ANNALS of Clinical and Translational Neurology
Open Access
AMERICAN NEUROLOGICAL ASSOCIATION

RESEARCH ARTICLE

Brain proteome-wide association study linking-genes in multiple sclerosis pathogenesis

Tingting Jia[1], Yanni Ma[1], Fengqin Qin[2], Feng Han[3] & Chengcheng Zhang[1]

[1]Mental Health Center and Psychiatric Laboratory, the State Key Laboratory of Biotherapy, West China Hospital of Sichuan University, Chengdu, Sichuan, China
[2]Department of Neurology, the 3rd Affiliated Hospital of Chengdu Medical College, Chengdu, Sichuan, China
[3]Department of Emergency Medicine, Hainan General Hospital, Hainan Affiliated Hospital of Hainan Medical University, Haikou, China

Correspondence
Chengcheng Zhang, Mental Health Center and Psychiatric Laboratory, the State Key Laboratory of Biotherapy, West China Hospital of Sichuan University, Chengdu, Sichuan 610041, China. Tel: 0086-028-85422716; Fax: 0086-028-85422716; E-mail: zhangcc@scu.edu.cn;

Received: 26 August 2022; Revised: 28 October 2022; Accepted: 2 November 2022

Annals of Clinical and Translational Neurology **2023; 10(1): 58–69**

doi: 10.1002/acn3.51699

Abstract

Objectives: To identify genes that confer MS risk via the alteration of cis-regulated protein abundance and verify their aberrant expression in human brain. **Methods**: Utilizing a two-stage proteome-wide association study (PWAS) design, MS GWAS data (N = 41,505) was respectively integrated with two distinct human brain proteomes from the dorsolateral prefrontal cortex, including ROSMAP (N = 376) in the discovery stage and Banner (N = 152) in the confirmation stage. In the following, Bayesian colocalization analysis was conducted for GWAS and protein quantitative trait loci signals to prioritize candidate genes. Differential expression analysis was then used to verify the dysregulation of risk genes in white matter and gray matter for evidence at the transcription level. **Results**: A total of 51 genes whose protein abundance had association with the MS risk were identified, of which 18 genes overlapped in the discovery and confirmation PWAS. Bayesian colocalization indicated six causal genes with genetic risk variants for the MS risk. The differential expression analysis of *SHMT1* ($P_{FDR} = 4.82 \times 10^{-2}$), *FAM120B* ($P_{FDR} = 8.13 \times 10^{-4}$) in white matter and *ICA1L* ($P_{FDR} = 3.44 \times 10^{-2}$) in gray matter confirmed the dysregulation at the transcription level. Further investigation of expression found *SHMT1* significantly up-regulated in white matter lesion, and *FAM120B* up-regulated in both white matter lesion and normal appearing white matter. *ICA1L* was down-regulated in both gray matter lesion and normal appearing gray matter. **Interpretation**: Dysregulation of *SHMT1*, *FAM120B* and *ICA1L* may confer MS risk. Our findings shed new light on the pathogenesis of MS and prioritized promising targets for future therapy research.

Introduction

Multiple sclerosis (MS) is a chronic inflammatory, demyelinating and neurodegenerative disease of the central nervous system that affects over 2.3 million people worldwide.[1] Identification of MS biomarkers to aid in the diagnosis and treatment at an early stage is extremely important due to the varied clinical characteristics of MS and the poor effectiveness of current treatments.[2] As the final products of genetic effects, protein biomarkers are ideal measurable molecules that provide a clue about the development of MS. Protein abundance alteration in human brain has been proved associated with MS. For instance, protein abundance of glial fibrillary acidic protein (GFAP),[3,4] myelin basic protein (MBP)[3,5] and thymosin β-4[6] was dysregulated in lesions from MS patients' brain, and these proteins have been used for disease severity prediction[7] and targeted therapy[8] lately. In addition, comparing with bodily fluid samples like cerebrospinal fluid[9–11] and plasma,[9,12] human brain tissue directly reflects the pathophysiology changes of MS and has become increasingly important in disease biomarker identification.[5,13] However, few studies focused on a specific subregion of human brain, which has ignored the possible differences in protein types and abundance between subregions with distinct functions.[14,15] Recent

transcriptomic[16] and epigenetic[17] investigations have linked the dorsolateral prefrontal cortex (dlPFC) to MS susceptibility. Insufficiency of dlPFC proteomic researches remains ambiguity for MS pathogenesis and treatment exploration.

Polygenic susceptibility is thought to be an important factor in MS pathogenesis, and MS heritability was estimated to be 0.64 (95%CI 0.36–0.76).[18] Post genome-wide association studies (GWAS) analyses provide an opportunity for this void, for example, proteome-wide association studies (PWAS) developed recently establish associations between proteome abundance and disease phenotype using protein quantitative trait loci (pQTL)[19] to explain the effects of genetic architectures in terms of downstream cis-regulated proteins. The integrative analysis combining PWAS and Bayesian colocalization to identify susceptibility genes for certain diseases has recently been used in several studies for risk loci excavation,[20–22] paving the way for prioritizing biomarkers that play a crucial role in pathogenesis of MS. In order to determine if susceptibility genes identified by integrative analysis were expressed differently in the MS postmortem brain, it is necessary to explore the expression of risk genes from two dimensions successively, distribution (white and gray matter) and lesion degree (lesion and normal appearing tissue). MS was formerly regarded as a demyelinating disease involving white matter, recently seizures[23] and cognitive impairment[24] are two symptoms of MS that may be brought on by inflammatory lesions that damage gray matter rich in neurites and neuron cell bodies. Thus, in addition to white matter, cortical lesions which have an important role in MS pathophysiology[25–27] also need to be explored.

In the current study, we combined high-throughput proteomics with genetic summary statistics in order to identify genomic architecture-associated protein biomarkers for MS and to provide promising targets for future pathogenesis studies. A three-step approach was used to systematically link protein biomarkers to MS. Figure 1 summarizes the overall analysis workflow applied in this study. First, we performed a two-stage PWAS analysis using two independent human brain protein quantitative trait loci (pQTL) data and summary statistics from a large-scale MS GWAS. At the discovery stage, we used human brain pQTL data from the dlPFC of 376 individuals collected by the Religious Order Research (ROS) or the Rush Memory and Aging Project (MAP) (ROSMAP dataset). At the confirmation stage, we used pQTL data from the dlPFC of 152 individuals collected at the Banner Sun Health Institute (Banner dataset). Second, Bayesian colocalization was used to highlight genes with pQTL/GWAS signals driven by shared causal variants. Finally, differential expression analysis was performed to explore the risk gene dysregulation in white matter and cortical gray matter in precentral gyrus for MS cases when compared with healthy controls.

