# Dysregulation of MS risk genes and pathways at distinct stages of disease

OPEN

Sundararajan Srinivasan, MSc
Marco Di Dario, MSc
Alessandra Russo, MSc
Ramesh Menon, PhD
Elena Brini, PhD
Marzia Romeo, MD
Francesca Sangalli, MD
Gloria Dalla Costa, MD
Mariaemma Rodegher, MD
Marta Radaelli, MD
Lucia Moiola, MD
Daniela Cantarella, MSc
Enzo Medico, PhD
Gianvito Martino, MD
Roberto Furlan, MD, PhD
Vittorio Martinelli, MD
Giancarlo Comi, MD
Cinthia Farina, PhD

Correspondence to Dr. Farina: farina.cinthia@hsr.it

## ABSTRACT

**Objective:** To perform systematic transcriptomic analysis of multiple sclerosis (MS) risk genes in peripheral blood mononuclear cells (PBMCs) of subjects with distinct MS stages and describe the pathways characterized by dysregulated gene expressions.

**Methods:** We monitored gene expression levels in PBMCs from 3 independent cohorts for a total of 297 cases (including clinically isolated syndromes (CIS), relapsing-remitting MS, primary and secondary progressive MS) and 96 healthy controls by distinct microarray platforms and quantitative PCR. Differential expression and pathway analyses for distinct MS stages were defined and validated by literature mining.

**Results:** Genes located in the vicinity of MS risk variants displayed altered expression in peripheral blood at distinct stages of MS compared with the healthy population. The frequency of dysregulation was significantly higher than expected in CIS and progressive forms of MS. Pathway analysis for each MS stage–specific gene list showed that dysregulated genes contributed to pathogenic processes with scientific evidence in MS.

**Conclusions:** Systematic gene expression analysis in PBMCs highlighted selective dysregulation of MS susceptibility genes playing a role in novel and well-known pathogenic pathways. ***Neurol Neuroimmunol Neuroinflamm* 2017;4:e337; doi: 10.1212/NXI.0000000000000337**

## GLOSSARY

**CIS** = clinically isolated syndromes; **cRNA** = complementary RNA; **DEG** = differentially expressed gene; **EAE** = experimental autoimmune encephalomyelitis; **FDR** = false discovery rate; **GWAS** = genome-wide association studies; **HC** = healthy control; **MHC** = major histocompatibility complex; **MS** = multiple sclerosis; **PBMCs** = peripheral blood mononuclear cells; **PP-MS** = primary progressive MS; **q-PCR** = quantitative PCR; **RR-MS** = relapsing-remitting MS; **RT-PCR** = real-time PCR; **SP-MS** = secondary progressive MS.

MS is an inflammatory demyelinating disease of the CNS with evidence of immune dysfunction.[1] The first clinical episode with features suggestive of MS is classified as clinically isolated syndrome (CIS), unless lesion dissemination in time and space ratifying MS diagnosis is evidenced.[2] Approximately, 85% of individuals develop the relapsing-remitting form of MS (RR-MS), whereas 15% of MS individuals experience a progressive course (primary progressive MS, PP-MS). After a variable time, most RR-MS subjects advance to a secondary progressive (SP) phase (SP-MS), where neurologic worsening takes place without periods of remission.[3] Several MS risk variants have been uncovered by genome-wide association studies (GWAS), and many of them are located close to immunologically relevant genes,[4] suggesting that immune dysfunction may be partly genetically determined. Although effective in highlighting MS risk alleles, GWAS have failed in dissecting the genetic components of the susceptibility to distinct MS clinical forms. In fact, MS is a multifactorial disorder determined by the complex interaction between genetic and environmental factors, whose integration occurs at the epigenetic level and

**Supplemental data at Neurology.org/nn**

From the Institute of Experimental Neurology (INSpe) (S.S., M.D.D., A.R., R.M., E.B., M. Romeo, F.S., G.D.C., M. Rodegher, M. Radaelli, L.M., G.M., R.F., V.M., G.C., C.F.), Division of Neuroscience, San Raffaele Scientific Institute, Milan; University Vita-Salute San Raffaele (S.S., E.B., G.C.), Milan; and Laboratory of Functional Genomics (D.C., E.M.), Institute for Cancer Research and Treatment (IRCC), University of Turin Medical School, Candiolo, Italy.

Funding information and disclosures are provided at the end of the article. Go to Neurology.org/nn for full disclosure forms. The Article Processing Charge was paid by Ospedale San Raffaele.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND), which permits downloading and sharing the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

 Copyright © 2017 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology 

determines gene expression. Our group has recently shown the importance of blood transcriptomics in uncovering gene expression changes and transcriptional regulators in MS.[5] In this study, we have (1) examined the expression levels of known MS susceptibility genes in peripheral blood mononuclear cells (PBMCs) of CIS, RR-MS, PP-MS, SP-MS, and control individuals, (2) identified a panel of blood transcriptional signatures for distinct MS forms, and (3) explored the pathways contributed by the dysregulated susceptibility genes.

