Brain. J.L.-S received travel compensation from Biogen, Merck, and Novartis; has been involved in clinical trials with Biogen, Merck, Novartis, and Roche; her institution has received honoraria for talks and advisory board service from Biogen, Merck, Novartis, and Roche. V.E.M. has accepted honoraria for presentations and research funds from Biogen and Merck.

References

1. International Multiple Sclerosis Genetics Consortium; Beecham, A.H.; Patsopoulos, N.A.; Xifara, D.K.; Davis, M.F.; Kemppinen, A.; Cotsapas, C.; Shah, T.S.; Spencer, C.; Booth, D.; et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* **2013**, *45*, 1353–1360. [CrossRef] [PubMed]
2. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* **2019**, *365*, eaav7188. [CrossRef] [PubMed]
3. Kellar-Wood, H.F.; Wood, N.W.; Holmans, P.; Clayton, D.; Robertson, N.; Compston, D.A. Multiple sclerosis and the HLA-D region: Linkage and association studies. *J. Neuroimmunol.* **1995**, *58*, 183–190. [CrossRef]
4. Sawcer, S.; Ban, M.; Maranian, M.; Yeo, T.W.; Compston, A.; Kirby, A.; Daly, M.J.; De Jager, P.L.; Walsh, E.; Lander, E.S.; et al. A high-density screen for linkage in multiple sclerosis. *Am. J. Hum. Genet.* **2005**, *77*, 454–467. [CrossRef] [PubMed]
5. Patsopoulos, N.A.; Barcellos, L.F.; Hintzen, R.Q.; Schaefer, C.; van Duijn, C.M.; Noble, J.A.; Raj, T.; Imsgc; Anzgene; Gourraud, P.A.; et al. Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet.* **2013**, *9*, e1003926. [CrossRef]
6. Attfield, K.E.; Jensen, L.T.; Kaufmann, M.; Friese, M.A.; Fugger, L. The immunology of multiple sclerosis. *Nat. Rev. Immunol.* **2022**, *22*, 734–750. [CrossRef]
7. Bar-Or, A.; Li, R. Cellular immunology of relapsing multiple sclerosis: Interactions, checks, and balances. *Lancet Neurol.* **2021**, *20*, 470–483. [CrossRef] [PubMed]
8. Ruprecht, K. The role of Epstein-Barr virus in the etiology of multiple sclerosis: A current review. *Expert. Rev. Clin. Immunol.* **2020**, *16*, 1143–1157. [CrossRef] [PubMed]
9. Dunn, N.; Kharlamova, N.; Fogdell-Hahn, A. The role of herpesvirus 6A and 6B in multiple sclerosis and epilepsy. *Scand. J. Immunol.* **2020**, *92*, e12984. [CrossRef] [PubMed]
10. Bjornevik, K.; Cortese, M.; Healy, B.C.; Kuhle, J.; Mina, M.J.; Leng, Y.; Elledge, S.J.; Niebuhr, D.W.; Scher, A.I.; Munger, K.L.; et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science* **2022**, *375*, 296–301. [CrossRef] [PubMed]
11. Hedstrom, A.K. Smoking and disability progression in multiple sclerosis. *Expert. Rev. Neurother.* **2020**, *20*, 739–741. [CrossRef] [PubMed]

12. Lucas, R.M.; Ponsonby, A.L.; Dear, K.; Valery, P.C.; Pender, M.P.; Taylor, B.V.; Kilpatrick, T.J.; Dwyer, T.; Coulthard, A.; Chapman, C.; et al. Sun exposure and vitamin D are independent risk factors for CNS demyelination. *Neurology* **2011**, *76*, 540–548. [CrossRef] [PubMed]
13. Ponsonby, A.L.; Lucas, R.M.; van der Mei, I.A.; Dear, K.; Valery, P.C.; Pender, M.P.; Taylor, B.V.; Kilpatrick, T.J.; Coulthard, A.; Chapman, C.; et al. Offspring number, pregnancy, and risk of a first clinical demyelinating event: The AusImmune Study. *Neurology* **2012**, *78*, 867–874. [CrossRef] [PubMed]
