Int J Clin Exp Med 2025;18(11):272-286
www.ijcem.com /ISSN:1940-5901/IJCEM0167342

Original Article

Integrated single cell RNA sequencing and proteome-wide Mendelian randomization identify therapeutic targets for multiple sclerosis

Changjiang Zhang[1], Ao Wang[2], Ke Yi[1,2]

[1]*Department of Cardiology, Hubei Provincial Key Laboratory of Occurrence and Intervention of Rheumatic Diseases/Hubei Provincial Clinical Research Center for Nephrology, Minda Hospital of Hubei Minzu University, Hubei Minzu University, Enshi, Hubei, China;* [2]*Key Laboratory of Obstetrics and Gynecologic and Pediatric Diseases and Birth Defects of Ministry of Education, West China Second University Hospital, Sichuan University, Chengdu, Sichuan, China*

Received July 9, 2025; Accepted October 28, 2025; Epub November 15, 2025; Published November 30, 2025

Abstract: Objectives: Multiple sclerosis (MS) is a chronic inflammatory demyelinating illness of the CNS that requires novel therapeutic targets and safety evaluations. This study aims to identify novel therapeutic protein targets for multiple sclerosis and evaluate possible side effects of druggable proteins. Methods: Causal links between MS risk and plasma proteins were explored by conducting a comprehensive proteome-wide MR analysis. The proteomic data were taken from the UK Biobank Pharma Proteomics Project (UKBPPP) and included 2,940 plasma proteins. Genetic data for MS were obtained from the Finnish database and the International Multiple Sclerosis Genetics Consortium (IMSGC). Colocalisation analysis was performed to identify shared causal variants between MS and plasma proteins. A phenome-wide association study (PheWAS) using the Finnish database was conducted to evaluate the potential side effects of the druggable proteins. Single-cell RNA sequencing (scRNA-seq) data from the GSE138266 dataset was also used to explore gene expression patterns across different cell types. Results: Six plasma proteins were identified with significant genetic correlations to an increased risk of MS and it was found that five of them - AIF1, AGER, TNFRSF14, CD58 and EVI5 - shared genetic variants with MS. This suggests their potential as direct therapeutic targets. PheWAS indicated the ability for these proteins to have adverse effects on other phenotypes (Pfdr < 0.05). ROC analysis of gene expression data from the GEO datasets (GSE131282, GSE21942) showed these proteins to have high diagnostic value, with AGER, AIF1, EVI5 and TNFRSF14 all displaying AUC values that were greater than 0.7. Single-cell RNA sequencing analysis revealed differential expressions of these key genes across various immune cell types, thereby highlighting their involvement in MS immunoregulation. Conclusions: This study explored the causal connections between MS and five plasma proteins, identifying possible treatment targets.

Keywords: Multiple sclerosis, plasma proteins, phenome-wide association study, Mendelian randomisation, single-cell RNA sequencing

Introduction

Multiple sclerosis (MS) is a chronic autoimmune disorder of the CNS marked by axonal loss, inflammation, and demyelination [1]. MS often manifests in adults aged between 20 and 40 and women are 2-3 times more frequently affected than men - this trend appears to be increasing on a global scale [2, 3]. The global median prevalence of the disorder is 33 per 100,000 individuals, but there are notable regional variations. Europe and North America have the highest prevalence rates with respective cases of 108 and 140 per 100,000 people, while sub-Saharan Africa and Asia have respective rates of 2.1 and 2.2 per 100,000 people, which are the lowest [4].

The diagnosis of MS is currently based on the 2017 McDonald criteria, which emphasize dissemination in space and time of CNS lesions on clinical and MRI evaluation, supplemented by cerebrospinal fluid analysis when required [5]. Clinically, MS typically presents with optic neuri-

https://doi.org/10.62347/LLNA4571

tis, motor and sensory deficits, ataxia, fatigue, and cognitive impairment, reflecting the multifocal nature of CNS demyelination [6, 7].

The pathophysiologic mechanisms of MS involve a multifaceted interaction between genetic, environmental and immunological factors [8]. The disruption of the blood-brain barrier (BBB) is central to MS pathology and this enables the penetration of the CNS by immune cells including T and B lymphocytes. Inflammatory responses and demyelination are then initiated by them [9, 10]. Microglia and macrophages are resident immune cells in the CNS and they are essential for lesion formation and the neurodegeneration procedure as they release pro-inflammatory cytokines, ROS and glutamate. These are contributors to axonal damage and neuronal death [11, 12].

Existing therapeutic options for MS focus mostly on immunomodulation and immunosuppression and they have served to improve disease management through the reduction of the severity and frequency of relapses and slowing disease development [13]. Current disease-modifying therapies (DMTs), including interferon-β, glatiramer acetate, oral agents (e.g., fingolimod, dimethyl fumarate, teriflunomide), and monoclonal antibodies (e.g., natalizumab, ocrelizumab, alemtuzumab), can reduce relapse rates and delay disability progression, particularly in relapsing-remitting MS [3, 14, 15]. Nevertheless, their efficacy in progressive MS remains limited, long-term use is associated with substantial adverse effects, and they mainly target inflammatory processes without adequately addressing neurodegeneration [16, 17]. As a result, MS remains incurable, and prognosis varies considerably depending on disease subtype and treatment response [18].

MR (Mendelian randomisation) analysis has gained traction recently as a powerful method for finding new therapeutic targets and repurposing existing drugs [19]. Specific single nucleotide polymorphisms (SNPs) that are located on chromosomes are instrumental in protein expression control and have been identified by GWAS (genome-wide association studies). These SNPs correspond to protein expression levels and are known as pQTLs (protein quantitative trait loci) [20]. MR leverages these pQTLs as essential variables, allowing researchers to examine the cause-and-effect links that exist between exposures and outcomes. This is invaluable in terms of new drug target and biomarker identification [21, 22].

MR provides a more reliable assessment of causality than traditional observational studies by mitigating the influence of confounding variables [23]. The integration of phenome-wide association studies (PheWAS) also enables the prediction of adverse reactions related to these targets [24]. Plasma proteins that are essential for several processes such as transport, signalling, repair, growth and immune responses often become dysregulated in various diseases, which makes them prime drug development candidates. While pursuing new MS treatments, a thorough proteome-wide MR study was conducted to identify promising therapeutic targets [25].

The approach began with a two-sample MR study to evaluate a causative role played by plasma proteins in MS. The credibility of the results was then confirmed by colocalization analyses. Finally, the potential adverse effects of the known proteins that could be targeted for MS treatment were assessed by utilising PheWAS.

Methods

Approval of research ethics and study blueprint

This study collected comprehensive GWAS summary statistics from original investigations, all of which were conducted with the informed consent of participants. Aggregated statistical data were used, which negated the need for any additional ethical approval.

Inclusion criteria were as follows: (1) plasma protein pQTLs reaching genome-wide significance ($P < 5 \times 10^{-8}$); (2) cis-pQTLs located within ±1 Mb of the corresponding gene region; (3) independent SNPs selected using LD clumping ($r^2 < 0.001$, 10,000 kb); (4) instrumental variables with F-statistics > 10 to minimize weak instrument bias; (5) GWAS summary data for multiple sclerosis obtained from large, well-characterized cohorts (FinnGen R10 and IMSGC).

Exclusion criteria included: (1) trans-pQTLs located outside the ±1 Mb window of the target gene; (2) SNPs failing quality control (e.g., low imputation quality, ambiguous alleles); (3) plasma proteins or phenotypes without valid instrumental variables; (4) duplicated samples or datasets without clear case-control classification.

Characterisation of plasma protein QTLs

pQTLs data was taken from the UK Biobank-PPP and this included proteomic data from 54,219 individuals (https://www.synapse.org/#!Synapse:syn51365303). Meticulous mapping was performed for 2,940 proteins [26]. pQTLs are typically situated near their respective genes, which are known as "cis-pQTLs", which suggests a regulatory effect via the proximate gene [27]. In contrast, "trans-pQTLs" are located further away or on different chromosomes and are presumed to exert effects through other genes. The distinction between cis- and trans-pQTLs often hinges on set distance thresholds, which are commonly 500 kb or 1,000 kb [28-30].

For this study, specific selection criteria for pQTLs were applied as follows: (1) only cis-pQTLs within ±1 Mb of the gene region were considered to ensure biological plausibility; (2) SNPs were required to achieve genome-wide significance ($P < 5 \times 10^{-8}$); (3) independent variants were selected using LD clumping with $r^2 < 0.001$ and a genetic distance of 10,000 kb; (4) instrumental variables with F-statistics > 10 were retained to minimize weak instrument bias [19].

Genetic association data for multiple sclerosis

Genetic data for multiple sclerosis was taken from the R10 release of the Finnish database (https://r10.finngen.fi/). This dataset included 488 patients and 364,071 controls [31]. The dataset of the IMSGC (International Multiple Sclerosis Genetics Consortium) included 68,374 control participants and 47,429 MS patients [32].

Conducting MR analysis

AS was used as the result and plasma proteins as the exposure for a two-sample MR analysis. The pQTLs were chosen based on the previously specified criteria. R package "TwoSampleMR" (version 0.6.0) was used for analysis, the IVW method was applied for multiple SNPs and the Wald ratio was used for single SNP scenarios [19]. *P*-values were adjusted for numerous tests using the false discovery rate (FDR) approach - Pfdr < 0.05 denoted statistical significance.

Analysis of colocalisation

The aim of colocalization analysis is ascertaining shared genetic variants that influence exposure and outcome, further corroborating MR findings. For proteins that have positive MR outcomes, SNPs within a ±1 MB window of the gene (cis-pQTLs) were scrutinised for their colocalization with multiple sclerosis [33]. Five hypotheses were tested in colocalization analysis in order to determine whether SNPs are related to the protein, the disease, both or neither. The focus of this study was on genes with a combined posterior probability (PPH3+PPH4) ≥ 0.8 due to the limited power of colocalization analyses [34].

Phenome-wide association analysis

PheWAS is the inverse of GWAS and is used to discern links between phenotypes or SNPs and a broad spectrum of phenotypic traits [13]. This methodology is particularly useful for identifying drug target side effects [22]. The focus of this study was on plasma proteins with positive MR outcomes as the exposure and the same criteria were maintained for instrumental variables. The outcome used phenotypic data from the Finnish database R10 version, which included 2,272 phenotypes across 46 categories, to conduct a comprehensive phenome-wide MR analysis. A Pfdr < 0.05 was an indicator of statistical significance.