## METHODS

**Human subjects and blood sampling.** Investigations were conducted according to the principles expressed in the Declaration of Helsinki, and peripheral blood was drawn after signing of the institutional informed consent. We recruited 142 patients with MS (46 CIS, 52 RR-MS, 23 PP-MS, and 21 SP-MS) and 40 healthy controls (HCs) of Italian origin for the generation of the main transcriptomic data set by Illumina platform. Demographic and clinical parameters are shown in table 1. A second distinct cohort comprising 21 RR-MS, 15 PP-MS, 13 SP-MS, and 27 HCs was enrolled according to the same inclusion criteria and used for validation with a distinct array platform (table 1). Finally, a third independent cohort comprising 31 CIS, 30 RR-MS, 24 PP-MS, 21 SP-MS, and 29 HCs was included and used for validation experiments by quantitative PCR (q-PCR) (table 1). Patients with MS were diagnosed according to McDonald criteria[6] and were not suffering from any other acute or chronic inflammatory diseases or other autoimmune disorders. Furthermore, they had not started any immunomodulatory therapy for MS yet, as recruitment was performed over the last 15 years, in a period when decision to treat was not established and widespread as nowadays. Blood sampling was performed between 30 and 90 days after the first clinical attack in patients with CIS, and at least 4 weeks after the last clinical attack or steroid treatment for RR-MS subjects. All participants had peripheral blood counts within the reference range. All blood samplings were performed between 9 and 12 AM

**PBMC isolation and RNA extraction.** PBMCs were isolated using a discontinuous density gradient (Lymphoprep; Nycomed, Oslo, Norway). Viable cells were counted by Trypan blue (Sigma-Aldrich, Milan, Italy) exclusion. Total RNA was extracted using Tri Reagent (Ambion; Applied Biosystems, Monza, Italy) and stored at −80°C.

**Generation of gene expression data sets.** RNA quality was checked using Bioanalyzer 2100 (Agilent, Milan, Italy). Complementary RNA (cRNA) synthesis was performed using the Illumina TotalPrep RNA Amplification Kit (Ambion) according to the manufacturer's protocol. Hybridization of the cRNA relative to the first casistics was performed on Illumina Human Ref-8 v2 arrays (Illumina, Son, Netherlands). GenomeStudio GX Software (Illumina, San Diego, CA) was used to extract the array raw data. All the raw data were background subtracted by NEC method and normalized using cubic spline normalization as implemented in the software. We performed batch correction using COMBAT to remove batch effects in the PBMC expression data.[7] The second independent transcriptomic data set was generated by Affymetrix GeneChip Human Genome U133 Plus 2.0 Array. All the necessary preprocessing steps were performed using robust multiarray average algorithm present in the Affy package in Bioconductor and keeping genes with multiple probes as individual transcripts. The MIAMI compliant microarray data have been deposited in the EBI ArrayExpress database

**Table 1** Subjects demographics and clinical information

| | HC | CIS | RR-MS | PP-MS | SP-MS |
|---|---|---|---|---|---|
| **Cohort for Illumina arrays** | | | | | |
| No. (female:male) | 40 (20:20) | 46 (23:23) | 52 (32:20) | 23 (09:14) | 21 (11:10) |
| Age, y | 33.3 ± 10.44 | 34.0 ± 10.17 | 37.4 ± 10.16 | 52.1 ± 11.70 | 54.3 ± 11.10 |
| Disease duration, y | — | — | 7.4 ± 6.8 | 12.41 ± 8.58 | 24.6 ± 9.27 |
| EDSS score | — | — | 1.72 ± 0.97 | 5.2 ± 2.08 | 6.7 ± 1.3 |
| **Cohort for Affymetrix arrays** | | | | | |
| No. (female:male) | 27 (13:14) | — | 21 (12:09) | 15 (09:06) | 13 (09:04) |
| Age, y | 42.4 ± 9.1 | — | 41.2 ± 8.1 | 52 ± 12.8 | 41.8 ± 9.3 |
| Disease duration, y | — | — | 4.8 ± 6.0 | 15.9 ± 8.7 | 15.4 ± 8.0 |
| EDSS score | — | — | 2.3 ± 1.1 | 6.3 ± 1.2 | 5.6 ± 1.2 |
| **Cohort for q-PCR assays** | | | | | |
| No. (female:male) | 29 (13:16) | 31 (19:12) | 30 (16:14) | 24 (12:12) | 21 (15:06) |
| Age, y | 33.7 ± 9.31 | 32.7 ± 8.05 | 36.8 ± 8.57 | 46.7 ± 7.96 | 51.2 ± 11.26 |
| Disease duration, y | — | — | 4.4 ± 6.2 | 8.5 ± 9.23 | 18.9 ± 6.52 |
| EDSS score | — | — | 1.64 ± 0.97 | 5.15 ± 1.54 | 5.34 ± 1.46 |