14. Maltby, V.E.; Graves, M.C.; Lea, R.A.; Benton, M.C.; Sanders, K.A.; Tajouri, L.; Scott, R.J.; Lechner-Scott, J. Genome-wide DNA methylation profiling of CD8+ T cells shows a distinct epigenetic signature to CD4+ T cells in multiple sclerosis patients. *Clin. Epigenetics* **2015**, *7*, 118. [CrossRef] [PubMed]
15. Maltby, V.E.; Lea, R.A.; Sanders, K.A.; White, N.; Benton, M.C.; Scott, R.J.; Lechner-Scott, J. Differential methylation at MHC in CD4(+) T cells is associated with multiple sclerosis independently of HLA-DRB1. *Clin. Epigenetics* **2017**, *9*, 71. [CrossRef]
16. Kular, L.; Liu, Y.; Ruhrmann, S.; Zheleznyakova, G.; Marabita, F.; Gomez-Cabrero, D.; James, T.; Ewing, E.; Linden, M.; Gornikiewicz, B.; et al. DNA methylation as a mediator of HLA-DRB1(star)15:01 and a protective variant in multiple sclerosis. *Nat. Commun.* **2018**, *9*, 2397. [CrossRef]
17. Maltby, V.E.; Lea, R.A.; Graves, M.C.; Sanders, K.A.; Benton, M.C.; Tajouri, L.; Scott, R.J.; Lechner-Scott, J. Genome-wide DNA methylation changes in CD19(+) B cells from relapsing-remitting multiple sclerosis patients. *Sci. Rep.* **2018**, *8*, 17418. [CrossRef]
18. Ewing, E.; Kular, L.; Fernandes, S.J.; Karathanasis, N.; Lagani, V.; Ruhrmann, S.; Tsamardinos, I.; Tegner, J.; Piehl, F.; Gomez-Cabrero, D.; et al. Combining evidence from four immune cell types identifies DNA methylation patterns that implicate functionally distinct pathways during Multiple Sclerosis progression. *eBioMedicine* **2019**, *43*, 411–423. [CrossRef]
19. Bos, S.D.; Page, C.M.; Andreassen, B.K.; Elboudwarej, E.; Gustavsen, M.W.; Briggs, F.; Quach, H.; Leikfoss, I.S.; Bjolgerud, A.; Berge, T.; et al. Genome-wide DNA methylation profiles indicate CD8+ T cell hypermethylation in multiple sclerosis. *PLoS ONE* **2015**, *10*, e0117403. [CrossRef]
20. Rhead, B.; Brorson, I.S.; Berge, T.; Adams, C.; Quach, H.; Moen, S.M.; Berg-Hansen, P.; Celius, E.G.; Sangurdekar, D.P.; Bronson, P.G.; et al. Increased DNA methylation of SLFN12 in CD4+ and CD8+ T cells from multiple sclerosis patients. *PLoS ONE* **2018**, *13*, e0206511. [CrossRef]
21. Campagna, M.P.; Xavier, A.; Lechner-Scott, J.; Maltby, V.; Scott, R.J.; Butzkueven, H.; Jokubaitis, V.G.; Lea, R.A. Epigenome-wide association studies: Current knowledge, strategies and recommendations. *Clin. Epigenetics* **2021**, *13*, 214. [CrossRef] [PubMed]
22. Lucas, R.; Ponsonby, A.L.; McMichael, A.; van der Mei, I.; Chapman, C.; Coulthard, A.; Dear, K.; Dwyer, T.; Kilpatrick, T.; Pender, M.; et al. Observational analytic studies in multiple sclerosis: Controlling bias through study design and conduct. The Australian Multicentre Study of Environment and Immune Function. *Mult. Scler.* **2007**, *13*, 827–839. [CrossRef] [PubMed]
23. Millstein, J.; Zhang, B.; Zhu, J.; Schadt, E.E. Disentangling molecular relationships with a causal inference test. *BMC Genet.* **2009**, *10*, 23. [CrossRef] [PubMed]
24. Choi, S.W.; O'Reilly, P.F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* **2019**, *8*, giz082. [CrossRef]
25. Zheng, S.C.; Breeze, C.E.; Beck, S.; Teschendorff, A.E. Identification of differentially methylated cell types in epigenome-wide association studies. *Nat. Methods* **2018**, *15*, 1059–1066. [CrossRef]