Materials and Methods

GWAS

We focused on the GWAS result conducted by the International Multiple Sclerosis Genetics Consortium (IMSGC)[28]in the current integrative study. Briefly, IMSGC conducted genome-wide meta-analysis through using 14,802 MS cases and 26,703 controls in discovery phase, containing 15 datasets including UK, CE, Medi, Nordic, US, AUS, FINLAND, GeneMSA DU, GeneMSA SW, GeneMSA US, IMSGC, BWH/MIGEN, ANZ, Berkeley and Rotterdam. The fixed effects inverse-variance meta-analysis was performed. More details about the sample description, genotyping, quality control, and statistical analyses could be found in original paper.[28]

pQTL in the discovery PWAS

The discovery ROSMAP dataset was generated by Wingo et al[20,29] using dlPFC tissues of 376 individuals of European ancestry. Removing the effects of clinical characteristics and technical factors (protein batch, MS2 versus MS3 reporter quantitation mode, sex, age at death, postmortem interval, and study (ROS vs. MAP)), the normalized abundance of 8356 proteins were calculated, among which 1475 protein could find significant cis associations with genetic variation. The weights for these 1475 protein were used for the discovery PWAS in the study. More details about the sample description, proteomic analysis, quality control, and statistical analyses are provided in the original paper.[20,29]

pQTL in the confirmation PWAS

Wingo et al produced Banner dataset[20,30] using dlPFC tissue samples of 152 individuals of European ancestry. The procedure of proteomic analysis applied to Banner dataset was the same as described above for ROSMAP dataset. Among 8168 proteins passed quality control, a total of 1139 was heritable. The weights of these 1139 proteins were used for confirmation PWAS in this study. More details about the sample description, proteomic analysis, quality control, and statistical analyses are provided in the original paper.[20,30]

Proteome-wide association studies

Genotypes and protein abundance were combined to acquire the weights to predict the protein abundance

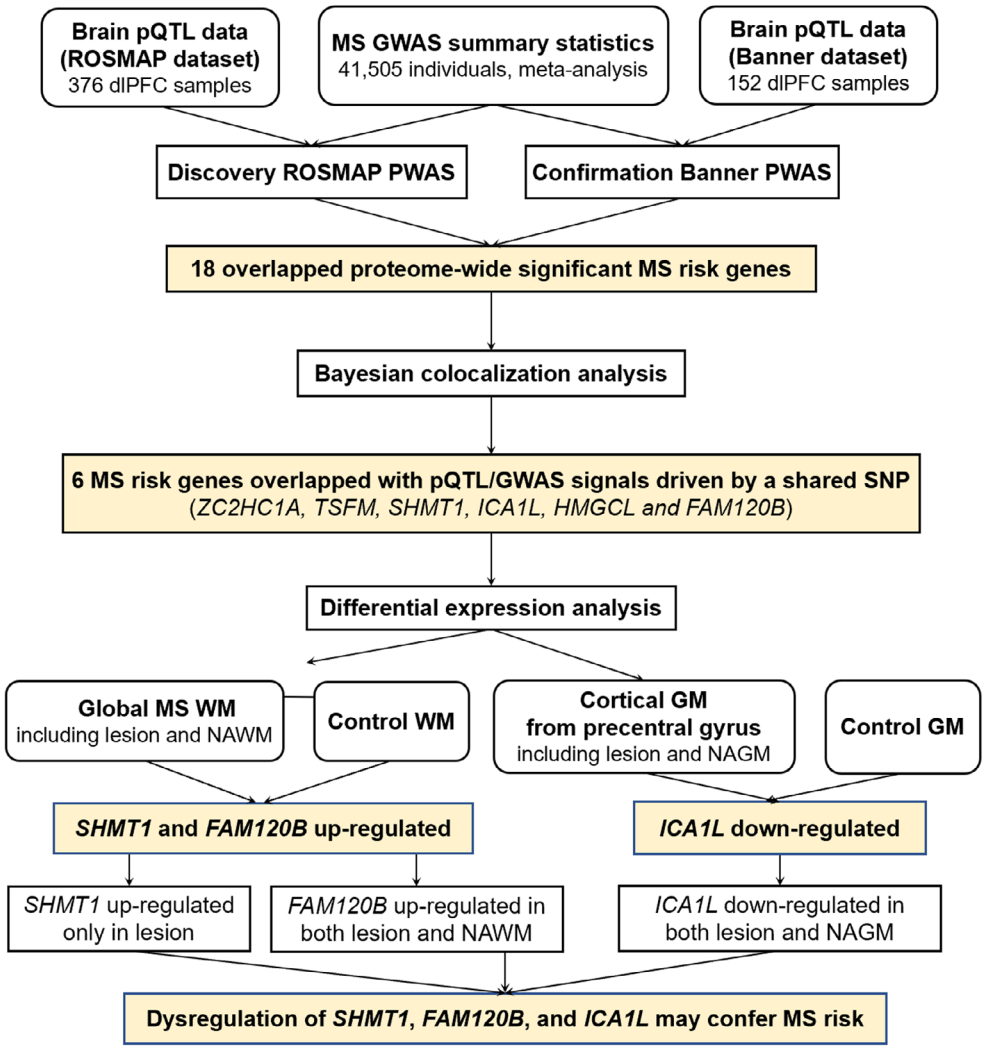

Figure 1. The integrated analysis approach used in present study to systematically link protein biomarkers to MS. Firstly, a two-stage PWAS analysis was performed using independent pQTL data (ROSMAP and Banner datasets) and summary statistics from a large-scale MS GWAS. Subsequently, we conducted Bayesian colocalization to identify risk genes with pQTL/GWAS signals driven by shared SNPs. Finally, differential expression analysis was carried out to explore the risk gene dysregulation in MS white matter and cortical gray matter with different histological manifestation (WM/GM lesions and NAWM/NAGM) as compared to healthy controls. pQTL, protein quantitative trait locus; ROSMAP, Religious Orders Study and Rush Memory and Aging Project; MS, multiple sclerosis; GWAS, genome-wide association study; PWAS, proteome-wide association study; SNP, single nucleotide polymorphism; WM, white matter; GM, gray matter; NAWM, normal appearing white matter; NAGM, normal appearing gray matter.

according to SNPs information in GWAS. We used FUSION.assoc_test.R default parameter to combine the genetic effect of MS (MS GWAS z-score) with the protein weights by calculating the linear sum of z-score×weight for the independent SNPs at the locus to perform the discovery and confirmation PWAS. The *P* value adjust for false discovery rate (FDR) using the Benjamini–Hochberg (BH) method were calculated to evaluate the significance level of each locus. Other settings complied with default FUSION parameters.[31]

Colocalization analysis

We also conducted colocalization analysis using the coloc R package.[32] Only genes with P_{FDR} <0.05 in discovery and confirmation PWAS were included in the subsequent analysis. In this approach, the association of signals from GWAS and pQTL with SNPs was assigned to five hypotheses (H_x) as follows: H_0, No association with either GWAS or pQTL; H_1, Association with GWAS, not with pQTL; H_2, Association with pQTL, not with GWAS; H_3,

Association with GWAS and pQTL, two independent SNPs; H_4, Association with GWAS and pQTL, one shared SNP. Coloc calculates the posterior probability of these five hypotheses based on Bayesian test. H_4>0.7 was considered as the threshold in colocalization analysis.