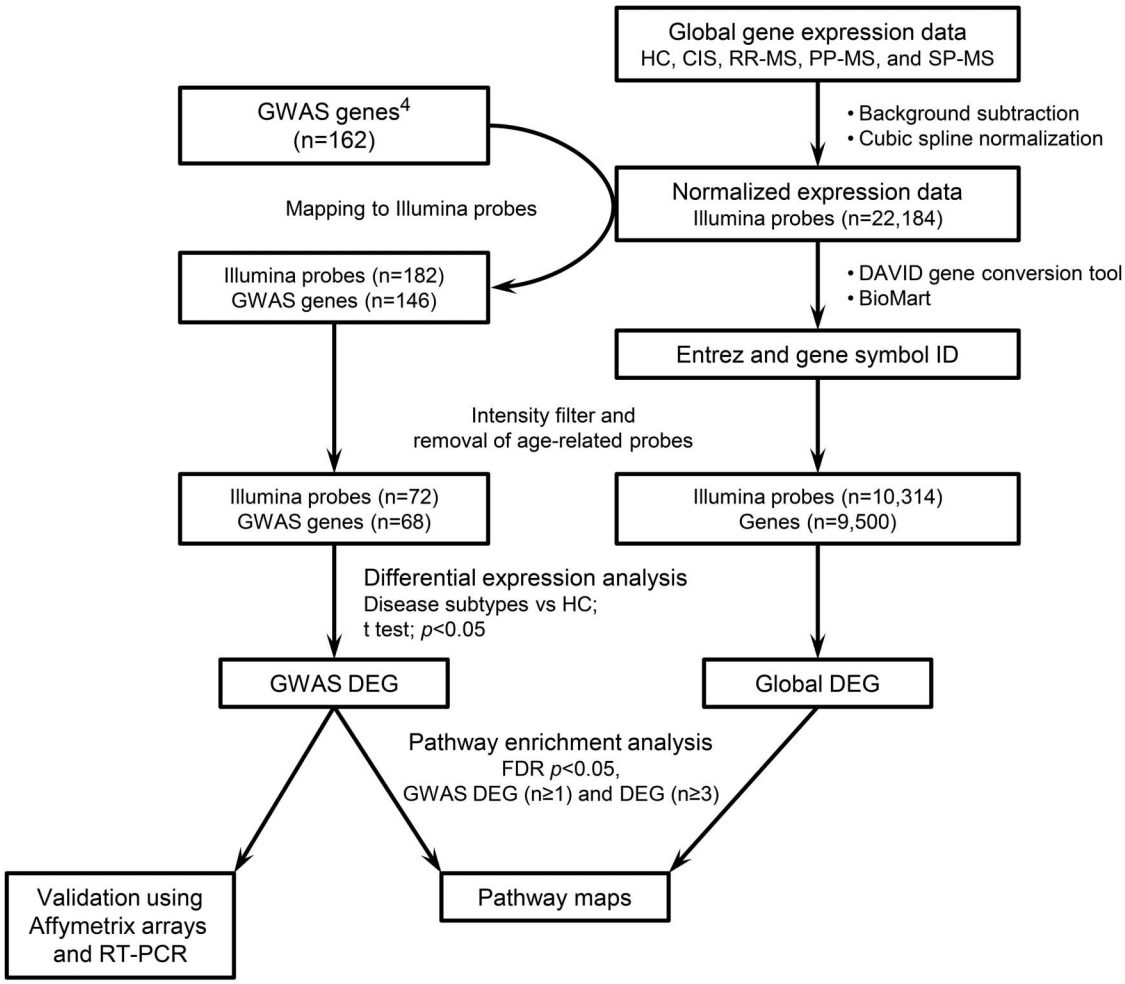
Abbreviations: CIS = clinically isolated syndrome; EDSS = Expanded Disability Status Scale; HC = healthy control; MS = multiple sclerosis; PP-MS = primary progressive MS; RR-MS = relapse-remitting MS; q-PCR = quantitative PCR; SP-MS = secondary progressive MS.
Numbers refer to mean ± SD.

(Accession numbers E-MTAB-4890 and E-MTAB-5151). Probes with a mean intensity value lower than 100 in all experimental groups were filtered out. Furthermore, we applied Pearson correlation to identify and remove probes correlating with age in the healthy population of the Illumina study, where age ranged from 22 to 57 years. To identify differentially expressed genes (DEGs) between each clinical form and the healthy population, we performed a 2-sample *t* test with a *p* value threshold of 0.05. A detailed flow diagram of this study is depicted in figure 1. To verify whether stage-specific transcriptomes were enriched in dysregulated risk genes, we measured the frequency of differential expression of GWAS genes and compared it to that of DEG in the global transcriptome by $\chi^2$ test with Yates' correction in GraphPad.

**Collection of GWAS genes and their mapping to gene expression data.** Genes located in the vicinity of MS susceptibility loci were collected from the supplementary tables 3 and 8 published by the International Multiple Sclerosis Genetics Consortium[4] and reporting 48 gene variants from previous GWAS studies[8–10] and 110 non–major histocompatibility complex (MHC) susceptibility variants described in 2013.[4] Figure e-1 at Neurology.org/nn summarizes the complete list of GWAS genes considered in our study. To map MS susceptibility genes to the relative Illumina probe identifiers, all the filtered probes in the Illumina data set were submitted to DAVID Gene Accession Conversion Tool,[11] using Illumina HumanRef-8 v2 as a background reference so to retrieve Entrez and gene symbol identifiers for each probe. The same probe list was submitted to BioMart ID,[12] conversion package in R Bioconductor,[13] as a further control. In this way, gene expression levels were retrieved for 182 Illumina probes corresponding to 146 GWAS genes (figure 1). Similarly, GWAS genes were annotated to the corresponding Affymetrix probes using BioMart ID,[12] conversion package in R Bioconductor,[13] and gene expression levels were retrieved for 344 Affymetrix probes relative to 149 MS susceptibility genes.