GEO data download

The datasets (GSE131282, GSE21942) were acquired from the NCBI website. GSE131282 had 184 grey matter, which included 42 healthy controls and 142 MS patients [35]. The GSE21942 dataset consisted of 12 patients with peripheral blood samples and 15 healthy controls [36].

Receiver Operating Characteristic (ROC) analysis

The "pROC" package was used to assess the diagnostic values and learn the possible clinical relevance of important genes. The area under the ROC curve (AUC) is an essential part of this analysis and this was calculated meticulously. The training dataset was GSE131282 and the testing dataset was GSE21942. AUC values of greater than 0.7 represented high diagnostic significance.

Quality control of single-cell datasets

All multiple sclerosis ("Multiple Sclerosis, multiple sclerosis, multiple sclerosis, multiple sclerosis, single-cell (scRNA-seq) dataset GSE-138266. MS") from the expression of the matrix was imported by using the "CreateSeuratObject" function of R package Seurat (Version 4.3.0) to create the Seurat object of the samples. Parameters were set to genes that were expressed in at least 3 cells and at least 200 genes expressed in each cell. Use the nCount RNA (Unique Molecular Identifier, UMI) is more than 500, greater than 200 nFeature RNA, Log10 (genes per UMI) is greater than 0.8, mitochondrial Genes accounted for (mitoRatio) is less than 0.05 for the filter, removing low quality of the cell. After quality control was performed, the cells were statistically analysed to calculate changes in the number of cells both before and after filtration. The sequencing depth of the scRNA-seq dataset GSE138266 was normalised by the "NormalizeData" function. The normalisation method was "LogNormalize". The "vst" method of the "FindVariableFeatures" function was used to identify the hypervariable genes in the dataset. The "ScaleData" function of the data was then used to zoom in and rule out the influence of sequencing depth.

Double cells are due to the experiment, in the unicellular microfluidic process such as a droplet containing two or more cells, and in the same cell in the subsequent analysis of droplet Cell with the same Barcode, which can be considered as a pseudo cell. The main characteristic of these pseudo-cells is that the number of UMI and genes that is detected is often two or more times more than that of normal cells. They also may carry classic marker genes of different cell types, thereby hindering cell type identification. The "scDblFinder" function of R package (Version 1.12.0) was used to evaluate each cell score of the double cells and identify whether for the cells and will double filtering cells.

Principal component analysis (PCA) was used to find the significant principal component (PC) and the "Elbowplot" function was used to visualise the standard deviation distribution. 40 PCs were chosen for Uniform Manifold Approximation and Projection (UMAP) analysis for dimension reduction. The "FindNeighbors" function was used for creating the k-nearest neighbours of the Euclidean distance in the base PCA space with 40 principal component (PC) dimension parameters.

Cell clustering and cell type annotation

R package Seurat (Version 4.3.0) was used to import the Seurat objects of the data quality and standardised single-cell Rna-Seq (scRNA-seq) dataset GSE138266. For possible batch effect between samples, using R package harmony (Version while) "RunHarmony" function in data integration and effect to batch processing, and under the resolution of 0.8 cells can be divided into different clustering.

By ScType algorithm, using specific marker genes from single-cell transcriptome data automatic cell type identification. Data sets of single-celled (scRNA - seq) cell type annotation, identification of cell types. The stacked bar chart was used to display the cell proportion of each cell type in all the samples and the relationship between cell type and clustering was then analysed.

Expression of key genes in each cell type

The expression differences of five key genes (AIF1, AGER, TNFRSF14, CD58 and EVI5) in various cell types in the single-cell Rna-Seq dataset GSE138266 were explored. The FeaturePlot function Feature the Plot was used for mapping, and violin plots were used to show Key Genes' (Key Genes) expression level in different cell types.

Observation indicators and evaluation methods

The observation indicators of this study and their evaluation methods were as follows: (1) Proteome-wide MR analysis - the primary indicator was the causal association between plas-

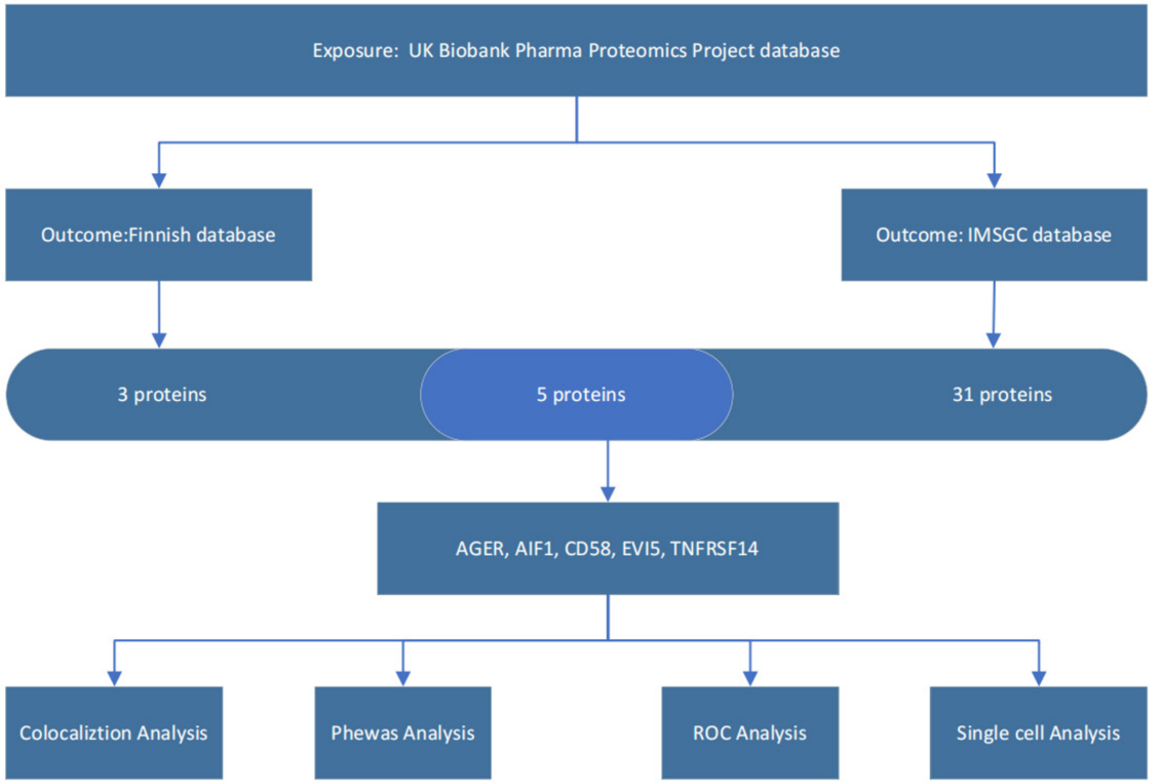


Figure 1. Flowchart of the study design.

ma protein levels and MS risk, assessed using a two-sample MR framework. The IVW method was applied when multiple SNPs were available, and the Wald ratio was used for single SNPs. (2) Colocalization analysis - shared causal variants between plasma proteins and MS were evaluated using Bayesian colocalization, with PPH3+PPH4 ≥ 0.8 considered as strong evidence. (3) Phenome-wide association study (PheWAS) - associations of identified proteins with 2,272 phenotypes from the FinnGen database were examined, with statistical significance defined as Pfdr < 0.05 after FDR correction. (4) Validation in GEO datasets - ROC curve analysis (using the "pROC" package) was performed to evaluate the diagnostic performance of key genes in MS. The AUC value was used as the primary diagnostic indicator, with GSE131282 as the training set and GSE21942 as the testing set. (5) Single-cell RNA sequencing analysis - the observation indicators included cell-type-specific expression patterns of the five key genes (AIF1, AGER, TNFRSF14, CD58, and EVI5). These were evaluated through clustering, cell type annotation, and visualizations including FeaturePlot, violin plots, and bubble charts.

Results

The schematic plot of the project is shown in **Figure 1**.

Plasma proteins related to multiple sclerosis

After the stringent instrumental variables screening criteria of the study were applied, 1,280 plasma proteins were subsequently included in the MR analysis of the Finnish database. Eight plasma proteins (Supplementary Table 1) were found to have causal relationships with MS in the MR analysis based on Wald or IVW ratio outcomes (Pfdr < 0.05) in the subset of 1,280 plasma proteins. 36 plasma proteins related to multiple sclerosis were found in the IMSGC database by MR analysis based on Wald or IVW ratio outcomes (Pfdr < 0.05) (Supplementary Table 2). The results from the Finnish database and the IMSGC dataset were intersected for determining proteins that were linked to multiple sclerosis in both