Differential expression analysis in brain

Differential expression analysis in global white matter

Differential expression analysis facilitates to validate the dysregulation of risk genes identified in integrated analysis at the level of transcription. Normalized gene expression profiles from white matter were obtained from two independent, well-characterized studies. (1) RNA-seq dataset contained the expression of white matter lesions from specific histological brain areas in progressive MS cases and controls without neurological disease.[33] A total of 98 snap-frozen white matter tissue blocks were harvested from UK Multiple Sclerosis Society Tissue Bank, including 52 white matter lesion, 21 normal appearing white matter (NAWM) and 25 white matter control blocks. The classification of tissue blocks in two pathologies (NAWM and demyelinated lesions) was based on myelin oligodendrocyte glycoprotein (MOG+) staining showing demyelinated lesions and HLA-DR+ staining showing the inflammatory state using the definition described previously.[34] Total RNA was isolated and sequenced by Illumina NextSeq550 after quality control by RNA integrity number. Transcripts were filtrated using Trimmomatic[35] and counted with strict mode in HTSeq-count.[36] Differential gene expression was performed using DEseq2 R package. The BH corrected threshold of P_{FDR} <0.05 was set for multiple comparison to determine if mRNA expression level of risk genes were statistically different between MS cases and controls. To determine whether the specificity of risk gene dysregulation in different histological areas, we compared expression in lesion and NAWM respectively with heathy controls[33] using Student's T Test. (2)The microarray dataset in white matter included tissue blocks from MS cases and healthy individuals provided by Netherlands Brain Bank.[37] In brief, 15 white matter lesion, 15 NAWM and 10 controls were harvest and classified according to activity measured by the staining of myelin proteolipid protein (PLP) and HLA-DP/Q/R defined previously.[38,39] Total RNA was extracted and hybridized on Agilent Human Gene Expression 4 × 44 K v2 Microarray, followed by normalization using loess and Gquantile algorithm in limma R package. Differential expression analysis was performed using limma R package. Investigation of risk genes expression in lesion and NAWM comparing with healthy controls[33] was performed with the same methodology mentioned above.

Differential expression analysis in cortical gray matter from precentral gyrus

Normalized gene expression profile in gray matter from precentral gyrus was generated using 20 cortical gray matter lesions, 20 normal appearing gray matter (NAGM) and 10 controls from UK Multiple Sclerosis Society Tissue Bank.[40] Tissue selection and classification were based on the results of MOG and MHC class II immunostaining as previously described.[41] Only subpial cortical lesions extending at least up to layer V were dissected, then Type III gray matter lesion and NAGM were prepared for RNA extraction. Total RNA was extracted and hybridized on Illumina whole genome HumanRef8 v2 BeadChip arrays, following normalization with Rosetta Biosoftware.[42] Differential gene expression analysis was performed with the same methodology mentioned above. Detailed information of expression profiles on each sample, including lesion classification, sequencing, quality control has been published previously.[33,37,40]

Results

Two-stage PWAS identified 18 overlapped proteome-wide significant risk genes for MS

We performed a two-stage PWAS via integrating two distinct human brain pQTL datasets and MS GWAS results. In the discovery stage, ROSMAP PWAS identified 35 proteome-wide significant (PWS) genes for MS at a FDR-adjusted *P* value (P_{FDR}) threshold of 0.05, indicating the protein abundance of these genes were associated with MS risk (Table 1). To further confirm the identified risk genes, we conducted Banner PWAS in the confirmation stage and found another 16 PWS genes. A total of 18 genes in the discovery ROSMAP PWAS successfully replicated in the Banner PWAS (14 genes up-regulated with positive *z*-score including *ZC2HC1A*, *TSFM, POGLUT1, TRAF3, DHRS11, SHMT1, GIMAP4, MTHFR, HMGCL, FAM120B, DOC2A, LRP4, WARS* and *GALC*, whereas 4 genes down-regulated with negative *z*-score including *ICA1L, TRIM47, AUH* and *PANK4*), and other 32 significant PWAS associations existed only in one dataset (ROSMAP or Banner). We focused on these 18 genes overlapped in the two-stage PWAS in the following integrated analysis, as they may confer MS risk by regulating the protein abundance.

Bayesian colocalization analysis highlighted 11 risk genes whose pQTL/GWAS signals were driven by shared genetic variants

To verify whether associations between MS and pQTL for the18 genes overlapped in previous two-stage PWAS were driven by a shared causal variant, we next performed

Table 1. The discovery ROSMAP PWAS identified 35 proteome-wide significant genes, of which 18 genes were replicated in Banner PWAS.