**cDNA synthesis and real-time PCR.** The purity and the quality of RNA were analyzed using NanoDrop8000 (Thermo-Scientific, Waltham, MA). Two-microgram aliquot of total RNA was reverse transcribed to cDNA using random primers and Superscript III reverse transcriptase (all from Invitrogen, Carlsbad, CA) following the manufacturer's protocol. Real-time (RT-PCR) was performed using TAQMan Universal Master Mix (Applied BioSystems). PSMB1 was used as housekeeping gene based on its low coefficient of variation in the Illumina transcriptomic data set. Transcript levels of the target gene CD86 were measured by the TAQMan assay corresponding to the gene sequence detected by the Illumina probe and expressed

**Figure 1** Flow chart of the study




CIS = clinically isolated syndrome; DEG = differentially expressed gene; FDR = false discovery rate; GWAS = genome-wide association studies; HC = healthy control; MS = multiple sclerosis; PP-MS = primary progressive MS; RR-MS = relapsing-remitting MS; RT-PCR = real-time PCR; SP-MS = secondary progressive MS.

as percentage of the housekeeping gene. Normality of data distribution was assessed by Kolmogorov-Smirnov statistics, and nonparametric Mann-Whitney test was used to compare mean values. All $p$ values were 2-sided and subjected to a significance level of 0.05.

**Pathway analysis and literature mining.** Pathway enrichment analysis was performed starting from the distinct dysregulated expression profiles isolated in each clinical form of disease. MS specialty modules in MetaCore (Thomson Reuters, New York, NY) are manually curated pathways associated with MS pathology, which are then ranked based on statistical significance. We selected pathways passing the false discovery rate (FDR)-corrected $p$ value threshold of 0.05 and containing 3 DEGs and at least 1 GWAS DEG. For literature mining, PubMed database was used to collect scientific evidence about the role of the identified biological pathways in MS pathogenesis by searching the name of the pathway and "Multiple Sclerosis" or "EAE."

## RESULTS

**Transcriptomic analysis in MS.** Genes located in the vicinity of MS susceptibility loci (hereon called as GWAS genes) were collected from the supplementary tables 3 and 8 published by the International Multiple Sclerosis Genetics Consortium[4] (see also figure e-1). To examine the expression levels of the 162 GWAS genes, we generated high-throughput gene expression data from PBMC of CIS, RR-MS, PP-MS, SP-MS individuals, and HCs (table 1) by Illumina arrays, recovered the gene symbols for GWAS genes, and mapped them to the relative Illumina probes. As shown in figure 1, there were 72 probes for 68 expressed GWAS genes passing signal intensity filter. Next, we tested whether these probes were differentially expressed in any disease group compared with healthy controls. We observed 29 GWAS DEGs in CIS, 19 in RR-MS, 35 in PP-MS, and 35 in SP-MS (table 2). Notably, 6 GWAS DEGs were concordantly dysregulated in all disease groups compared with HCs, and several other GWAS DEGs were common to at least 2 MS forms (figure 2A). Among them, NDFIP1, AHI1, FCRL1, and STAT4 were concordantly regulated in CIS and RR-MS but not in progressive forms of disease, whereas 11 MS risk genes, including TYK2, IL7R, and TNFRSF1A, were differentially expressed in PP-MS and SP-MS but not in CIS and RR-MS subjects (figure 2A). Finally, some GWAS transcripts were altered at single stages of disease, suggesting distinct pathogenic alterations in distinct disease courses (figure 2A). To verify whether MS transcriptome was enriched in dysregulated GWAS genes, we measured the frequency of differential gene expression in the list of expressed GWAS genes and in the global transcriptome for each disease subtype (figure 1, table 2) and verified whether the frequency of GWAS DEG was significantly higher in the GWAS gene list than the expected frequency of dysregulation in a random selection of transcripts using $\chi^2$ statistics. As shown in table 2, we found a clear and significant enrichment of GWAS DEG in CIS, PP-MS, SP-MS, and not in RR-MS, demonstrating that the frequency of dysregulation in genes located in the vicinity of MS susceptibility variants is higher than expected in at least 3 distinct stages of disease. Then, we checked the expression of the identified GWAS DEG in PBMC of an independent cohort of subjects including 21 RR-MS, 15 PP-MS, 13 SP-MS, and 27 HCs, for which gene expression was generated using the Affymetrix array platform (table 1). After applying the same statistical threshold of the Illumina data set, several dysregulated GWAS DEGs were identified for each MS form in the Affymetrix data set. In fact, a total of 26 GWAS DEGs were concordantly upregulated (or downregulated) compared with the healthy population in 2 independent transcriptomic data sets (figure 2B). Finally, we used a third case-control cohort to validate the expression of CD86, one of the GWAS genes upregulated in all forms of disease in the Illumina and Affymetrix data sets, by q-PCR (figures 2C and e-2).