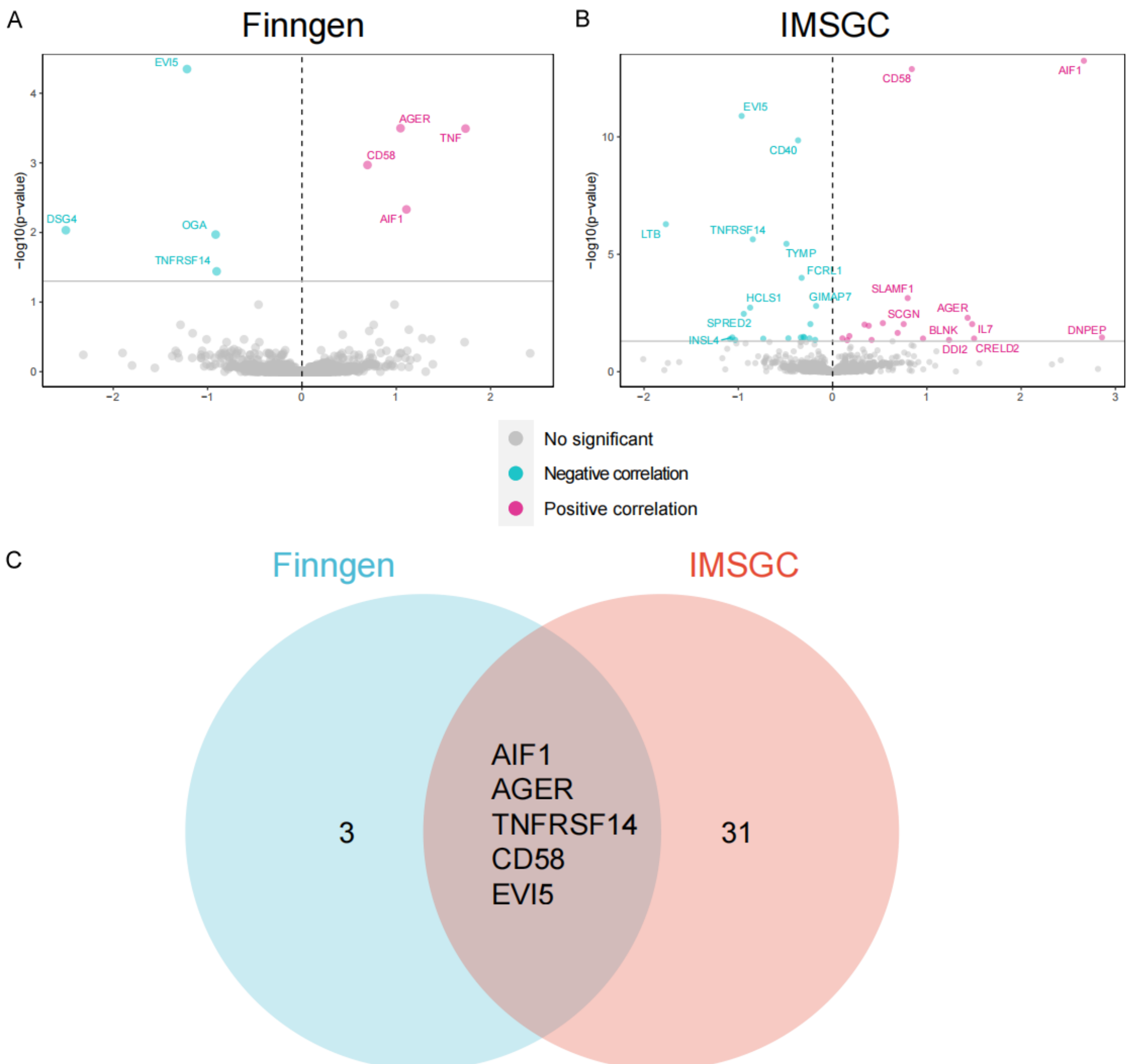


Figure 2. Volcano plot of MR results: Causal relationship between plasma proteins and multiple sclerosis (A) Finngen database, (B) IMSGC database. (C) Venn diagram: Intersection of plasma proteins in two datasets. IMSGC: International Multiple Sclerosis Genetics Consortium.

datasets. Five plasma proteins are shown in the Venn diagram: AIF1 (allograft inflammatory factor 1), AGER (advanced glycosylation end product-specific receptor), TNFRSF14 (tumour necrosis factor receptor superfamily member 14), CD58 (CD58 molecule) and EVI5 (ecotropic viral integration site 5). The following are the odds ratios and 95% confidence intervals for these plasma proteins in the Finnish database: AGER 2.84 (95% CI: 1.91-4.23), AIF1 3.03 (95% CI: 1.81-5.05), CD58 2.00 (95% CI: 1.51-2.66), EVI5 0.30 (95% CI: 0.19-0.46) and TNFRSF14 0.40 (95% CI: 0.25-0.66). The following are the odds ratios and 95% confidence intervals for the plasma proteins in the IMSGC database: AGER 4.19 (95% CI: 2.14-8.20), AIF1 10.72 (95% CI: 8.00-14.37), CD58 2.27 (95% CI: 1.73-2.97), EVI5 0.38 (95% CI: 0.30-0.49) and TNFRSF14 0.43 (95% CI: 0.32-0.57). More information is shown in **Figures 2** and **3**.

Sensitivity analysis for plasma proteins related to MS

Gene colocalization analysis was performed for these five plasma proteins within a range of ±1 MB upstream and downstream of their respective genes as a means of investigating potential connections with MS. From the findings, it was suggested that a causative mutation in

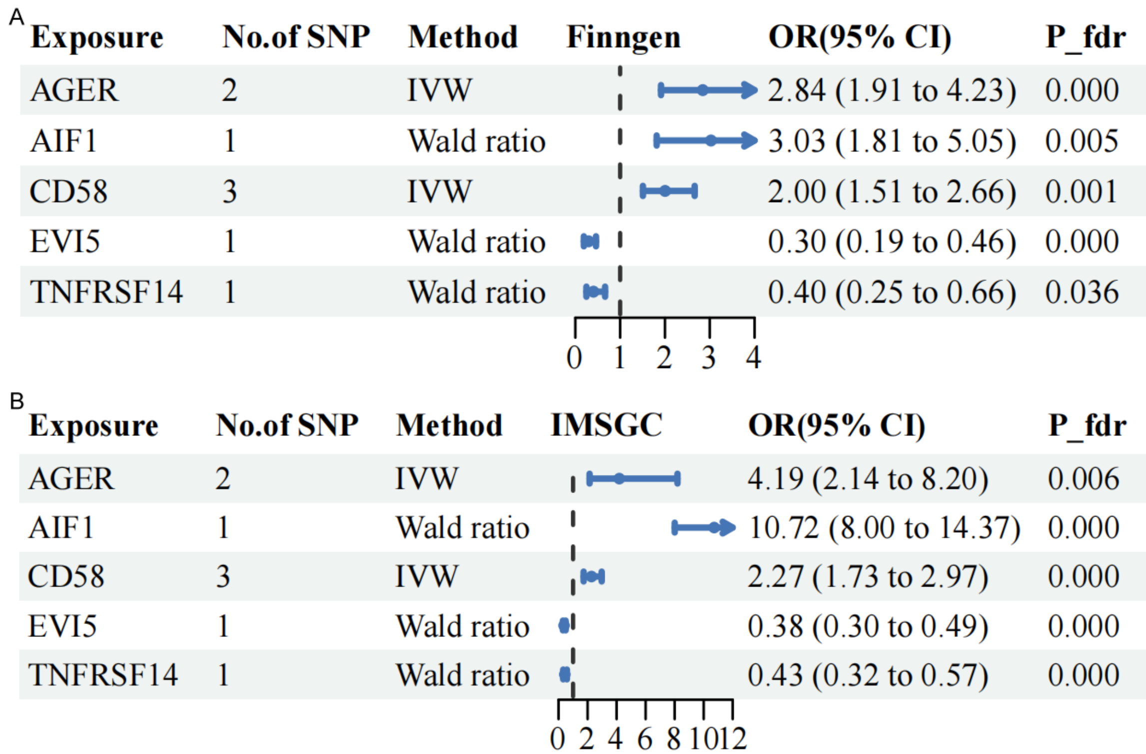


Figure 3. Forest plot of the MR results: Effects of 5 plasma proteins on multiple sclerosis (A) Finngen database, (B) IMSGC database. CI: confidence interval; OR: odds ratio; IMSGC: International Multiple Sclerosis Genetics Consortium.

this region (PPH3+PPH4 > 0.8) could be shared by all potential plasma proteins. Comprehensive details are presented in Supplementary Table 3. This implies that all five plasma proteins may be useful therapeutic targets for MS.

Phenome-wide association analysis for 5 plasma proteins related to MS

A phenome-wide association analysis was conducted and 2,408 phenotypes were screened across 45 categories from the Finnish database (version R10) to assess the potential positive or negative impacts of the five plasma proteins that were linked to MS on other phenotypes. Noteworthy causal relationships were observed between AGER and 53 phenotypes (Pfdr < 0.05), AIF1 and 120 phenotypes (Pfdr < 0.05), CD58 and 2 phenotypes (Pfdr < 0.05), EVI5 and 7 phenotypes (Pfdr < 0.05) and TNFRSF14 and 19 phenotypes (Pfdr < 0.05). More details are provided in **Figure 4**. These phenotypes may have harmful consequences on the target protein or they could serve as therapeutic targets.

Validate the role of plasma proteins in MS

The GEO dataset was used for validating the role of these protein-coding genes in MS. The GSE131282 dataset was used as the training dataset and these results were obtained from ROC analysis: AGER (AUC = 0.812), AIF1 (AUC = 0.626), CD58 (AUC = 0.723), EVI5 (AUC = 0.765) and TNFRSF14 (AUC = 0.777). The GSE21942 dataset was used as a test dataset for the validation of the diagnostic value of these five plasma proteins and the following results were yielded: AGER (AUC = 0.879), AIF1 (AUC = 0.795), CD58 (AUC = 0.438), EVI5 (AUC = 0.769) and TNFRSF14 (AUC = 0.717) (**Figure 5**).

Quality control of single-cell datasets

The data quality had to be integrated and controlled in order to guarantee that the following analysis was founded on the best quality data. The expression matrices of all multiple sclerosis (MS) samples in the single-cell (scRNA-seq) dataset GSE138266 were imported and then