			Discovery PWAS			Confirmation PWAS			Evidence for confirmation
	Gene	Chromosome	PWAS z-score	PWAS *P*	PWAS FDR *P*	PWAS z-score	PWAS *P*	PWAS FDR *P*	
1	*ZC2HC1A*	8	5.65	1.61E-08	5.86E-06	5.68	1.38E-08	2.24E-06	Yes
2	*TSFM*	12	6.42	1.38E-10	6.70E-08	6.49	8.41E-11	1.91E-08	Yes
3	*POGLUT1*	3	4.76	1.98E-06	4.81E-04	3.89	9.84E-05	5.33E-03	Yes
4	*TRAF3*	14	3.84	1.21E-04	1.18E-02	4.36	1.32E-05	1.15E-03	Yes
5	*DHRS11*	17	4.17	3.02E-05	4.40E-03	4.37	1.27E-05	1.15E-03	Yes
6	*SHMT1*	17	3.82	1.32E-04	1.20E-02	4.16	3.19E-05	2.42E-03	Yes
7	*GIMAP4*	7	3.97	7.27E-05	8.15E-03	3.91	9.08E-05	5.16E-03	Yes
8	*MTHFR*	1	3.79	1.52E-04	1.30E-02	4.77	1.81E-06	2.29E-04	Yes
9	*ICA1L*	2	−3.92	8.89E-05	9.25E-03	−4.48	7.40E-06	8.41E-04	Yes
10	*HMGCL*	1	3.71	2.04E-04	1.47E-02	3.71	2.04E-04	9.55E-03	Yes
11	*FAM120B*	6	4.05	5.10E-05	6.19E-03	4.05	5.10E-05	3.22E-03	Yes
12	*DOC2A*	16	4.43	9.28E-06	1.69E-03	4.15	3.40E-05	2.42E-03	Yes
13	*TRIM47*	17	−3.68	2.35E-04	1.56E-02	−3.46	5.38E-04	2.18E-02	Yes
14	*LRP4*	11	3.47	5.12E-04	2.82E-02	3.71	2.10E-04	9.55E-03	Yes
15	*WARS*	14	3.49	4.87E-04	2.82E-02	3.98	6.81E-05	4.08E-03	Yes
16	*GALC*	14	3.44	5.93E-04	2.98E-02	4.37	1.25E-05	1.15E-03	Yes
17	*AUH*	9	−3.35	8.14E-04	3.71E-02	−3.43	5.97E-04	2.34E-02	Yes
18	*PANK4*	1	−3.30	9.67E-04	4.03E-02	−4.10	4.05E-05	2.71E-03	Yes
19	*C4A*	6	7.83	4.93E-15	3.59E-12	−14.16	1.65E-45	9.38E-43	No[a]
20	*HLA-DRB5*	6	20.18	1.57E-90	2.29E-87	-	-	-	No
21	*PREX1*	20	5.30	1.15E-07	3.35E-05	-	-	-	No
22	*TMEM160*	19	−4.61	4.05E-06	8.43E-04	-	-	-	No
23	*LMAN2*	5	4.19	2.85E-05	4.40E-03	-	-	-	No
24	*LLGL1*	17	4.14	3.49E-05	4.62E-03	-	-	-	No
25	*FUCA1*	1	−3.71	2.08E-04	1.47E-02	-	-	-	No
26	*RAB24*	5	−3.71	2.08E-04	1.47E-02	-	-	-	No
27	*FKBP2*	11	−3.70	2.12E-04	1.47E-02	-	-	-	No
28	*PRICKLE1*	12	−3.52	4.29E-04	2.72E-02	-	-	-	No
29	*PDE2A*	11	−3.48	4.98E-04	2.82E-02	-	-	-	No
30	*SLC44A2*	19	−3.47	5.22E-04	2.82E-02	-	-	-	No
31	*CARM1*	19	−3.45	5.51E-04	2.87E-02	-	-	-	No
32	*WBP2*	17	−3.39	6.90E-04	3.35E-02	-	-	-	No
33	*STX1A*	7	−3.38	7.30E-04	3.43E-02	-	-	-	No
34	*SH3GL1*	19	−3.31	9.20E-04	4.03E-02	-	-	-	No
35	*TYW5*	2	3.30	9.60E-04	4.03E-02	-	-	-	No
36	*C4B*	6	-	-	-	−16.92	3.28E-64	3.73E-61	No
37	*PSMB9*	6	-	-	-	7.63	2.28E-14	8.64E-12	No
38	*IFI30*	19	-	-	-	7.08	1.40E-12	3.98E-10	No
39	*OS9*	12	-	-	-	6.04	1.57E-09	2.98E-07	No
40	*OGFOD2*	12	-	-	-	5.01	5.48E-07	7.79E-05	No
41	*CHCHD2*	7	-	-	-	−4.33	1.46E-05	1.19E-03	No
42	*ACOX1*	17	-	-	-	3.83	1.31E-04	6.77E-03	No
43	*SMIM8*	6	-	-	-	3.72	1.99E-04	9.55E-03	No
44	*MERTK*	2	-	-	-	−3.60	3.21E-04	1.40E-02	No
45	*MADD*	11	-	-	-	3.48	5.07E-04	2.14E-02	No
46	*ATXN3*	14	-	-	-	3.39	7.10E-04	2.69E-02	No
47	*ALG11*	13	-	-	-	3.35	8.04E-04	2.95E-02	No
48	*RAB5C*	17	-	-	-	3.27	1.07E-03	3.80E-02	No
49	*ARHGEF25*	12	-	-	-	3.25	1.15E-03	3.96E-02	No
50	*IQGAP1*	15	-	-	-	−3.20	1.38E-03	4.61E-02	No
51	*ARF4*	3	-	-	-	3.17	1.51E-03	4.91E-02	No

The table provides the z-scores for proteome-wide significant genes (PWAS FDR P < 0.05) with their corresponding P values and FDR-adjusted P values in the ROSMAP and Banner PWAS. A total of 51 proteome-wide significant associations were identified in ROSMAP discovery and Banner confirmation PWAS, among which 18 PWAS associations could be replicated, and 32 existed only in one dataset.
[a]C4A was not regarded as replicated due to the z-scores in opposite directions in the two-stage PWAS.

Table 2. Bayesian colocalization analysis found 11 risk genes with pQTL/GWAS signals driven by a shared causal SNP, 6 of which overlapped in ROSMAP dataset and Banner dataset.

	Gene	Chromosome	ROSMAP		Banner		Evidence for confirmation
			H_4[a]	Causal variant	H_4	Causal variant	
1	*ZC2HC1A*	8	0.987	Yes	0.980	Yes	Yes
2	*TSFM*	12	0.957	Yes	0.974	Yes	Yes
3	*SHMT1*	17	0.965	Yes	0.963	Yes	Yes
4	*ICA1L*	2	0.869	Yes	0.877	Yes	Yes
5	*HMGCL*	1	0.757	Yes	0.757	Yes	Yes
6	*FAM120B*	6	0.913	Yes	0.808	Yes	Yes
7	*TRAF3*[b]	14	0.991	Yes	-	-	No
8	*DOC2A*[b]	16	0.707	Yes	-	-	No
9	*MTHFR*[b]	1	0.707	Yes	-	-	No
10	*WARS*[c]	14	-	-	0.791	Yes	No
11	*DHRS11*[c]	17	-	-	0.751	Yes	No

The table provides the results of Bayesian colocalization analysis for 11 genes that found causal variants, 6 of which overlapped in ROSMAP and Banner datasets.

[a]H_4 is the Bayesian posterior probability based on hypothesis that pQTL and GWAS signals are regulated by a shared causal variant.

[b]Genes without evidence of colocalization in the Banner dataset.

[c]Genes without evidence of colocalization in the ROSMAP dataset.

Bayesian colocalization analysis, the results are as follows (Table 2). In ROSMAP dataset, we observed strong posterior probability ($H_4 > 0.7$) for hypothesis 4 (H_4, pQTL/GWAS signals driven by a shared causal variant) in 9 genes, suggesting MS risk may be mediated by causal genetic variants that have effects on protein abundance. In Banner dataset, the strong H_4 was found for 8 genes. Notably, a total of 6 genes including *ZC2HC1A, TSFM, SHMT1, ICA1L, HMGCL* and *FAM120B* were identified independently in both ROSMAP and Banner datasets, indicating associations between these genes and MS risk were of high confidence. In general, 6 genes out of 18 PWS genes overlapped in the two-stage PWAS provided evidence of colocalization. We defined these 6 genes as candidate risk genes.

Differential expression analysis validated dysregulation of SHMT1, FAM120B and ICA1L in MS cases

To validate if the 6 candidate risk genes (*ZC2HC1A, TSFM, SHMT1, ICA1L, HMGCL* and *FAM120B*) were dysregulated at transcriptional level in MS cases, we examined the expression in MS cases comparing with healthy controls.