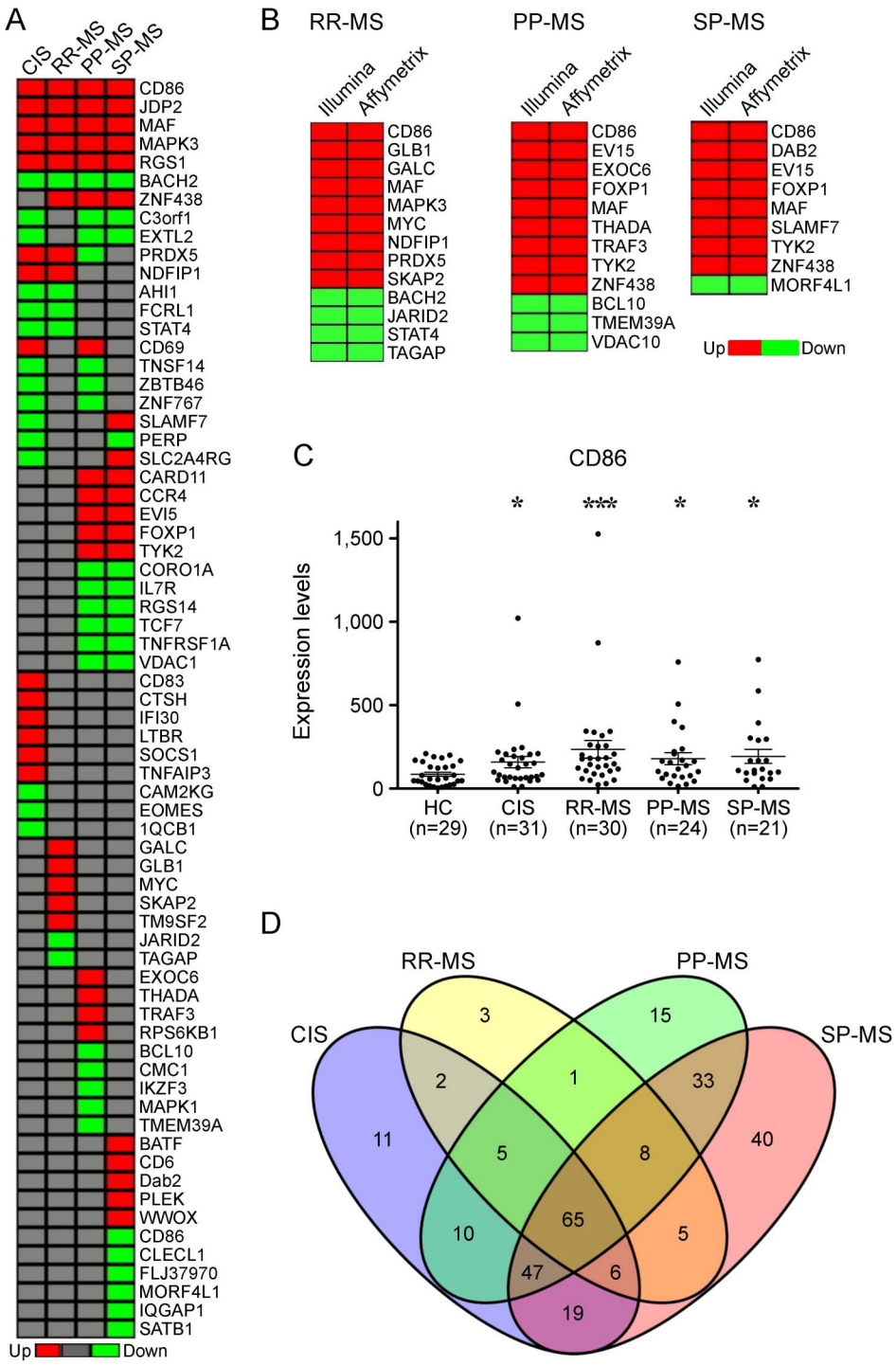
**Biological pathways contributed by MS GWAS genes.** To verify whether the dysregulated GWAS genes were involved in specific biological pathways, we applied the MetaCore pathway analysis program which uses a curated annotation database of biological functions and identifies the significantly enriched pathways for a given DEG list. Critically considering the limited

**Table 2** Probability of GWAS DEG enrichment at distinct stages of MS

| | GWAS DEG | Expressed GWAS genes | GWAS DEG, % | Global DEG | Expressed global genes | Global DEG, % | $\chi^2$ | $p$ Value |
|---|---|---|---|---|---|---|---|---|
| CIS | 29 | 72 | 40.2 | 2,649 | 10,314 | 25.68 | 7.39 | 0.006 |
| RR-MS | 19 | 72 | 26.3 | 2,442 | 10,314 | 23.67 | 0.16 | 0.686 |
| PP-MS | 35 | 72 | 48.6 | 3,487 | 10,314 | 33.08 | 6.44 | 0.011 |
| SP-MS | 35 | 72 | 48.6 | 3,323 | 10,314 | 32.21 | 19.08 | 0.0001 |

Abbreviations: CIS = clinically isolated syndrome; DEG = differentially expressed gene; GWAS = genome-wide association studies; HC = healthy control; MS = multiple sclerosis; PP-MS = primary progressive MS; RR-MS = relapsing-remitting MS; SP-MS = secondary progressive MS.