26. Ong, L.T.C.; Schibeci, S.D.; Fewings, N.L.; Booth, D.R.; Parnell, G.P. Age-dependent VDR peak DNA methylation as a mechanism for latitude-dependent multiple sclerosis risk. *Epigenetics Chromatin* **2021**, *14*, 9. [CrossRef]
27. Marabita, F.; Almgren, M.; Sjoholm, L.K.; Kular, L.; Liu, Y.; James, T.; Kiss, N.B.; Feinberg, A.P.; Olsson, T.; Kockum, I.; et al. Smoking induces DNA methylation changes in Multiple Sclerosis patients with exposure-response relationship. *Sci. Rep.* **2017**, *7*, 14589. [CrossRef]
28. Christopoulos, P.F.; Gjolberg, T.T.; Kruger, S.; Haraldsen, G.; Andersen, J.T.; Sundlisaeter, E. Targeting the Notch Signaling Pathway in Chronic Inflammatory Diseases. *Front. Immunol.* **2021**, *12*, 668207. [CrossRef]
29. Lee, W.S.; Lee, W.H.; Bae, Y.C.; Suk, K. Axon Guidance Molecules Guiding Neuroinflammation. *Exp. Neurobiol.* **2019**, *28*, 311–319. [CrossRef]
30. Dominguez-Romero, M.E.; Slater, P.G. Unraveling Axon Guidance during Axotomy and Regeneration. *Int. J. Mol. Sci.* **2021**, *22*, 8344. [CrossRef]
31. Wang, S.Q.; Wang, Y.X.; Zou, S.Q. A Glance at the Molecules That Regulate Oligodendrocyte Myelination. *Curr. Issues Mol. Biol.* **2022**, *44*, 2194–2216. [CrossRef] [PubMed]
32. Graves, M.; Benton, M.; Lea, R.; Boyle, M.; Tajouri, L.; Macartney-Coxson, D.; Scott, R.; Lechner-Scott, J. Methylation differences at the HLA-DRB1 locus in CD4+ T-Cells are associated with multiple sclerosis. *Mult. Scler.* **2013**, *20*, 1033–1041. [CrossRef]
33. Ma, Q.; Caillier, S.J.; Muzic, S.; University of California San Francisco MS-EPIC Team; Wilson, M.R.; Henry, R.G.; Cree, B.A.C.; Hauser, S.L.; Didonna, A.; Oksenberg, J.R. Specific hypomethylation programs underpin B cell activation in early multiple sclerosis. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2111920118. [CrossRef]
34. Sabatino, J.J., Jr.; Zamvil, S.S.; Hauser, S.L. B-Cell Therapies in Multiple Sclerosis. *Cold Spring Harb. Perspect. Med.* **2019**, *9*, a032037. [CrossRef]
35. Bittner, S.; Ruck, T.; Wiendl, H.; Grauer, O.M.; Meuth, S.G. Targeting B cells in relapsing-remitting multiple sclerosis: From pathophysiology to optimal clinical management. *Ther. Adv. Neurol. Disord.* **2017**, *10*, 51–66. [CrossRef]
36. Parisi, L.; Gini, E.; Baci, D.; Tremolati, M.; Fanuli, M.; Bassani, B.; Farronato, G.; Bruno, A.; Mortara, L. Macrophage Polarization in Chronic Inflammatory Diseases: Killers or Builders? *J. Immunol. Res.* **2018**, *2018*, 8917804. [CrossRef]

37. Bar-Or, A.; Nuttall, R.K.; Duddy, M.; Alter, A.; Kim, H.J.; Ifergan, I.; Pennington, C.J.; Bourgoin, P.; Edwards, D.R.; Yong, V.W. Analyses of all matrix metalloproteinase members in leukocytes emphasize monocytes as major inflammatory mediators in multiple sclerosis. *Brain* **2003**, *126*, 2738–2749. [CrossRef] [PubMed]
38. D'Amico, E.; Zanghi, A.; Parrinello, N.L.; Romano, A.; Palumbo, G.A.; Chisari, C.G.; Toscano, S.; Raimondo, F.D.; Zappia, M.; Patti, F. Immunological Subsets Characterization in Newly Diagnosed Relapsing-Remitting Multiple Sclerosis. *Front. Immunol.* **2022**, *13*, 819136. [CrossRef] [PubMed]