We examined the expression in white matter and gray matter to explore the specificity of risk genes in brain parenchyma. In the white matter, *SHMT1* (P_{FDR} = 4.82E-02) and *FAM120B* (P_{FDR} = 8.13E-04) were significantly up-regulated in effect direction consistent with PWAS (positive z-score in PWASs means up-regulated, negative means down-regulated) (Fig. 2A,C). In the gray matter, *ICA1L* (P_{FDR} = 3.44E-02) was considerably down-regulated (Fig. 2F). Besides, we also examined the expression of genes other than candidate risk genes that had one co-localized signal and found *DOC2A* (P_{FDR} = 2.56E-02), *WARS* (P_{FDR} = 2.72E-05) and *MTHFR* (P_{FDR} = 1.11E-03) were also significantly up-regulated in MS as compared to controls (Table 3, Fig. S1). Afterwards, we examined the expression in lesion and NAWM/NAGM respectively comparing with healthy controls to further investigate the dysregulation of *SHMT1*, *FAM120B* and *ICA1L* in tissue blocks with varying degrees of pathological changes. According to the Student's T Test results, *SHMT1* (P = 2.51E-02) was dramatically up-regulated in white matter lesions instead of NAWM comparing with healthy controls (Fig. 2B). Further investigation of specific lesion types revealed *SHMT1* was mainly up-regulated in active lesion (P = 4.00E-02) and inactive lesion (P = 2.61E-02) (Fig. S2). Whereas the up-regulation of *FAM120B* could be observed in both lesion (P = 2.83E-04) and NAWM (P = 1.19E-03) comparing to controls (Fig. 2D,E). Similarly, *ICA1L* was dramatically down-regulated in both lesions (P = 4.41E-03) and NAGM (P = 1.70E-02) comparing to controls (Fig. 2G,H). Dysregulation of *SHMT1*, *FAM120B* and *ICA1L* was confirmed at transcription level, further indicating these risk genes affected protein abundance via the process of transcription and confer MS risk.

Discussion

Elucidating the pathogenesis of disease is a key goal of human genetics research, especially for neurodegenerative

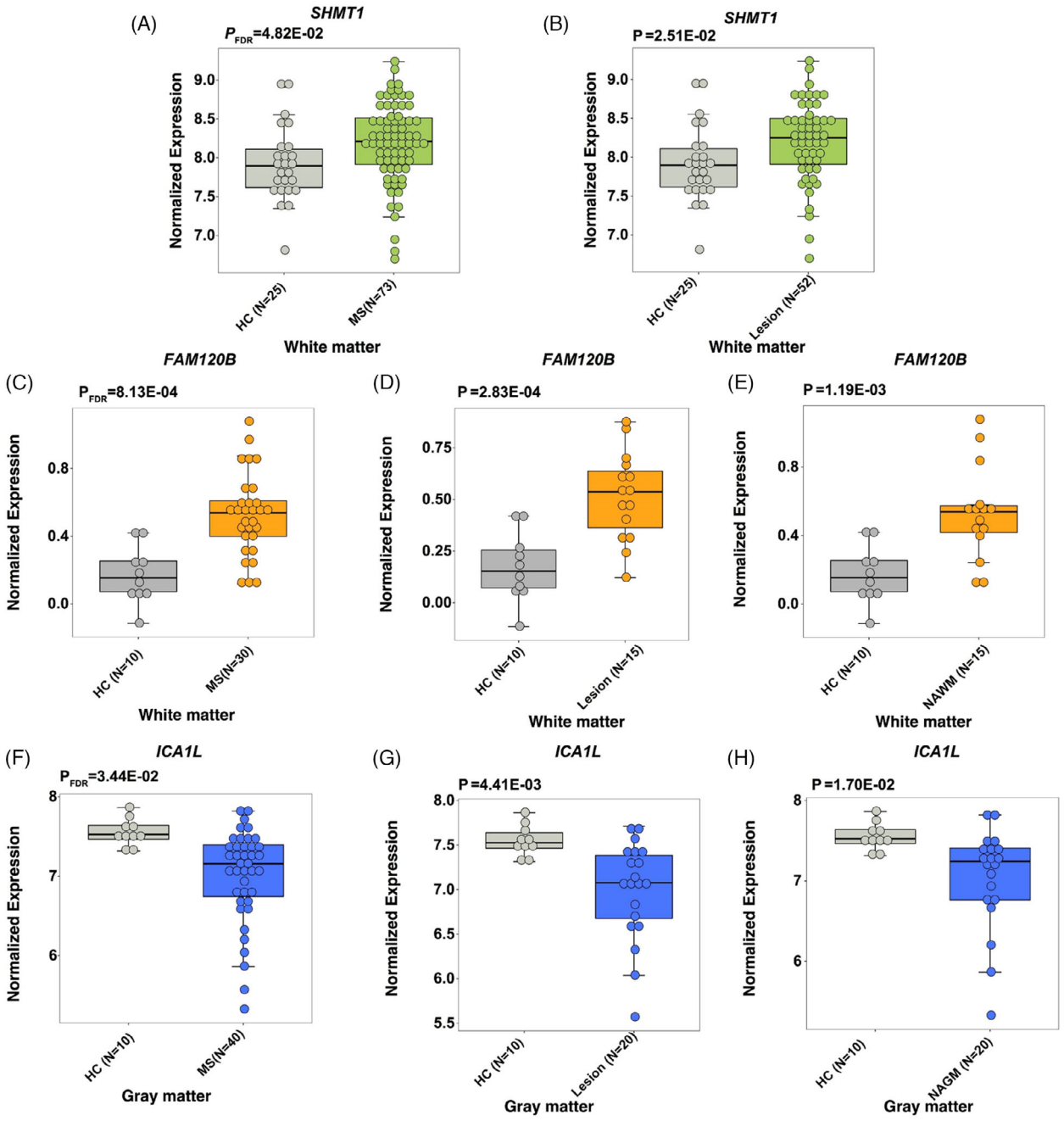

Figure 2. Differential expression analysis for *SHMT1, FAM120B* and *ICA1L* validated dysregulation of risk genes at transcription level. The boxplot shows the differential expression analysis results of *SHMT1, FAM120B* and *ICA1L* out of 6 candidate risk genes verified by coloc. Firstly, comparisons of expression between MS tissue blocks and controls were conducted in white matter and gray matter. Next, we further explored dysregulation in different lesion types, including lesion, NAWM/NAGM and control. (A) *SHMT1* was significantly up-regulated in white matter in MS cases comparing with healthy controls; (B) Further investigation of subgroups showed *SHMT1* was mainly up-regulated in white matter lesion instead of NAWM when comparing with healthy controls; (C) *FAM120B* was significantly up-regulated in white matter in MS cases comparing with healthy controls; (D) Further investigation revealed *FAM120B* was up-regulated in white matter lesion as compared to healthy controls; (E) Further investigation revealed *FAM120B* was up-regulated in NAWM as compared to healthy controls; (F) *ICA1L* was significantly down-regulated in gray matter in MS cases comparing with healthy controls; (G) Further investigation revealed *ICA1L* was down-regulated in gray matter lesion comparing with healthy controls; (H) Further investigation revealed *ICA1L* was down-regulated in NAGM comparing with healthy controls. MS, multiple sclerosis; HC, healthy control; NAWM, normal appearing white matter; NAGM, normal appearing gray matter.

Table 3. Overview of the analysis performed and genes identified in the study.