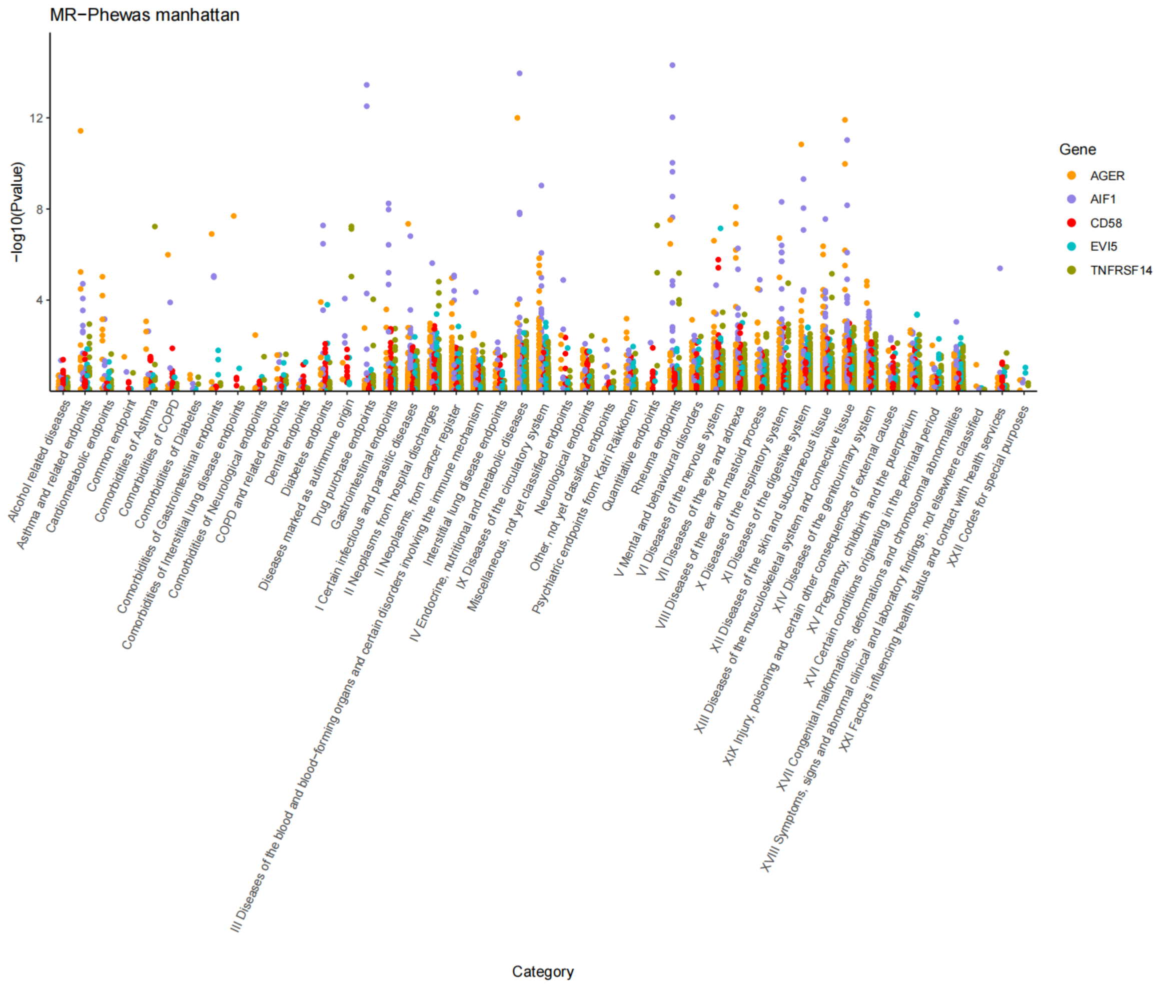


Figure 4. Manhattan plot of result of PheWAS analysis of associations between 5 plasma proteins and other disease outcomes.

created as Seurat objects using the "Create-SeuratObject" function of R package Seurat. The parameter was set to at least three cells in each cell of the expression of genes and at least 200 gene expressions.

Using nCount RNA (UMI) is greater than 500, greater than 200 nFeature RNA, Log_{10} (Genes per UMI) is greater than 0.8, (mitoRatio) is less than 0.05 of mitochondrial Genes accounted for the filter, removing low quality of the cell. Map nCount RNA (UMI) violin shows a single cell detected the number of RNA sequencing Reads (Supplementary Figure 1A), which is used to measure RNA expression level in every cell. The higher the cell number of Reads, the higher the RNA expression level in that cell. The nFeature RNA violin plot shows the number of unique genes found in a single cell (Supplementary Figure 1B) and the diversity of genes expressed in each single-cell sample. The more unique genes detected in a cell, the richer the RNA sequencing data for that cell was. The Log_{10} (genes per UMI) violin diagram shows each UMI detected gene (Supplementary Figure 1C). UMI is used for distinguishing the same sequence in the RNA sequencing of identifier, Log_{10} (Genes per UMI) value is higher, on behalf of the detected under the same number of UMI gene number, the more reflect the diversity of gene expression in cells. Map (mitoRatio) of mitochondrial genes accounted the violin show mitochondrial genes accounted for (Supplementary Figure 1D), mitochondrial genes can affect the accuracy of the data in the cells. The density diagram shows the distribution of the above data (Supplementary Figure

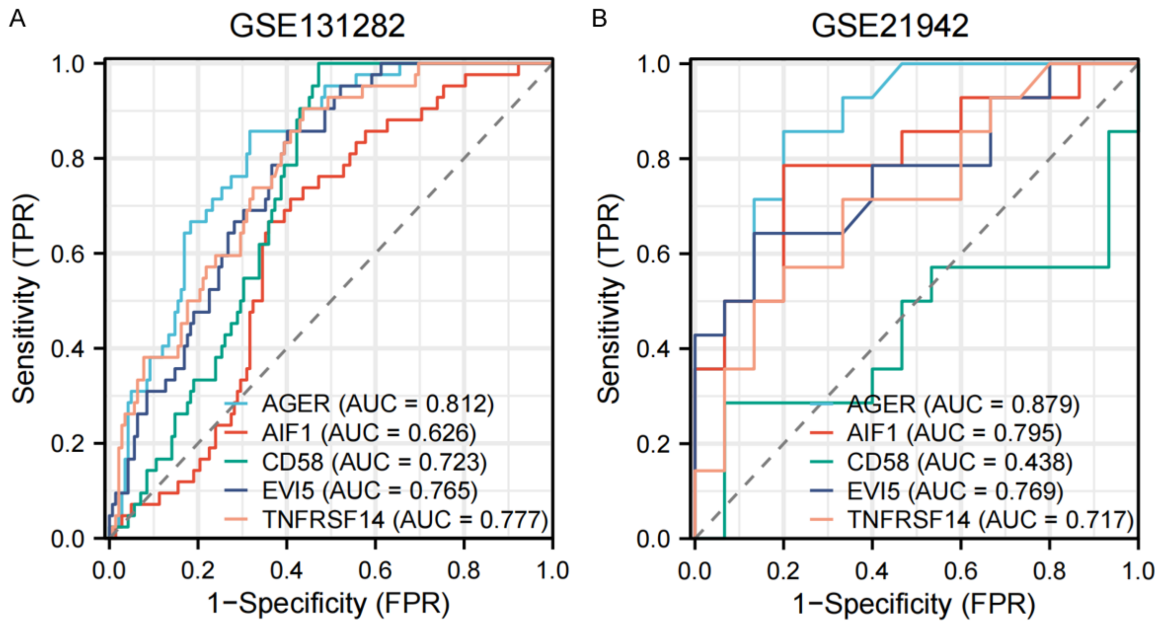


Figure 5. ROC validation of the diagnostic effects in multiple sclerosis dataset. A. Training dataset: GSE131282. B. Testing dataset: GSE21942.

2A-D). Through double filtering, data integration and quality control, 19545 cells were selected

Through the "NormalizeData" function of single-cell (scRNA - seq) data set GSE138266 sequencing depth for standardization, standardization of methods for "LogNormalize", The "vst" method of the "FindVariableFeatures" function was employed for identifying the hypervariable genes in the dataset. Principal component analysis (PCA) found important PCs and 40 of them were then chosen for UMAP analysis for dimensionality reduction. The "FindNeighbors" function was used to create the k-nearest neighbours of the Euclidean distance in the base PCA space with 40 principal component (PC) dimension parameters.

Cell clustering and cell type annotation

The "RunHarmony" function in R package harmony single-celled (scRNA - seq) GSE138266 was used for data integration and data sets to processing of the batch effect, and for each sample ' clustering mapping (Supplementary Figure 3A). Then, under the resolution of 0.8, cells could be divided into different clustering. Each sample cell's clustering was mapped out (Supplementary Figure 3B). The results show that 19,545 cells were divided into 22 clusters under the resolution of 0.8 and they were identified as being a total of 15 cell types by the ScType algorithm (Supplementary Figure 3C). These cell types were classic mononuclear cells (classical monocytes), initial CD4 + T cells (naive CD4 + T cells), initial B cells (Naive B cells), effect of CD4 + T cells (effector CD4 + T cells), plasma cells, dendritic cells (plasmacytoid dendritic cells), platelets, natural killer cells, effect of CD8 + T cells (effector CD8 + T cells), initial CD8 + T cells (naive CD8 + T cells), atypical mononuclear cells (non-classical monocytes), CD8 + sample of NKT cells (CD8 + NKT-like cells), myeloid dendritic cells, plasma B cells, and progenitor cells. The 15 cell types are presented in the stacked bar chart (Supplementary Figure 3D).

Key gene expression of various cell types

"FeaturePlot" function was used to verify five Key Genes (AIF1, AGER, TNFRSF14, CD58, EVI5) of GSE138266 single-celled data sets with 15 kinds of cell types in the expression Feature Plot (**Figure 6**). Then, the bubble chart was made to show the Key Genes expression in the 15 types of cells (**Figure 7**).

Discussion

This study used Mendelian randomisation analysis as a means of identifying significant

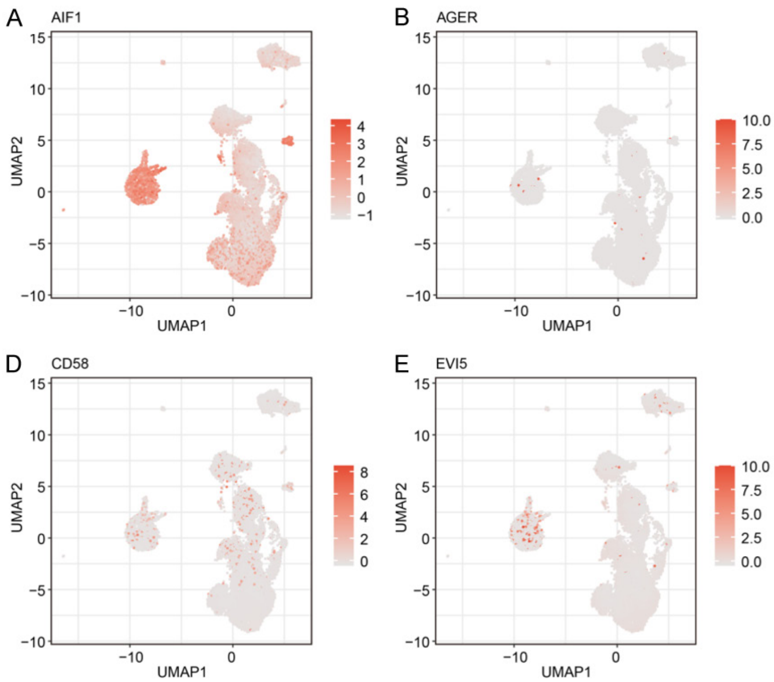


Figure 6. Feature plot of Key Genes expression in different cell types: Expression of Key Genes AIF1 (A), AGER (B), TNFRSF14 (C), CD58 (D), EVI5 (E) in 15 cell types in scRNA-seq dataset GSE138266 Feature Plot. UMAP: Uniform Manifold Approximation and Projection.