**Figure 2** **Dysregulated MS risk genes and pathways at distinct stages of disease**




(A) MS susceptibility genes dysregulated at distinct MS stages compared with as measured by Illumina microarrays. (B) Dysregulated MS susceptibility genes validated in 2 distinct array platforms. For A and B, fold-change heatmaps represent upregulated (red), downregulated (green), and unchanged (gray) expressions in the disease group compared with the healthy population. (C) Real-time PCR validation of CD86 transcript in peripheral blood mononuclear cells of a third case-control cohort. Refer also to figure e-2 for data elaboration after removal of a few outliers (samples with expression levels above 500). *$p < 0.05$, ***$p < 0.005$. (D) Venn diagram representing number of common and unique enriched pathways for each disease stage. CIS = clinically isolated syndrome; HC = healthy control; MS = multiple sclerosis; PP-MS = primary progressive MS; RR-MS = relapsing-remitting MS; SP-MS = secondary progressive MS.

number of dysregulated GWAS genes for a pathway analysis, for each disease group, we enriched the list of GWAS DEG with those dysregulated genes which were not located in the vicinity of MS susceptibility variants, performed pathway analysis using the global DEG list as input, and selected those pathways passing

**Table 3** Shared pathways across MS stages

| | | FDR *p* value | | | | |
|---|---|---|---|---|---|---|
| | **Pathways** | **CIS** | **RR-MS** | **PP-MS** | **SP-MS** | **PubMed ID** |
| **Transcription and translation** | Regulation of EIF2 activity | 0.0161 | 0.0146 | 2.97E-04 | 2.17E-05 | 15160394 |
| | Regulation of EIF4F activity | 1.49E-03 | 2.08E-04 | 1.43E-03 | 4.86E-03 | |
| | CREB pathway | 4.58E-03 | 0.0449 | 4.30E-03 | 5.90E-03 | 11571785 |
| **Cell signaling** | S1P1 receptor signaling via beta-arrestin | 8.46E-03 | 3.33E-03 | 2.68E-04 | 4.50E-04 | |
| | G-protein alpha-12 signaling pathway | 2.17E-03 | 0.0108 | 9.41E-03 | 1.71E-03 | |
| | G-CSF signaling | 9.89E-03 | 9.89E-03 | 7.47E-04 | 5.90E-03 | 25559893 |
| | A2A receptor signaling | 4.99E-03 | 0.0255 | 9.99E-03 | 2.15E-03 | 22529293 |
| | A3 receptor signaling | 0.0219 | 0.0231 | 0.0496 | 1.21E-03 | 26960979 |
| | IP3 signaling | 0.0219 | 0.0449 | 0.0234 | 5.87E-04 | 27378687 |
| | Activation of PKC via G-protein–coupled receptor | 6.16E-03 | 0.0307 | 1.19E-03 | 6.05E-05 | 26205402 |
| | LTBR1 signaling | 0.0143 | 8.23E-03 | 9.99E-03 | 1.34E-03 | 16870269 |
| | IL-33 signaling pathway | 1.00E-03 | 0.0467 | 0.0304 | 0.0334 | 22189043 |
| | C5a signaling | 2.25E-03 | 0.0255 | 0.0124 | 1.34E-03 | 19850104 |
| | Fc epsilon RI pathway | 7.79E-04 | 5.72E-03 | 4.72E-03 | 3.00E-03 | 11841694 |
| | IGF-1 receptor signaling | 9.7E-05 | 2.07E-04 | 9.7E-06 | 6.05E-05 | 24718491 |
| | Lipid metabolism_Insulin signaling | 0.0176 | 8.23E-03 | 5.14E-04 | 1.47E-04 | |
| | Receptor-mediated HIF regulation | 0.0161 | 3.18E-03 | 4.79E-03 | 6.15E-03 | 18056737 |
| | Insulin regulation of translation | 4.47E-03 | 5.22E-03 | 5.09E-04 | 3.99E-04 | |
| | PTEN pathway | 6.82E-03 | 0.0155 | 2.55E-04 | 3.58E-04 | 18349128 |
| **Cell proliferation** | EGFR signaling pathway | 2.68E-04 | 0.0109 | 2.68E-04 | 7.87E-04 | 26805386 |
| | Nucleocytoplasmic transport of CDK/cyclins | 4.99E-03 | 0.0336 | 0.012 | 3.58E-03 | |
| | Erk interactions: inhibition of Erk | 0.0215 | 3.33E-03 | 6.54E-04 | 0.0359 | 16728393 |
| | Beta-adrenergic receptors transactivation of EGFR | 0.011 | 0.0278 | 1.20E-03 | 0.011 | 15342205 |
| **Cancer** | Signaling cascades in multiple myeloma cells | 5.30E-04 | 8.37E-03 | 4.42E-04 | 1.50E-03 | |
| | IGF family signaling in colorectal cancer | 6.82E-03 | 8.23E-03 | 2.55E-04 | 1.34E-03 | |
| | Androgen receptor in prostate cancer | 0.0132 | 7.36E-04 | 5.19E-03 | 9.46E-04 | |
| | Ovarian cancer (main signaling cascades) | 0.0208 | 0.0108 | 0.0326 | 0.019 | |
| | Schwann cells transformation in neurofibromatosis | 7.86E-04 | 5.79E-03 | 1.14E-03 | 4.50E-04 | |
| | Prostaglandin E2 in gastric cancer | 0.0219 | 9.89E-03 | 0.0234 | 0.0264 | 23200567 |
| | PGE2 pathways in cancer | 0.0183 | 7.40E-04 | 8.44E-04 | 3.00E-03 | 23200567 |
| | Cancer independent of coagulation protease | 0.0235 | 0.0466 | 0.0158 | 0.0182 | |
| **Cytoskeleton remodeling** | TGF, WNT and cytoskeletal remodeling | 0.0158 | 3.6E-05 | 8.1E-08 | 3.36E-04 | |
| | Cytoskeleton remodeling | 2.47E-03 | 2.74E-03 | 6.54E-04 | 6.64E-03 | |
| | FAK signaling | 4.99E-03 | 0.0466 | 6.59E-03 | 4.50E-04 | 27076682 |
| **Cell migration** | C5a-induced chemotaxis | 0.0111 | 0.0255 | 0.0499 | 0.012 | 8994115 |
| | Chemokines and adhesion | 0.0134 | 0.0105 | 0.0135 | 5.56E-03 | 25559893 |
| | Chemotaxis_Leukocyte chemotaxis | 0.0251 | 2.83E-03 | 6.99E-03 | 1.23E-03 | 25903727 |
| | CXCR4 signaling pathway | 8.45E-03 | 0.0415 | 2.67E-04 | 6.15E-03 | 17114479 |
| | VEGF signaling via VEGFR2-generic cascades | 4.17E-03 | 9.90E-03 | 5.77E-04 | 5.84E-04 | 12387457 |
| | Role of IL-8 in angiogenesis | 0.0223 | 0.012 | 8.21E-04 | 6.05E-05 | 15342208 |
| | SDF-1 signaling in hematopoietic stem cell homing | 6.13E-03 | 0.012 | 3.99E-03 | 2.14E-03 | 23192675 |
| **Apoptosis** | Role of PKR in stress-induced apoptosis | 9.7E-05 | 8.23E-03 | 8.1E-08 | 3.47E-05 | 15160394 |
| | BAD phosphorylation | 1.76E-03 | 5.45E-04 | 2.67E-04 | 1.34E-04 | 12507784 |