		PWAS		Colocalization analysis		Differential expression analysis
	Gene	Discovery	Confirmation	Discovery	Confirmation	Brain
1	*SHMT1*	Yes	Yes	Yes	Yes	Yes[a]
2	*FAM120B*	Yes	Yes	Yes	Yes	Yes[b]
3	*ICA1L*	Yes	Yes	Yes	Yes	Yes[c]
4	*MTHFR*	Yes	Yes	Yes	No	Yes
5	*DOC2A*	Yes	Yes	Yes	No	Yes
6	*WARS*	Yes	Yes	No	Yes	Yes

[a]Global white matter lesions.
[b]Global white matter including lesions and normal appearing white matter.
[c]Cortical gray matter lesions and normal appearing gray matter from precentral gyrus.

diseases like MS whose pathogenesis still remains obscure. In this study, we performed a two-stage PWAS by integrating human brain pQTL data and genome-wide associations, identifying 51 MS risk genes whose protein abundance levels were associated with disease phenotype. By comparing the results of PWAS in the discovery and replication stages, we identified 18 overlapped MS risk genes (14 genes up-regulated including *ZC2HC1A*, *TSFM, POGLUT1, TRAF3, DHRS11, SHMT1, GIMAP4, MTHFR, HMGCL, FAM120B, DOC2A, LRP4, WARS, GALC*, whereas 4 genes down-regulated including *ICA1L, TRIM47, AUH* and *PANK4*). These genes showed consistent associations with MS in the two-stage PWAS. The following integrative analysis of bayesian colocalization identified 6 genes with colocalization evidence in both ROSMAP and Banner datasets. Our results suggested that efficient integration of brain pQTLs and MS-associated GWAS signals facilitates the excavation for MS risk genes whose genetic variation confer MS risk through modulating protein abundance. Three MS risk genes of high-confidence including *SHMT1, ICA1L* and *FAM120B* were validated at transcription level through differential expression analysis.

In the present results, we not only replicated and highlights genes already identified in current MS studies such as *SHMT1*, but also identified novel risk genes like *FAM120B* and *ICA1L*. As for *SHMT1*, the gene encodes a serine hydroxymethyl transferase that plays an important role in the folate cycle. It catalyzes the transfer of carbon units for subsequent synthesis of nucleotides and methionine.[43,44] Previous GWAS studies have identified *SHMT1* as a novel susceptibility locus for MS.[45] We observed *SHMT1* significantly up-regulated in MS cases, which was consistent with effect direction from eQTL and mQTL analysis in previous publication.[45] Interestingly, we noticed that both *SHMT1* and *MTHFR* (a PWS MS risk gene identified in PWAS and provided colocalization evidence in ROSMAP dataset) are both key enzymes in the folate metabolic pathway. *MTHFR* encodes methylenetetrahydrofolate reductase, which catalyzes the conversion of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate. The association of *MTHFR* polymorphisms with MS has been validated in individuals of various descent.[46–48] Our findings indicated that the disorder of the one-carbon metabolic pathway may be closely associated with the pathogenesis of MS, and more researches are in need in the future.

FAM120B and *ICA1L* are novel MS risk genes. *FAM120B* has been confirmed as a transcriptional co-activator of PPARγ,[49] and PPARγ activation could suppress the inflammatory state of macrophages as previous study has shown.[50] Recent studies have related PPARγ downregulation with MS and emphasize PPARγ agonists as a promising treatment in MS.[51] What role the up-regulation of *FAM120B* play indeed in PPARγ pathway need further study in the future. *ICA1L* has been identified as a risk gene for cerebrovascular disease by several GWAS,[52–54] and recent PWAS found that diseases associated with *ICA1L* include in cerebrovascular disease.[55] A causal relationship between MS and a range of cardiovascular diseases has been found in a mendelian randomization study,[56] providing insights into the network of diseases and their interactions, which could attribute to some shared etiological pathways such as immune system-related inflammatory responses and their contribution to neurodegeneration.[57] In general, our findings, as well as the extant literature, revealed an important role for *SHMT1, ICA1L, FAM120B* in the central nervous system. They may confer MS risk by their protein-abundance related effects on specific metabolic pathways.

Our study found that risk gene dysregulation had specificity in distribution, as *SHMT1* and *FAM120B* were significantly up-regulated in the white matter, while *ICA1L* was considerably down-regulated in the gray matter. Several studies have explored differences of genetic effects between white and gray matter in MS and have reported the differences at both the proteome[13] and

transcriptome levels.[58,59] This may be due to the various pathophysiological mechanisms of demyelination in gray and white matter,[60] as white matter lesions are accompanied by activation of local glial cells and infiltration of peripheral leucocytes, whereas gray matter lesions show a lack of activated glial cells and few infiltration of peripheral leucocytes.[61] Another finding was that risk gene dysregulation was associated with the degree of pathological changes, as up-regulation of *SHMT1* could only be observed in macroscopic lesions comparing with healthy controls, but not in NAWM. Further investigation of lesion types found *SHMT1* was considerably up-regulated only in active lesions and inactive lesions, indicating its potential role in the early formation of plaque, which could be used for the MS progression prediction.[62] Meanwhile, dysregulation might exist extensively in different histological areas, since *FAM120B* was up-regulated in both NAWM and lesion comparing with controls, likewise *ICA1L* down-regulated in NAGM and lesion, suggesting that the presence of dysregulation may be widespread and persist from early pathologic stages to development of lesion.

The present study has several strengths. Firstly, to the best of our knowledge, this is the first PWAS study for MS that utilized pQTL derived from human dlPFC, taking into account factors of post-transcriptional regulation and translation levels. As only 3 (50%) of the 6 candidate risk genes provided evidence at mRNA level, our results reflected the essentiality for carrying out studies at protein level. Secondly, through differential expression analysis, we found that MS susceptiblity gene dysregulation has specificity in distribution (white matter and gray matter) and degree of lesions, which indicated the materials sources as an important issue to pay attention to in future research.

The study has several limitations. First, age and ethnic limitation between the GWAS and pQTL studies may be the key factor leading to deviation. Since most of the samples/summary statistics used in present study are of European descent, our conclusions may be constricted to specific population, and larger PWAS studies using cross-ethnic sample are in need for the future. Second, the small sample size included in the proteomic analysis resulted in the identification of a limited number of pQTLs. Therefore, the number of PWAS genes identified in this study was relatively small. Third, 32 significant PWAS associations could not be replicated, which means they only existed in one dataset (ROSMAP or Banner), indicating that more confirmation work to be done when using the integration methodology of PWAS. Finally, we recognized that not all candidate PWS risk genes with colocalized evidence were validated dysregulation in expression, and there are a variety of possible reasons: (1)Different tissue sources used in PWAS and differential expression analysis may have various expression characteristics of risk genes, since pQTLs were derived from dlPFC, while differential expression analysis was selected from lesions and nearby NAWM/NAGM. (2) Potential methodological differences in tissue collection, classification and analysis pipeline also affected the results.

Overall, we identified three genetic predictive effects (*SHMT1*, *FAM120B*, and *ICA1L*) across the genome using a distinct but integrated analysis pipeline (PWAS, Bayesian co-localization and differential expression analysis). These findings implicated causal biological pathways involved in the pathogenesis of MS, which illuminates the direction for further exploration in the future. More studies are warranted to discover the underlying mechanism of abnormal changes in these proteins and their related pathways in MS. Meanwhile, our results also prioritized unique protein biomarkers and potential therapeutic targets that could aid in MS diagnosis and advance the development of new intervention.