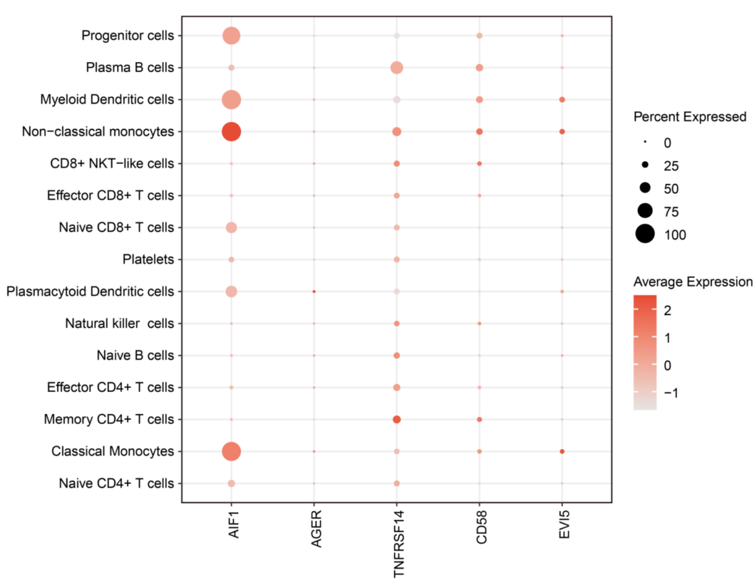


Figure 7. Bubble plot of expression of Key Genes in 15 cell types in scRNA-seq dataset GSE138266. The abscissa represents the Key Genes, the ordinate represents the cell type, the bubble size represents the percentage of expression, and the bubble color represents the average expression level.

associations between five key plasma proteins (AGER, AIF1, CD58, EVI5 and TNFRSF14) and MS. Strong causal relationships were found between these proteins and MS in two independent datasets, the Finnish and IMSGC databases. Their therapeutic target potential was further substantiated by sensitivity analyses and colocalization studies. These findings provide new perspective on the molecular processes underlying MS and imply that these plasma proteins may become therapeutic targets or novel biomarkers for the disease. The robust statistical significance of these findings highlights their clinical relevance while also paving the way for future research and therapeutic development.

Our findings can be compared with those of a recent MR study that identified potential drug targets for MS using plasma and CSF proteins (FCRL3, TYMP, AHSG, MMEL1, SLAMF7, and CD5L) [37]. In that report, several proteins such as FCRL3, TYMP, and SLAMF7 demonstrated protective associations with MS, while MMEL1 increased MS risk. In contrast, our study highlighted different targets (AGER, AIF1, CD58, EVI5, and TNFRSF14), that were consistently supported across both the Finnish and IMSGC cohorts. The differences may arise from distinct proteomic datasets, the inclusion of scRNA-seq data in our study, and variations in analytical design. Importantly, the convergence of findings from both studies underscores that multiple immune-related proteins may play causal roles in MS pathogenesis, suggesting that complementary therapeutic strategies could be explored.

Furthermore, while the previous study incorporated CSF proteins, our study focused on plasma proteins combined with single-cell transcriptomic validation, thereby providing additional insight into the cell-type-specific expression patterns of candidate targets. This integrative approach highlights not only their genetic associations but also their functional relevance in immune regulation at the cellular level, which may enhance the translational potential of our findings.

AGER is a multi-ligand receptor that has been implicated in a variety of pathologic conditions by the role it plays in mediating inflammatory responses [38, 39]. In the context of MS, AGER binds to AGEs (advanced glycation end products), which triggers a cascade of pro-inflammatory signals that activate NF-κB and produce ROS and cytokines [40, 41]. This process contributes to the chronic inflammation and neuronal damage characteristic of MS, thereby promoting the autoimmune destruction of myelin sheaths in the CNS [42, 43]. The significant association between AGER and MS suggests its use as a diagnostic biomarker, which could enable the early detection and better monitoring of the disease. Elevated levels of AGER or its ligands could be indicators of heightened disease activity, which may guide therapeutic intervention.

AIF1 is a 17-kDa protein predominantly expressed in microglia and macrophages that is upregulated in a variety of inflammatory conditions. In functional terms, AIF1 enhances the activation, proliferation and migration of macrophages and T cells, thereby contributing to a pro-inflammatory environment. In EAE (experimental autoimmune encephalomyelitis), which is a mouse model for MS, AIF1-deficient mice exhibit reduced disease severity, marked by lower CNS leukocyte infiltration, demyelination and pro-inflammatory cytokine production [44, 45]. This suggests that AIF1 promotes encephalitogenic CD4 T cell activation and expansion, facilitating their infiltration into the CNS and exacerbating inflammation. In clinical terms, targeting AIF1 could provide a therapeutic avenue for MS. The inflammatory response could be reduced, T-cell proliferation could be limited and microglial activation could be decreased by inhibiting AIF1 expression or function [46]. This may lead to the development of diagnostic markers or treatment strategies that modulate AIF1 activity, improving MS patient outcomes.

The cell adhesion molecule CD58, which is more commonly referred to as lymphocyte function-associated antigen 3 (LFA-3), is mostly expressed in B cells and antigen-presenting cells (APCs). Its main function is to bind to CD2 on T cells, which strengthens the adhesion between T cells and APCs and enhances T cell activation [47]. The CD58 gene is associated with protective and risk alleles with MS. Higher CD58 expression is linked to the protective allele and this increases the regulatory T cell (Treg) function that is essential for the maintenance of immunological tolerance and reducing autoimmune responses [48]. Conversely, the risk allele causes decreased CD58 expression, thereby impairing Treg function and contributing to the autoimmune attack on myelin sheaths in the CNS [49].

EVI5 is a protein involved in the regulation of cell cycle progression and mitosis that particularly influences the transition from the G2 phase to mitosis [50]. Recent studies have highlighted its role in immune cell function, specifically in T cell activation and proliferation [51, 52]. Genetic variations in EVI5 have been linked to a growing autoimmune disease risk, suggesting that it plays a significant role in the immune response [53]. EVI5 influences the pathogenesis of MS by modulating T cell function and cytokine production, contributing to the autoimmune attack on myelin sheaths in the CNS [54]. This results in the inflammation and neurodegeneration characteristic of MS. Investigations have suggested that there is a link between polymorphisms in the EVI5 gene and higher MS susceptibility, thereby reinforcing its possible use as a biomarker for the diagnosis and monitoring of the disorder [55, 56].

The TNFRSF14 protein is also referred to as herpes virus entry mediator (HVEM) and this is essential for immune response modulation. TNFRSF14 interacts with a variety of ligands, which include LIGHT and BTLA, to balance stimulatory and inhibitory signals in T cells [57]. This is essential for the prevention of excessive immune responses and the maintenance of immune homeostasis [58]. With MS, the role of TNFRSF14 involves influencing T-cell activation and cytokine production, which is a contributor to the inflammatory environment that is often seen in the disease [59]. Investigations have found a link between TNFRSF14 polymorphisms and MS risk, particularly in people with active human herpesvirus 6 (HHV-6) replication [60, 61], which suggests involvement of TNFRSF14 in the viral mechanisms that exacerbate MS. It is an additional candidate disease activity and progression biomarker.

Single-cell RNA sequencing and the GEO dataset were used for the further validation of the results of this study. All hub genes coding these plasma proteins were found to have a good diagnostic effect in the training dataset and all genes in the test dataset also showed high diagnostic effects, except CD58.

The key strengths of the study included using MR design to minimise bias from reverse causality and potential confounders. It also included cis-pQTLs that are able to increase the strength of the evidence (cis-pQTL > trans-pQTL > eQTL) and gene colocalization analyses to enhance statistical efficiency and result validity. This investigation used scRNA-seq data for the analysis of the expression of five key genes (AIF1, AGER, TNFRSF14, CD58 and EVI5) across various cell types. The results found significant expression differences of these genes in distinct cell types, particularly within immune cells. This suggested that they play crucial roles in MS immunoregulation. Complete phenotypic association analysis enabled a thorough investigation of the side effects of potential therapeutics. It is hoped that other researchers will use the PheWAS technique for examining the adverse effects of pharmacologic targets to expand the body of knowledge in the field.

However, this study has some shortcomings. Firstly, the fact that all GWAS participants were European may have impacted the generalisability of its findings. Secondly, despite the fact that the UKB-PPP data included 2,940 plasma proteins, the MR study only included 1,280 plasma proteins due to the instrumental variable limitations. Thirdly, the investigation was confined to cases where the combined posterior association probability (PPH3+PPH4) was greater than or equal to 0.8 due to colocalization analyses having less power. Fourthly, realistic settings may make it impossible to supplement animal and cell experiments, so these trials perhaps should be included in further studies in this field. Finally, this study used a multiple regression analysis (MR) method to search for causal connections between MS and plasma proteins but did not use a genetics-led drug target prioritisation method (priority index, PI) to prioritise less-explored targets.

Conclusion

This study has examined the causal link between five plasma proteins (AIF1, AGER, TNFRSF14, CD58 and EVI5) and MS to identify new targets for MS therapy. It is hoped that future research can examine these drug targets, their potential therapeutic strategies, and the ways in which they might affect MS treatment.

Disclosure of conflict of interest

None.

Address correspondence to: Ke Yi, Department of Cardiology, Hubei Provincial Key Laboratory of Occurrence and Intervention of Rheumatic Diseases/Hubei Provincial Clinical Research Center for Nephrology, Minda Hospital of Hubei Minzu University, Hubei Minzu University, No. 2, Wufengshan Road, Enshi 445000, Hubei, China. E-mail: yikeltt@hotmail.com

References

[1] Thompson AJ, Baranzini SE, Geurts J, Hemmer B and Ciccarelli O. Multiple sclerosis. Lancet 2018; 391: 1622-1636.
[2] Oh J, Vidal-Jordana A and Montalban X. Multiple sclerosis: clinical aspects. Curr Opin Neurol 2018; 31: 752-759.
[3] Jakimovski D, Bittner S, Zivadinov R, Morrow SA, Benedict RH, Zipp F and Weinstock-Guttman B. Multiple sclerosis. Lancet 2024; 403: 183-202.
[4] Ward M and Goldman MD. Epidemiology and pathophysiology of multiple sclerosis. Continuum (Minneap Minn) 2022; 28: 988-1005.

[5] Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, Correale J, Fazekas F, Filippi M, Freedman MS, Fujihara K, Galetta SL, Hartung HP, Kappos L, Lublin FD, Marrie RA, Miller AE, Miller DH, Montalban X, Mowry EM, Sorensen PS, Tintoré M, Traboulsee AL, Trojano M, Uitdehaag BMJ, Vukusic S, Waubant E, Weinshenker BG, Reingold SC and Cohen JA. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol 2018; 17: 162-173.