*Continued*

**Table 3** Continued

| | | FDR *p* value | | | | |
|---|---|---|---|---|---|---|
| | **Pathways** | **CIS** | **RR-MS** | **PP-MS** | **SP-MS** | **PubMed ID** |
| | Ceramides signaling pathway | 7.2E-06 | 0.0384 | 3.33E-04 | 7.40E-03 | 22544924 |
| | Anti-apoptotic action of gastrin | 0.0251 | 0.0255 | 0.0499 | 2.15E-03 | 12507784 |
| | HTR1A signaling | 4.99E-03 | 6.43E-03 | 2.39E-04 | 6.98E-04 | |
| | IL-4—antiapoptotic action | 1.70E-03 | 0.0257 | 0.0171 | 1.16E-03 | 23826999 |
| **Immunity** | Inhibitory PD-1 signaling in T cells | 4.8E-05 | 4.37E-03 | 3.52E-03 | 4.86E-03 | 17304234 |
| | HSP60 and HSP70/TLR signaling pathway | 9.8E-05 | 0.0372 | 4.05E-03 | 2.56E-03 | 25153885 |
| | Immunologic synapse formation | 5.37E-04 | 8.23E-03 | 3.33E-04 | 6.11E-04 | 22312112 |
| | ICOS pathway in T-helper cell | 1.39E-03 | 8.23E-03 | 4.44E-04 | 3.58E-03 | 19291374 |
| | CD28 signaling | 4.58E-03 | 0.0102 | 0.0127 | 0.0295 | 24412596 |
| | IL-15 signaling | 9.95E-03 | 5.31E-04 | 7.19E-04 | 0.037 | 21911607 |
| | NFAT in immune response | 0.0123 | 0.0281 | 4.45E-04 | 3.58E-03 | 25630465 |
| | TLR2 and TLR4 signaling pathways | 4.61E-04 | 0.0109 | 2.79E-03 | 8.36E-03 | 20434372 |
| | BCR pathway | 1.73E-03 | 8.37E-03 | 7.47E-04 | 2.94E-04 | 12244307 |
| | IL-2 activation and signaling pathway | 4.58E-03 | 0.045 | 1.81E-03 | 0.0131 | 24376757 |
| | MIF the neuroendocrine-macrophage connector | 6.82E-03 | 0.0345 | 6.65E-03 | 7.76E-04 | 23797673 |
| | EPO-induced MAPK pathway | 6.16E-03 | 0.0307 | 0.0304 | 1.33E-03 | 17728357 |
| | Thrombopoietin-regulated cell processes | 6.16E-03 | 0.0307 | 0.0135 | 1.33E-03 | |
| | Thromboxane A2 signaling pathway | 7.86E-04 | 9.89E-03 | 3.32E-04 | 1.21E-03 | |
| **Nervous system** | Dynein-dynactin motor complex in axonal transport | 0.0347 | 8.36E-03 | 0.0206 | 0.012 | |
| | Oligodendrocyte precursor cell proliferation | 0.0499 | 0.0307 | 0.0226 | 0.0143 | |
| | Ligand-independent activation of ESR1 and ESR2 | 0.0143 | 7.56E-03 | 8.45E-04 | 7.09E-03 | 15161628 |
| | CDK5 in neuronal death and survival | 0.0499 | 0.0415 | 0.0333 | 0.0156 | 20937706 |

Abbreviations: CIS = clinically isolated syndrome; MS = multiple sclerosis; PP-MS = primary progressive MS; RR-MS = relapsing-remitting MS; SP-MS = secondary progressive MS.
List of the 65 significantly enriched pathways shared across distinct MS stages and classified according to the main biological categories. Pathways with relevant scientific evidence of a pathogenic role in MS were provided with PubMed identifiers.