Acknowledgements

All authors are grateful for participation in our research. This work was partly funded by the China Postdoctoral Science Foundation (2020 M673247), Key R & D projects of Science and Technology Department of Sichuan Province (2021YFS0248), the Postdoctoral Foundation of West China Hospital (2020HXBH163), and College Students' innovation and entrepreneurship training program (C2022121154).

Conflict of Interest

Nothing to report.

Author Contributions

Conceptualization, C.Z.; formal analysis, T.J., Y.M., F.Q. and F.H.; writing—original draft preparation, T.J. and Y.M.; writing—review and editing, C.Z.; visualization, T.J., Y.M., F.Q. and F.H.; funding acquisition, C.Z. All authors have read and agreed to the published version of the manuscript.

References

1. Walton C, King R, Rechtman L, et al. Rising prevalence of multiple sclerosis worldwide: insights from the atlas of MS, third edition. Mult Scler. 2020;26(14):1816-1821.
2. Reich DS, Lucchinetti CF, Calabresi PA. Multiple sclerosis. N Engl J Med. 2018;378(2):169-180.

3. Fissolo N, Haag S, de Graaf KL, et al. Naturally presented peptides on major histocompatibility complex I and II molecules eluted from central nervous system of multiple sclerosis patients. Mol Cell Proteomics. 2009;8(9):2090-2101.
4. Ly L, Barnett MH, Zheng YZ, Gulati T, Prineas JW, Crossett B. Comprehensive tissue processing strategy for quantitative proteomics of formalin-fixed multiple sclerosis lesions. J Proteome Res. 2011;10(10):4855-4868.
5. Broadwater L, Pandit A, Clements R, et al. Analysis of the mitochondrial proteome in multiple sclerosis cortex. Biochim Biophys Acta. 2011;1812(5):630-641.
6. Maccarrone G, Nischwitz S, Deininger SO, et al. MALDI imaging mass spectrometry analysis-a new approach for protein mapping in multiple sclerosis brain lesions. J Chromatogr B Analyt Technol Biomed Life Sci. 2017;1047:131-140.
7. Högel H, Rissanen E, Barro C, et al. Serum glial fibrillary acidic protein correlates with multiple sclerosis disease severity. Mult Scler. 2020;26(2):210-219.
8. Severa M, Zhang J, Giacomini E, et al. Thymosins in multiple sclerosis and its experimental models: moving from basic to clinical application. Mult Scler Relat Disord. 2019;27:52-60.
9. Huang J, Khademi M, Fugger L, et al. Inflammation-related plasma and CSF biomarkers for multiple sclerosis. Proc Natl Acad Sci USA. 2020;117(23):12952-12960.
10. Mosleth EF, Vedeler CA, Liland KH, et al. Cerebrospinal fluid proteome shows disrupted neuronal development in multiple sclerosis. Sci Rep. 2021;11(1):4087.
11. Kroksveen AC, Opsahl JA, Guldbrandsen A, et al. Cerebrospinal fluid proteomics in multiple sclerosis. Biochim Biophys Acta. 2015;1854(7):746-756.
12. Malekzadeh A, Leurs C, van Wieringen W, et al. Plasma proteome in multiple sclerosis disease progression. Ann Clin Transl Neurol. 2019;6(9):1582-1594.
13. Rai NK, Singh V, Li L, Willard B, Tripathi A, Dutta R. Comparative proteomic profiling identifies reciprocal expression of mitochondrial proteins between White and gray matter lesions from multiple sclerosis brains. Front Neurol. 2021;12:779003.
14. Lee D, Rushworth MF, Walton ME, Watanabe M, Sakagami M. Functional specialization of the primate frontal cortex during decision making. J Neurosci. 2007;27(31):8170-8173.
15. Rushworth MF, Noonan MP, Boorman ED, Walton ME, Behrens TE. Frontal cortex and reward-guided learning and decision-making. Neuron. 2011;70(6):1054-1069.
16. Olah M, Patrick E, Villani AC, et al. A transcriptomic atlas of aged human microglia. Nat Commun. 2018;9(1):539.
17. Sanchez-Mut JV, Heyn H, Vidal E, et al. Whole genome grey and white matter DNA methylation profiles in dorsolateral prefrontal cortex. Synapse. 2017;71(6):e21959.
18. Westerlind H, Ramanujam R, Uvehag D, et al. Modest familial risks for multiple sclerosis: a registry-based study of the population of Sweden. Brain. 2014;137(Pt 3):770-778.
19. Brandes N, Linial N, Linial M. PWAS: proteome-wide association study-linking genes and phenotypes by functional variation in proteins. Genome Biol. 2020;21(1):173.
20. Wingo AP, Liu Y, Gerasimov ES, et al. Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer's disease pathogenesis. Nat Genet. 2021;53(2):143-146.
21. Liu J, Li X, Luo XJ. Proteome-wide association study provides insights into the genetic component of protein abundance in psychiatric disorders. Biol Psychiatry. 2021;90(11):781-789.
22. Ou YN, Yang YX, Deng YT, et al. Identification of novel drug targets for Alzheimer's disease by integrating genetics and proteomes from brain and blood. Mol Psychiatry. 2021;26(10):6065-6073.
23. Uribe-San-Martín R, Ciampi-Díaz E, Suarez-Hernández F, Vásquez-Torres M, Godoy-Fernández J, Cárcamo-Rodríguez C. Prevalence of epilepsy in a cohort of patients with multiple sclerosis. Seizure. 2014;23(1):81-83.
24. Morgen K, Sammer G, Courtney SM, et al. Evidence for a direct association between cortical atrophy and cognitive impairment in relapsing-remitting MS. Neuroimage. 2006;30(3):891-898.
25. Kutzelnigg A, Lucchinetti CF, Stadelmann C, et al. Cortical demyelination and diffuse white matter injury in multiple sclerosis. Brain. 2005;128(Pt 11):2705-2712.
26. Lucchinetti CF, Popescu BF, Bunyan RF, et al. Inflammatory cortical demyelination in early multiple sclerosis. N Engl J Med. 2011;365(23):2188-2197.
27. Popescu BF, Lucchinetti CF. Meningeal and cortical grey matter pathology in multiple sclerosis. BMC Neurol. 2012;12:11.
28. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. Science. 2019;365(6460):eaav7188.
29. Wingo AP, Fan W, Duong DM, et al. Shared proteomic effects of cerebral atherosclerosis and Alzheimer's disease on the human brain. Nat Neurosci. 2020;23(6):696-700.
30. Beach TG, Adler CH, Sue LI, et al. Arizona study of aging and neurodegenerative disorders and brain and body donation program. Neuropathology. 2015;35(4):354-389.
31. Gusev A, Ko A, Shi H, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet. 2016;48(3):245-252.
32. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10(5):e1004383.