[6] Lefeuvre JA, Guy JR, Luciano NJ, Ha SK, Leibovitch E, Santin MD, Silva AC, Jacobson S, Lehéricy S, Reich DS and Sati P. The spectrum of spinal cord lesions in a primate model of multiple sclerosis. Mult Scler 2020; 26: 284-293.

[7] Yousuf MS, Noh MC, Friedman TN, Zubkow K, Johnson JC, Tenorio G, Kurata HT, Smith PA and Kerr BJ. Sensory neurons of the dorsal root ganglia become hyperexcitable in a T-cell-mediated MOG-EAE model of multiple sclerosis. eNeuro 2019; 6: ENEURO.0024-19.2019.

[8] Olsson T, Barcellos LF and Alfredsson L. Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. Nat Rev Neurol 2017; 13: 25-36.

[9] Attfield KE, Jensen LT, Kaufmann M, Friese MA and Fugger L. The immunology of multiple sclerosis. Nat Rev Immunol 2022; 22: 734-750.

[10] Lassmann H. Multiple sclerosis pathology. Cold Spring Harb Perspect Med 2018; 8: a028936.

[11] Mey GM, Mahajan KR and DeSilva TM. Neurodegeneration in multiple sclerosis. WIREs Mech Dis 2023; 15: e1583.

[12] Rodríguez Murúa S, Farez MF and Quintana FJ. The immune response in multiple sclerosis. Annu Rev Pathol 2022; 17: 121-139.

[13] Callegari I, Derfuss T and Galli E. Update on treatment in multiple sclerosis. Presse Med 2021; 50: 104068.

[14] Charabati M, Wheeler MA, Weiner HL and Quintana FJ. Multiple sclerosis: neuroimmune crosstalk and therapeutic targeting. Cell 2023; 186: 1309-1327.

[15] Vasileiou ES and Fitzgerald KC. Multiple sclerosis pathogenesis and updates in targeted therapeutic approaches. Curr Allergy Asthma Rep 2023; 23: 481-496.

[16] Duan H, Jing Y, Li Y, Lian Y, Li J and Li Z. Rehabilitation treatment of multiple sclerosis. Front Immunol 2023; 14: 1168821.

[17] Hauser SL and Cree BAC. Treatment of multiple sclerosis: a review. Am J Med 2020; 133: 1380-1390.e1382.

[18] Amin M and Hersh CM. Updates and advances in multiple sclerosis neurotherapeutics. Neurodegener Dis Manag 2023; 13: 47-70.

[19] Shi F, Zhang G, Li J, Shu L, Yu C, Ren D, Zhang Y and Zheng P. Integrated analysis of single cell-RNA sequencing and Mendelian randomization identifies lactate dehydrogenase B as a target of melatonin in ischemic stroke. CNS Neurosci Ther 2024; 30: e14741.

[20] Chen J, Xu F, Ruan X, Sun J, Zhang Y, Zhang H, Zhao J, Zheng J, Larsson SC, Wang X, Li X and Yuan S. Therapeutic targets for inflammatory bowel disease: proteome-wide Mendelian randomization and colocalization analyses. EBioMedicine 2023; 89: 104494.

[21] Zhang W, Ma L, Zhou Q, Gu T, Zhang X and Xing H. Therapeutic targets for diabetic kidney disease: proteome-wide Mendelian randomization and colocalization analyses. Diabetes 2024; 73: 618-627.

[22] Ference BA. Interpreting the clinical implications of drug-target Mendelian randomization studies. J Am Coll Cardiol 2022; 80: 663-665.

[23] Walker VM, Zheng J, Gaunt TR and Smith GD. Phenotypic causal inference using genome-wide association study data: Mendelian randomization and beyond. Annu Rev Biomed Data Sci 2022; 5: 1-17.

[24] Jerome RN, Joly MM, Kennedy N, Shirey-Rice JK, Roden DM, Bernard GR, Holroyd KJ, Denny JC and Pulley JM. Leveraging human genetics to identify safety signals prior to drug marketing approval and clinical use. Drug Saf 2020; 43: 567-582.

[25] Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, Burgess S, Jiang T, Paige E, Surendran P, Oliver-Williams C, Kamat MA, Prins BP, Wilcox SK, Zimmerman ES, Chi A, Bansal N, Spain SL, Wood AM, Morrell NW, Bradley JR, Janjic N, Roberts DJ, Ouwehand WH, Todd JA, Soranzo N, Suhre K, Paul DS, Fox CS, Plenge RM, Danesh J, Runz H and Butterworth AS. Genomic atlas of the human plasma proteome. Nature 2018; 558: 73-79.

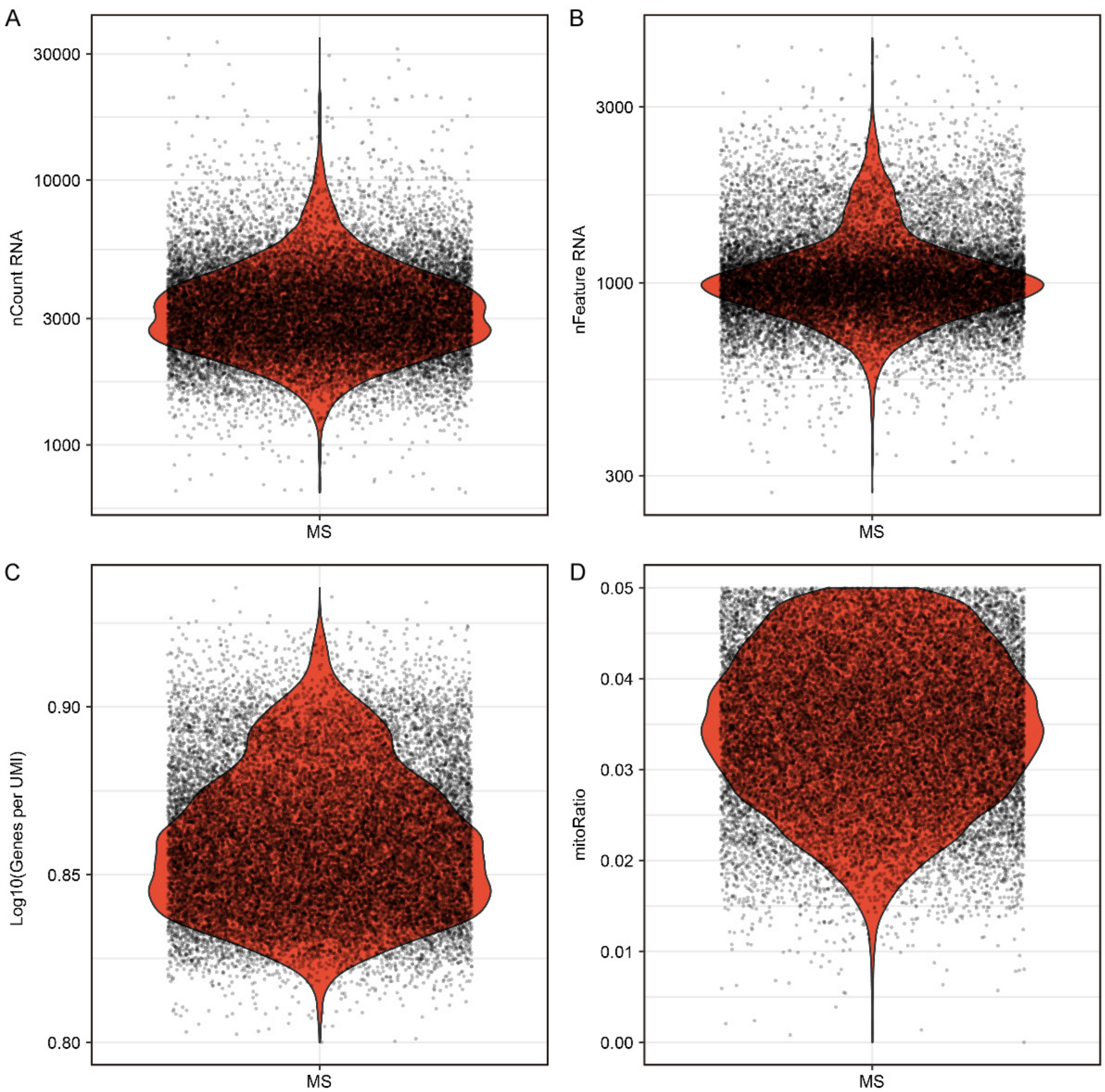
[26] Sun BB, Chiou J, Traylor M, Benner C, Hsu YH, Richardson TG, Surendran P, Mahajan A, Robins C, Vasquez-Grinnell SG, Hou L, Kvikstad EM, Burren OS, Davitte J, Ferber KL, Gillies CE, Hedman ÅK, Hu S, Lin T, Mikkilineni R, Pendergrass RK, Pickering C, Prins B, Baird D, Chen CY, Ward LD, Deaton AM, Welsh S, Willis CM, Lehner N, Arnold M, Wörheide MA, Suhre K, Kastenmüller G, Sethi A, Cule M and Raj A; Alnylam Human Genetics; AstraZeneca Genomics Initiative; Biogen Biobank Team; Bristol Myers Squibb; Genentech Human Genetics; GlaxoSmithKline Genomic Sciences; Pfizer Integrative Biology; Population Analytics of Janssen Data Sciences; Regeneron Genetics Center; Burkitt-Gray L, Melamud E, Black MH, Fauman EB, Howson JMM, Kang HM, McCarthy MI, Nioi P, Petrovski S, Scott RA, Smith EN,