39. Durelli, L.; Conti, L.; Clerico, M.; Boselli, D.; Contessa, G.; Ripellino, P.; Ferrero, B.; Eid, P.; Novelli, F. T-Helper 17 Cells Expand in Multiple Sclerosis and Are Inhibited by Interferon-beta. *Ann. Neurol.* **2009**, *65*, 499–509. [CrossRef]
40. Zhang, X.; Tao, Y.Z.; Chopra, M.; Dujmovic-Basuroski, I.; Jin, J.P.; Tang, Y.N.; Drulovic, J.; Markovic-Pese, S. IL-11 Induces Th17 Cell Responses in Patients with Early Relapsing-Remitting Multiple Sclerosis. *J. Immunol.* **2015**, *194*, 5139–5149. [CrossRef]
41. Pender, M.P.; Csurhes, P.A.; Pfluger, C.M.M.; Burrows, S.R. Deficiency of CD8(+) effector memory T cells is an early and persistent feature of multiple sclerosis. *Mult. Scler. J.* **2014**, *20*, 1825–1832. [CrossRef]
42. Caruana, P.; Lemmert, K.; Ribbons, K.; Lea, R.; Lechner-Scott, J. Natural killer cell subpopulations are associated with MRI activity in a relapsing-remitting multiple sclerosis patient cohort from Australia. *Mult. Scler. J.* **2017**, *23*, 1479–1487. [CrossRef] [PubMed]
43. Rock, K.L.; Reits, E.; Neefjes, J. Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends Immunol.* **2016**, *37*, 724–737. [CrossRef]
44. Salgado, F.J.; Lojo, J.; Fernandez-Alonso, C.M.; Vinuela, J.; Cordero, O.J.; Nogueira, M. Interleukin-dependent modulation of HLA-DR expression on CD4and CD8 activated T cells. *Immunol. Cell Biol.* **2002**, *80*, 138–147. [CrossRef] [PubMed]
45. Diniz, S.N.; da Silva, C.F.; de Almeida, I.T.; da Silva Costa, F.E.; de Oliveira, E.M.L. INFbeta treatment affects global DNA methylation in monocytes of patients with multiple sclerosis. *J. Neuroimmunol.* **2021**, *355*, 577563. [CrossRef] [PubMed]
46. Souren, N.Y.; Gerdes, L.A.; Lutsik, P.; Gasparoni, G.; Beltran, E.; Salhab, A.; Kumpfel, T.; Weichenhan, D.; Plass, C.; Hohlfeld, R.; et al. DNA methylation signatures of monozygotic twins clinically discordant for multiple sclerosis. *Nat. Commun.* **2019**, *10*, 2094. [CrossRef]
47. Thompson, A.J.; Banwell, B.L.; Barkhof, F.; Carroll, W.M.; Coetzee, T.; Comi, G.; Correale, J.; Fazekas, F.; Filippi, M.; Freedman, M.S.; et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol.* **2017**, *17*, 162–173. [CrossRef]
48. Campagna, M.P.; Xavier, A.; Lea, R.A.; Stankovich, J.; Maltby, V.E.; Butzkueven, H.; Lechner-Scott, J.; Scott, R.J.; Jokubaitis, V.G. Whole-blood methylation signatures are associated with and accurately classify multiple sclerosis disease severity. *Clin. Epigenetics* **2022**, *14*, 194. [CrossRef]
49. Yan, L.; Ma, C.; Wang, D.; Hu, Q.; Qin, M.; Conroy, J.M.; Sucheston, L.E.; Ambrosone, C.B.; Johnson, C.S.; Wang, J.; et al. OSAT: A tool for sample-to-batch allocations in genomics experiments. *BMC Genom.* **2012**, *13*, 689. [CrossRef]
50. Morris, T.J.; Butcher, L.M.; Feber, A.; Teschendorff, A.E.; Chakravarthy, A.R.; Wojdacz, T.K.; Beck, S. ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* **2014**, *30*, 428–430. [CrossRef]
51. Tian, Y.; Morris, T.J.; Webster, A.P.; Yang, Z.; Beck, S.; Feber, A.; Teschendorff, A.E. ChAMP: Updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics* **2017**, *33*, 3982–3984. [CrossRef]
52. Teschendorff, A.E.; Marabita, F.; Lechner, M.; Bartlett, T.; Tegner, J.; Gomez-Cabrero, D.; Beck, S. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **2013**, *29*, 189–196. [CrossRef] [PubMed]