33. Elkjaer ML, Frisch T, Reynolds R, et al. Molecular signature of different lesion types in the brain white matter of patients with progressive multiple sclerosis. Acta Neuropathol Commun. 2019;7(1):205.
34. Reynolds R, Roncaroli F, Nicholas R, Radotra B, Gveric D, Howell O. The neuropathological basis of clinical progression in multiple sclerosis. Acta Neuropathol. 2011;122(2):155-170.
35. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-2120.
36. Anders S, Pyl PT, Huber W. HTSeq--a python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166-169.
37. Hendrickx DAE, van Scheppingen J, van der Poel M, et al. Gene expression profiling of multiple sclerosis pathology identifies early patterns of demyelination surrounding chronic active lesions. Front Immunol. 2017;8:1810.
38. Hendrickx DA, Koning N, Schuurman KG, et al. Selective upregulation of scavenger receptors in and around demyelinating areas in multiple sclerosis. J Neuropathol Exp Neurol. 2013;72(2):106-118.
39. van der Valk P, De Groot CJ. Staging of multiple sclerosis (MS) lesions: pathology of the time frame of MS. Neuropathol Appl Neurobiol. 2000;26(1):2-10.
40. Magliozzi R, Howell OW, Durrenberger P, et al. Meningeal inflammation changes the balance of TNF signalling in cortical grey matter in multiple sclerosis. J Neuroinflammation. 2019;16(1):259.
41. Magliozzi R, Howell OW, Reeves C, et al. A gradient of neuronal loss and meningeal inflammation in multiple sclerosis. Ann Neurol. 2010;68(4):477-493.
42. Weng L, Dai H, Zhan Y, He Y, Stepaniants SB, Bassett DE. Rosetta error model for gene expression analysis. Bioinformatics. 2006;22(9):1111-1121.
43. MacFarlane AJ, Liu X, Perry CA, et al. Cytoplasmic serine hydroxymethyltransferase regulates the metabolic partitioning of methylenetetrahydrofolate but is not essential in mice. J Biol Chem. 2008;283(38):25846-25853.
44. Herbig K, Chiang EP, Lee LR, Hills J, Shane B, Stover PJ. Cytoplasmic serine hydroxymethyltransferase mediates competition between folate-dependent deoxyribonucleotide and S-adenosylmethionine biosyntheses. J Biol Chem. 2002;277(41):38381-38389.
45. Andlauer TF, Buck D, Antony G, et al. Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation. Sci Adv. 2016;2(6):e1501678.
46. Cakina S, Ocak O, Ozkan A, Yucel S, Ozısık Karaman HI. Relationship between genetic polymorphisms MTHFR (C677T, A1298C), MTR (A2756G) and MTRR (A66G) genes and multiple sclerosis: a case-control study. Folia Neuropathol. 2019;57(1):36-40.
47. Dashti M, Ateyah K, Alroughani R, Al-Temaimi R. Replication analysis of variants associated with multiple sclerosis risk. Sci Rep. 2020;10(1):7327.
48. Naghibalhossaini F, Ehyakonandeh H, Nikseresht A, Kamali E. Association between MTHFR genetic variants and multiple sclerosis in a southern Iranian population. Int J Mol Cell Med. 2015;4(2):87-93.
49. Li D, Kang Q, Wang DM. Constitutive coactivator of peroxisome proliferator-activated receptor (PPARgamma), a novel coactivator of PPARgamma that promotes adipogenesis. Mol Endocrinol. 2007;21(10):2320-2333.
50. Wouters E, Grajchen E, Jorissen W, et al. Altered PPARγ expression promotes myelin-induced foam cell formation in macrophages in multiple sclerosis. Int J Mol Sci. 2020;21(23):9329.
51. Vallée A, Lecarpentier Y, Guillevin R, Vallée JN. Demyelination in multiple sclerosis: reprogramming energy metabolism and potential PPARγ agonist treatment approaches. Int J Mol Sci. 2018;19(4):1212.
52. Chung J, Marini S, Pera J, et al. Genome-wide association study of cerebral small vessel disease reveals established and novel loci. Brain. 2019;142(10):3176-3189.
53. Wang X, Mo X, Zhang H, Zhang Y, Shen Y. Identification of phosphorylation associated SNPs for blood pressure, coronary artery disease and stroke from genome-wide association studies. Curr Mol Med. 2019;19(10):731-738.
54. Traylor M, Persyn E, Tomppo L, et al. Genetic basis of lacunar stroke: a pooled analysis of individual patient data and genome-wide association studies. Lancet Neurol. 2021;20(5):351-361.
55. Cullell N, Gallego-Fábrega C, Cárcel-Márquez J, et al. ICA1L is associated with small vessel disease: a proteome-wide association study in small vessel stroke and intracerebral Haemorrhage. Int J Mol Sci. 2022;23(6):3161.
56. Yang F, Hu T, He K, Ying J, Cui H. Multiple sclerosis and the risk of cardiovascular diseases: a mendelian randomization study. Front Immunol. 2022;13:861885.
57. Paolini Paoletti F, Simoni S, Parnetti L, Gaetani L. The contribution of small vessel disease to neurodegeneration: focus on Alzheimer's disease, Parkinson's disease and multiple sclerosis. Int J Mol Sci. 2021;22(9):4958.
58. van Wageningen TA, Gerrits E, Brouwer N, et al. Distinct gene expression in demyelinated white and grey matter areas of patients with multiple sclerosis. Brain Commun. 2022;4(2):fcac005.
59. Chai K, Zhang X, Tang H, et al. The application of consensus weighted gene Co-expression network analysis to comparative transcriptome meta-datasets of multiple sclerosis in gray and White matter. Front Neurol. 2022;13:807349.
60. Mallucci G, Peruzzotti-Jametti L, Bernstock JD, Pluchino S. The role of immune cells, glia and neurons in white and gray matter pathology in multiple sclerosis. Prog Neurobiol. 2015;127-128:1-22.
61. Prins M, Schul E, Geurts J, van der Valk P, Drukarch B, van Dam AM. Pathological differences between white and grey matter multiple sclerosis lesions. Ann N Y Acad Sci. 2015;1351:99-113.

62. Frischer JM, Weigand SD, Guo Y, et al. Clinical and pathological insights into the dynamic nature of the white matter multiple sclerosis plaque. Ann Neurol. 2015;78 (5):710-721.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Differential expression analysis of genes other than candidate risk genes that had one co-localized signal revealed dysregulation of *DOC2A*, *WARS* and *MTHFR*. (A) *DOC2A* was significantly up-regulated in white matter in MS cases comparing with healthy controls. (B) *WARS* was dramatically up-regulated in white matter in MS cases comparing with healthy controls. (C) *MTHFR* was considerably up-regulated in white matter in MS cases comparing with healthy controls.

Figure S2. Further investigation of differential expression analysis revealed specific lesion types that could observe *SHMT1* dysregulation. The boxplot shows the differential expression analysis results of *SHMT1* in different lesion types, including active lesion and inactive lesion. There was no significant difference between other lesion types and healthy controls. (A) *SHMT1* was dramatically up-regulated in active lesion in MS cases comparing with healthy controls. (B) *SHMT1* was significantly up-regulated in inactive lesion in MS cases comparing with healthy controls. AL, active lesion; IL, inactive lesion.