Szalma S, Waterworth DM, Mitnaul LJ, Szustakowski JD, Gibson BW, Miller MR and Whelan CD. Plasma proteomic associations with genetics and health in the UK Biobank. Nature 2023; 622: 329-338.
[27] Fauman EB and Hyde C. An optimal variant to gene distance window derived from an empirical definition of cis and trans protein QTLs. BMC Bioinformatics 2022; 23: 169.
[28] Wang L, Tassiulas I, Park-Min KH, Reid AC, Gil-Henn H, Schlessinger J, Baron R, Zhang JJ and Ivashkiv LB. 'Tuning' of type I interferon-induced Jak-STAT1 signaling by calcium-dependent kinases in macrophages. Nat Immunol 2008; 9: 186-193.
[29] Pietzner M, Wheeler E, Carrasco-Zanini J, Cortes A, Koprulu M, Wörheide MA, Oerton E, Cook J, Stewart ID, Kerrison ND, Luan J, Raffler J, Arnold M, Arlt W, O'Rahilly S, Kastenmüller G, Gamazon ER, Hingorani AD, Scott RA, Wareham NJ and Langenberg C. Mapping the proteo-genomic convergence of human diseases. Science 2021; 374: eabj1541.
[30] Stacey D, Fauman EB, Ziemek D, Sun BB, Harshfield EL, Wood AM, Butterworth AS, Suhre K and Paul DS. ProGeM: a framework for the prioritization of candidate causal genes at molecular quantitative trait loci. Nucleic Acids Res 2019; 47: e3.
[31] Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, Reeve MP, Laivuori H, Aavikko M, Kaunisto MA, Loukola A, Lahtela E, Mattsson H, Laiho P, Della Briotta Parolo P, Lehisto AA, Kanai M, Mars N, Rämö J, Kiiskinen T, Heyne HO, Veerapen K, Rüeger S, Lemmelä S, Zhou W, Ruotsalainen S, Pärn K, Hiekkalinna T, Koskelainen S, Paajanen T, Llorens V, Gracia-Tabuenca J, Siirtola H, Reis K, Elnahas AG, Sun B, Foley CN, Aalto-Setälä K, Alasoo K, Arvas M, Auro K, Biswas S, Bizaki-Vallaskangas A, Carpen O, Chen CY, Dada OA, Ding Z, Ehm MG, Eklund K, Färkkilä M, Finucane H, Ganna A, Ghazal A, Graham RR, Green EM, Hakanen A, Hautalahti M, Hedman ÅK, Hiltunen M, Hinttala R, Hovatta I, Hu X, Huertas-Vazquez A, Huilaja L, Hunkapiller J, Jacob H, Jensen JN, Joensuu H, John S, Julkunen V, Jung M, Junttila J, Kaarniranta K, Kähönen M, Kajanne R, Kallio L, Kälviäinen R and Kaprio J; FinnGen; Kerimov N, Kettunen J, Kilpeläinen E, Kilpi T, Klinger K, Kosma VM, Kuopio T, Kurra V, Laisk T, Laukkanen J, Lawless N, Liu A, Longerich S, Mägi R, Mäkelä J, Mäkitie A, Malarstig A, Mannermaa A, Maranville J, Matakidou A, Meretoja T, Mozaffari SV, Niemi MEK, Niemi M, Niiranen T, O Donnell CJ, Obeidat ME, Okafo G, Ollila HM, Palomäki A, Palotie T, Partanen J, Paul DS, Pelkonen M, Pendergrass RK, Petrovski S, Pitkäranta A, Platt A, Pulford D, Punkka E, Pussinen P, Raghavan N, Rahimov F, Rajpal D, Renaud NA, Riley-Gillis B, Rodosthenous R, Saarentaus E, Salminen A, Salminen E, Salomaa V, Schleutker J, Serpi R, Shen HY, Siegel R, Silander K, Siltanen S, Soini S, Soininen H, Sul JH, Tachmazidou I, Tasanen K, Tienari P, Toppila-Salmi S, Tukiainen T, Tuomi T, Turunen JA, Ulirsch JC, Vaura F, Virolainen P, Waring J, Waterworth D, Yang R, Nelis M, Reigo A, Metspalu A, Milani L, Esko T, Fox C, Havulinna AS, Perola M, Ripatti S, Jalanko A, Laitinen T, Mäkelä TP, Plenge R, McCarthy M, Runz H, Daly MJ and Palotie A. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature 2023; 613: 508-518.
[32] International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. Science 2019; 365: eaav7188.
[33] Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C and Plagnol V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 2014; 10: e1004383.
[34] Su WM, Gu XJ, Dou M, Duan QQ, Jiang Z, Yin KF, Cai WC, Cao B, Wang Y and Chen YP. Systematic druggable genome-wide Mendelian randomisation identifies therapeutic targets for Alzheimer's disease. J Neurol Neurosurg Psychiatry 2023; 94: 954-961.
[35] Enz LS, Zeis T, Schmid D, Geier F, van der Meer F, Steiner G, Certa U, Binder TMC, Stadelmann C, Martin R and Schaeren-Wiemers N. Increased HLA-DR expression and cortical demyelination in MS links with HLA-DR15. Neurol Neuroimmunol Neuroinflamm 2019; 7: e656.
[36] Kemppinen AK, Kaprio J, Palotie A and Saarela J. Systematic review of genome-wide expression studies in multiple sclerosis. BMJ Open 2011; 1: e000053.
[37] Lin J, Zhou J and Xu Y. Potential drug targets for multiple sclerosis identified through Mendelian randomization analysis. Brain 2023; 146: 3364-3372.
[38] Chen R, Zhu S, Zeng L, Wang Q, Sheng Y, Zhou B, Tang D and Kang R. AGER-mediated lipid peroxidation drives caspase-11 inflammasome activation in sepsis. Front Immunol 2019; 10: 1904.
[39] Fang X, Dai E, Bai L, Liu J, Kang R, Zhao Y, Han L and Tang D. The HMGB1-AGER-STING1 pathway mediates the sterile inflammatory response to alkaliptosis. Biochem Biophys Res Commun 2021; 560: 165-171.
[40] Gutowska K, Koźniewski K, Wąsowski M, Jonas MI, Bartoszewicz Z, Lisik W, Jonas M, Binda A, Jaworski P, Tarnowski W, Noszczyk B, Puzianowska-Kuźnicka M, Czajkowski K and Kuryłowicz A. AGER-1 long non-coding RNA levels correlate with the expression of the ad-

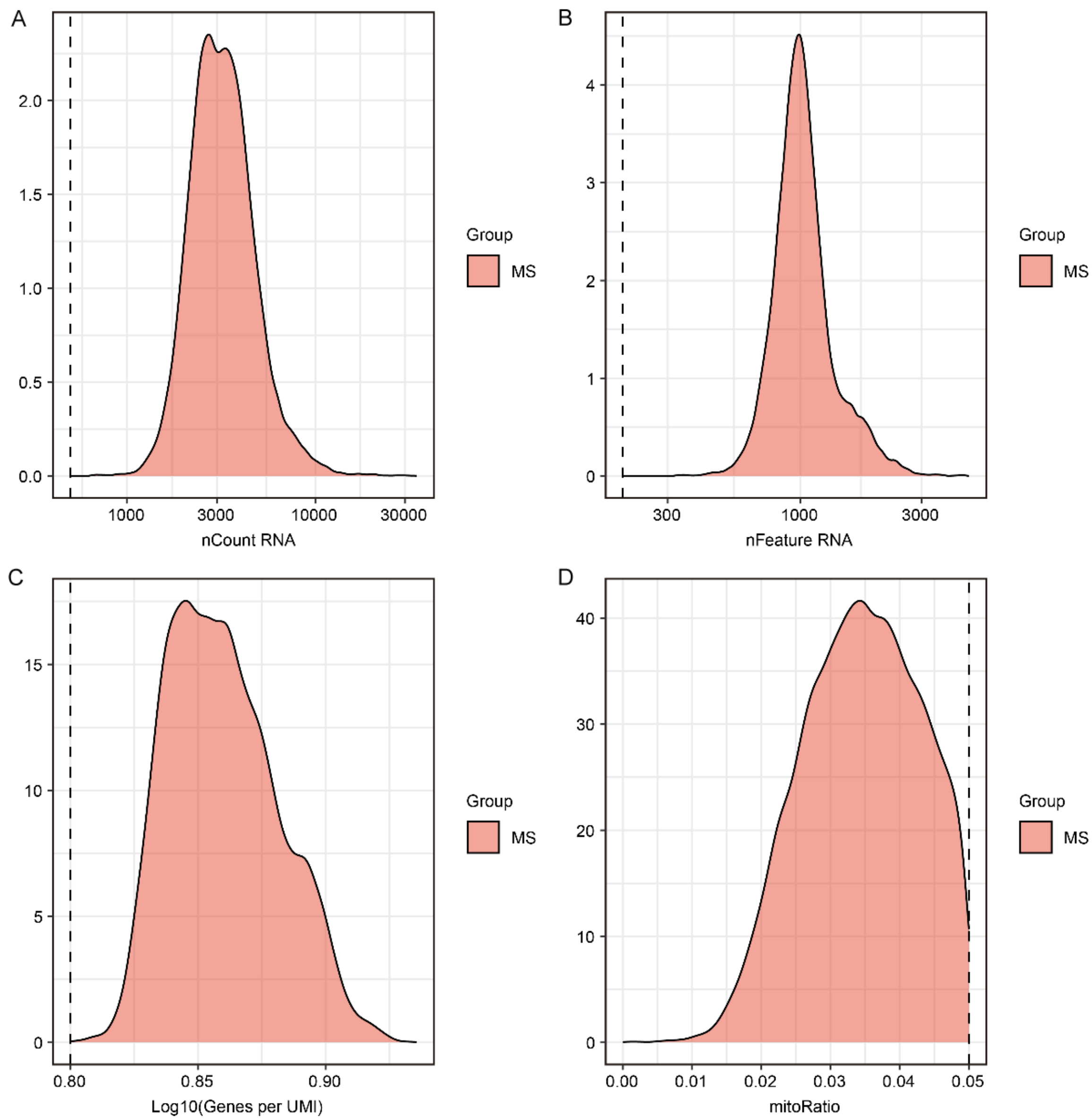
vanced glycosylation end-product receptor, a regulator of the inflammatory response in visceral adipose tissue of women with obesity and type 2 diabetes mellitus. Int J Mol Sci 2023; 24: 17447.
[41] Liu J, Zhu S, Zeng L, Li J, Klionsky DJ, Kroemer G, Jiang J, Tang D and Kang R. DCN released from ferroptotic cells ignites AGER-dependent immune responses. Autophagy 2022; 18: 2036-2049.
[42] Vlassara H. The AGE-receptor in the pathogenesis of diabetic complications. Diabetes Metab Res Rev 2001; 17: 436-443.
[43] Zeng H, Cheng L, Lu DZ, Fan S, Wang KX, Xu LL, Cai B, Zhou MW and Wang JW. Unbiased multitissue transcriptomic analysis reveals complex neuroendocrine regulatory networks mediated by spinal cord injury-induced immunodeficiency. J Neuroinflammation 2023; 20: 219.
[44] Chinnasamy P, Lutz SE, Riascos-Bernal DF, Jeganathan V, Casimiro I, Brosnan CF and Sibinga NE. Loss of allograft inflammatory factor-1 ameliorates experimental autoimmune encephalomyelitis by limiting encephalitogenic CD4 T-cell expansion. Mol Med 2015; 21: 233-241.
[45] Peres DS, Theisen MC, Fialho MFP, Dalenogare DP, Rodrigues P, Kudsi SQ, Bernardes LB, Ruviaro da Silva NA, Lückemeyer DD, Sampaio TB, Pereira GC, Mello FK, Ferreira J, Bochi GV, Oliveira SM, de David Antoniazzi CT and Trevisan G. TRPA1 involvement in depression- and anxiety-like behaviors in a progressive multiple sclerosis model in mice. Brain Res Bull 2021; 175: 1-15.
[46] Gray E, Kemp K, Hares K, Redondo J, Rice C, Scolding N and Wilkins A. Increased microglial catalase activity in multiple sclerosis grey matter. Brain Res 2014; 1559: 55-64.
[47] Zhang Y, Liu Q, Yang S and Liao Q. CD58 immunobiology at a glance. Front Immunol 2021; 12: 705260.
[48] Ahmed A, Rawshan AEM, Tishe ZH, Shawkat S, Popy MN, Shohag MH, Hossain M and Mostaid MS. Association of CD58 rs12044852 and rs2300747 polymorphisms with the risk of multiple sclerosis: a systematic review and meta-analysis. Mult Scler Relat Disord 2024; 82: 105411.
[49] Zhou W, Hu W, Tang L, Ma X, Liao J, Yu Z, Qi M, Chen B and Li J. Meta-analysis of the selected genetic variants in immune-related genes and multiple sclerosis risk. Mol Neurobiol 2024; 61: 8175-8187.
[50] Cai T, Zhou J, Zeng Y, Du W, Zhang Y, Liu T, Fu Y, Huang JA, Qian Q, Zhu J, Ling C and Liu Z. EVI5 is an oncogene that regulates the proliferation and metastasis of NSCLC cells. J Exp Clin Cancer Res 2020; 39: 84.
[51] Baron BW, Anastasi J, Bies J, Reddy PL, Joseph L, Thirman MJ, Wroblewski K, Wolff L and Baron JM. GFI1B, EVI5, MYB--additional genes that cooperate with the human BCL6 gene to promote the development of lymphomas. Blood Cells Mol Dis 2014; 52: 68-75.
[52] Liao X, Du Y, Morse HC 3rd, Jenkins NA and Copeland NG. Proviral integrations at the Evi5 locus disrupt a novel 90 kDa protein with homology to the Tre2 oncogene and cell-cycle regulatory proteins. Oncogene 1997; 14: 1023-1029.
[53] Liu J, Liu X, Liu Y, Deng S, Huang H, Chen Q, Liu W and Huang Z. Association of EVI5 rs11808092, CD58 rs2300747, and CIITA rs3087456 polymorphisms with multiple sclerosis risk: a meta-analysis. Meta Gene 2016; 9: 97-103.
[54] Lim YS and Tang BL. The Evi5 family in cellular physiology and pathology. FEBS Lett 2013; 587: 1703-1710.
[55] Hoppenbrouwers IA, Aulchenko YS, Ebers GC, Ramagopalan SV, Oostra BA, van Duijn CM and Hintzen RQ. EVI5 is a risk gene for multiple sclerosis. Genes Immun 2008; 9: 334-337.
[56] Mazdeh M, Ghafouri-Fard S, Noroozi R, Sayad A, Khani M, Taheri M and Davood Omrani M. Ecotropic Viral Integration Site 5 (EVI5) variants are associated with multiple sclerosis in Iranian population. Mult Scler Relat Disord 2017; 18: 15-19.
[57] Bobik A and Kalinina N. Tumor necrosis factor receptor and ligand superfamily family members TNFRSF14 and LIGHT: new players in human atherogenesis. Arterioscler Thromb Vasc Biol 2001; 21: 1873-1875.
[58] Soroosh P, Doherty TA, So T, Mehta AK, Khorram N, Norris PS, Scheu S, Pfeffer K, Ware C and Croft M. Herpesvirus entry mediator (TNFRSF14) regulates the persistence of T helper memory cell populations. J Exp Med 2011; 208: 797-809.
[59] Steinberg MW, Cheung TC and Ware CF. The signaling networks of the herpesvirus entry mediator (TNFRSF14) in immune regulation. Immunol Rev 2011; 244: 169-187.
[60] Blanco-Kelly F, Alvarez-Lafuente R, Alcina A, Abad-Grau MM, de Las Heras V, Lucas M, de la Concha EG, Fernández O, Arroyo R, Matesanz F and Urcelay E. Members 6B and 14 of the TNF receptor superfamily in multiple sclerosis predisposition. Genes Immun 2011; 12: 145-148.
[61] Shui JW, Steinberg MW and Kronenberg M. Regulation of inflammation, autoimmunity, and infection immunity by HVEM-BTLA signaling. J Leukoc Biol 2011; 89: 517-523.