53. Leek, J.T.; Johnson, W.E.; Parker, H.S.; Jaffe, A.E.; Storey, J.D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **2012**, *28*, 882–883. [CrossRef]
54. Guo, Y.; He, J.; Zhao, S.; Wu, H.; Zhong, X.; Sheng, Q.; Samuels, D.C.; Shyr, Y.; Long, J. Illumina human exome genotyping array clustering and quality control. *Nat. Protoc.* **2014**, *9*, 2643–2662. [CrossRef] [PubMed]
55. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.; Daly, M.J.; et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [CrossRef] [PubMed]
56. Zivkovic, M.; Stankovic, A.; Dincic, E.; Popovic, M.; Popovic, S.; Raicevic, R.; Alavantic, D. The tag SNP for HLA-DRB1*1501, rs3135388, is significantly associated with multiple sclerosis susceptibility: Cost-effective high-throughput detection by real-time PCR. *Clin. Chim. Acta* **2009**, *406*, 27–30. [CrossRef]
57. Sing, T.; Sander, O.; Beerenwinkel, N.; Lengauer, T. ROCR: Visualizing classifier performance in R. *Bioinformatics* **2005**, *21*, 3940–3941. [CrossRef]
58. Teschendorff, A.E.; Breeze, C.E.; Zheng, S.J.C.; Beck, S. A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. *BMC Bioinform.* **2017**, *18*, 105. [CrossRef]
59. Edgar, R.; Domrachev, M.; Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, *30*, 207–210. [CrossRef]
60. Dunning, M.J.; Smith, M.L.; Ritchie, M.E.; Tavare, S. beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics* **2007**, *23*, 2183–2184. [CrossRef]
61. Aryee, M.J.; Jaffe, A.E.; Corrada-Bravo, H.; Ladd-Acosta, C.; Feinberg, A.P.; Hansen, K.D.; Irizarry, R.A. Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **2014**, *30*, 1363–1369. [CrossRef] [PubMed]

62. Marabita, F.; Almgren, M.; Lindholm, M.E.; Ruhrmann, S.; Fagerstrom-Billai, F.; Jagodic, M.; Sundberg, C.J.; Ekstrom, T.J.; Teschendorff, A.E.; Tegner, J.; et al. An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform. *Epigenetics* **2013**, *8*, 333–346. [CrossRef] [PubMed]
63. Houseman, E.A.; Accomando, W.P.; Koestler, D.C.; Christensen, B.C.; Marsit, C.J.; Nelson, H.H.; Wiencke, J.K.; Kelsey, K.T. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinform.* **2012**, *13*, 86. [CrossRef]
64. Pidsley, R.; CC, Y.W.; Volta, M.; Lunnon, K.; Mill, J.; Schalkwyk, L.C. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genom.* **2013**, *14*, 293. [CrossRef] [PubMed]
65. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47. [CrossRef] [PubMed]
66. Fabregat, A.; Sidiropoulos, K.; Viteri, G.; Forner, O.; Marin-Garcia, P.; Arnau, V.; D'Eustachio, P.; Stein, L.; Hermjakob, H. Reactome pathway analysis: A high-performance in-memory approach. *BMC Bioinform.* **2017**, *18*, 142. [CrossRef] [PubMed]
67. Orchard, S.; Ammari, M.; Aranda, B.; Breuza, L.; Briganti, L.; Broackes-Carter, F.; Campbell, N.H.; Chavali, G.; Chen, C.; del-Toro, N.; et al. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **2013**, *42*, D358–D363. [CrossRef] [PubMed]
68. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*, 2nd ed.; Springer: New York, NY, USA, 2016; ISBN 978-3-319-24277-4.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.