Supplementary Table 3. Colocalization results for the five MS-associated plasma proteins within ±1 Mb of their corresponding genes (PPH3+PPH4 > 0.8)

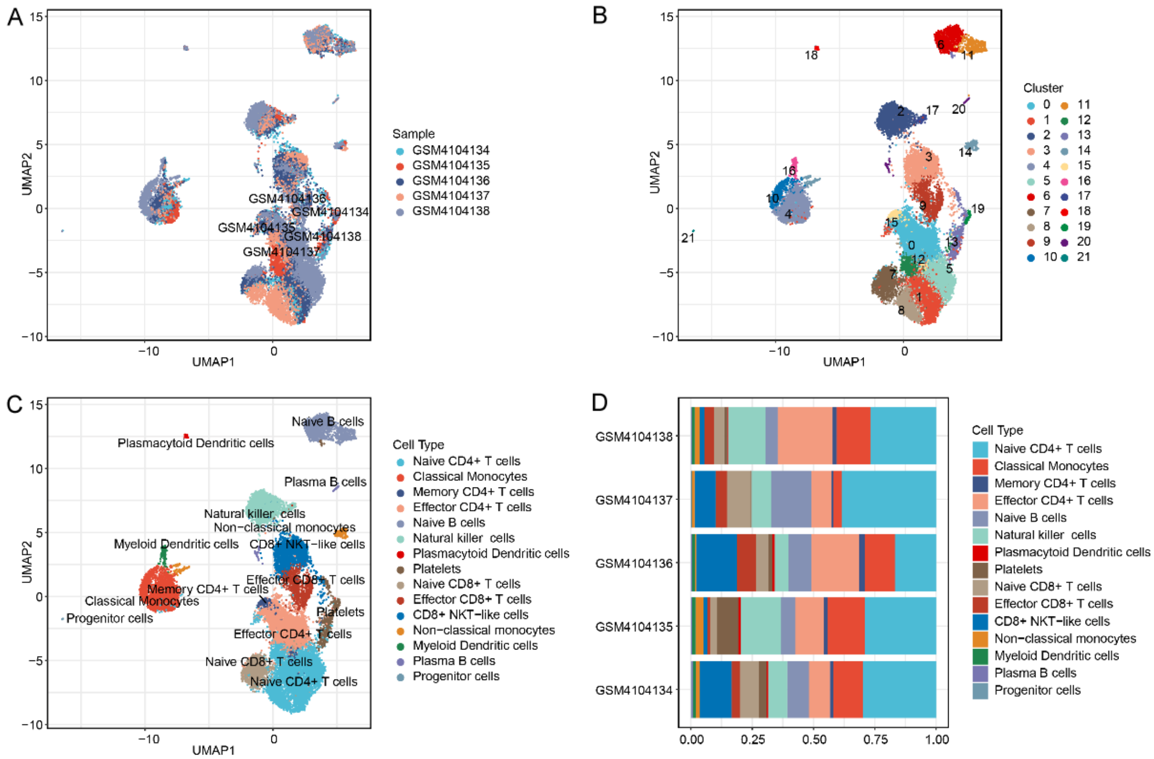
EVI5	coloc_res$summary
nsnps	8076
PP.H0.abf	8.62E-62
PP.H1.abf	6.99E-05
PP.H2.abf	1.04E-58
PP.H3.abf	0.083890148
PP.H4.abf	0.916039909
CD58	coloc_res$summary
nsnps	7552
PP.H0.abf	1.47E-110
PP.H1.abf	2.91E-03
PP.H2.abf	6.94E-110
PP.H3.abf	0.012765808
PP.H4.abf	0.984325889
AIF1	coloc_res$summary
nsnps	19823
PP.H0.abf	9.15E-175
PP.H1.abf	1.35E-135
PP.H2.abf	6.79E-40
PP.H3.abf	1
PP.H4.abf	9.65895E-39
TNFRSF14	coloc_res$summary
nsnps	9065
PP.H0.abf	7.87E-49
PP.H1.abf	6.97E-02
PP.H2.abf	6.63E-48
PP.H3.abf	0.58662422
PP.H4.abf	0.343703657
AGER	coloc_res$summary
nsnps	25637
PP.H0.abf	3.94E-231
PP.H1.abf	6.15E-136
PP.H2.abf	6.41E-96
PP.H3.abf	1
PP.H4.abf	1.67E-87



Supplementary Figure 1. Violin plot of single-cell data quality control. (A-D) Single-celled (scRNA - seq) data sets GSE138266 nCount RNA (UMI) (A), nFeature RNA (B), Log10 (Genes per UMI) (C), Mitochondrial genes accounted for (mitoRatio) (D) the data of quality control chart on the violin. Red samples are multiple sclerosis (MS) samples. MS, and Multiple Sclerosis; UMI: Unique Molecular Identifier.



Supplementary Figure 2. Density plot of single-cell data quality control. (A-D) Single-celled (scRNA - seq) data sets GSE138266 nCount RNA (UMI) (A), nFeature RNA (B), Log10 (Genes per UMI) (C), Data distribution density plot of mitochondrial gene proportion (mitoRatio) (D). Red samples are multiple sclerosis (MS) samples. MS, and Multiple Sclerosis; UMI: Unique Molecular Identifier.



Supplementary Figure 3. Single-cell cluster and cell type annotation: A. Single-celled (scRNA - seq) data set GSE138266 cells express situation UMAP graph in different samples. Different colors represent different samples. B. UMAP map of cell expression in different cell clusters. Different colors indicate different clusters. C. Cell expression in different cell types of the UMAP figure. Different colors indicate different cell types. D. Single-celled (scRNA - seq) data set GSE138266 cell percentage histogram in different samples. MS: Multiple Sclerosis; UMAP: Uniform Manifold Approximation and the Projection.