FDR-corrected *p* value cut-off of 0.05 and containing 3 global DEGs and at least 1 GWAS DEG for each disease form (figure 1). We found 165, 95, 184, and 223 significant pathways for CIS, RR-MS, PP-MS, and SP-MS respectively, with 65 pathways shared by all the disease courses and others common to at least 2 or 3 clinical forms (figure 2D, table e-1). Of interest, the 65 shared pathways belonged to major biological themes relative to transcription and translation, cell signaling and proliferation, cancer, cytoskeleton remodeling and cell migration, apoptosis, immunity, and nervous system (table 3). Functions related to immunity and cell signaling were predominant and included pathways regulating innate and adaptive immunity, and triggered by lipids, cytokines, and growth factors. An interactive link in the table e-1 opens the graphical maps representing the overall genes and the dysregulated MS transcripts involved in each pathway. It is important that when searching for scientific evidence about a role of the pathways in MS or its animal model by literature mining, we found it in most cases (table 3). Thus, bioinformatical annotation of gene expression data led to the identification of a core of pathogenetic functions common to all MS stages and characterized by dysregulation in genes including MS susceptibility genes.

**DISCUSSION** In this study, we provide evidence that genes located in the vicinity of MS risk variants display dysregulated expression in peripheral blood of subjects with distinct MS stages compared with healthy individuals, that the frequency of dysregulation is significantly higher than expected, and that those genes contribute to pathogenic pathways in MS.

The most recent Immunochip study has highlighted more than 100 novel non-MHC risk variants in MS and has underlined that several genes close to the identified polymorphisms play a key role in immunity, thus suggesting genetic susceptibility to a dysimmune state.[4] On the other hand, genetics contributes only to part of the risk of developing MS and has been unsuccessful in determining

predisposition to distinct clinical courses of MS, indicating an important role for environmental factors. Thus, the potential effect of the identified GWAS genes during distinct stages of MS remains elusive and needs clarification. As a first step into this direction, we have verified whether the GWAS genes postulated by the Immunochip analysis have altered transcript expression in peripheral mononuclear cells in MS subjects starting from the early phase of disease (CIS) to the distinct clinical courses most of the patients experience. Here, we show that although a core of GWAS genes is concordantly dysregulated in all MS stages when compared with the healthy population, each disease group is characterized by a distinct set of dysregulated GWAS genes, clearly indicating that alteration in the expression of MS risk genes is not a stable feature reproduced throughout the distinct phases and forms of disease, but is a selective event for each disease stage. Notably, 26 MS susceptibility genes were concordantly regulated in 2 independent cohorts measured with distinct array platforms which use different probes to detect gene expression, emphasizing the robustness of several dysregulations. A GWAS DEG in our study was CD86 (B7-2), a costimulatory molecule expressed by a wide range of leukocytes[14] and critically involved in T-lymphocyte activation and proliferation.[15,16] Monocytes of patients with RR-MS display higher expression of CD86 compared with healthy controls and interferon β treatment triggers reversal regulation of CD86.[17] Here, we demonstrate the overexpression of CD86 transcripts in PBMC at all the stages of MS compared with HCs in 2 independent microarray data sets and in a third independent cohort using q-PCR. Further, we report that the expression of JDP2, MAF, MAPK3, and RGS1 is upregulated in MS. Of interest, RGS1 is also upregulated by pathogenic Th17 T cells during experimental autoimmune encephalomyelitis (EAE).[18] Conversely, BACH2, a critical transcriptional factor regulating the balance between tolerance and immunity by repressing $CD4^+$ T-cell differentiation,[19] is downregulated in pathogenic T cells in EAE mice[18] and in the blood of RR-MS subjects.[20] Our study confirms this observation and extends it to the other MS stages. Very limited information is present in the literature about the expression of other GWAS genes in MS blood cells. A recent report has highlighted the downregulation of the transcription factor EOMES in whole blood of CIS and a pooled group of MS subjects with various disease courses compared with HCs by both RNA-seq and RT-PCR analysis.[21] We confirm the low transcript levels in CIS but not in the distinct clinical forms when analyzed separately. Furthermore, our study dissects the previously reported differential regulation of IKZF3, FOXP1, ZNF438 in CIS, and all MS clinical forms[21] and reproduces IL7R downregulation in PP-MS.[22] The regulation of the other GWAS genes during MS finds the first description in our study. Concerning the frequency of dysregulation of MS susceptibility genes, a recent work shows that CD4-positive pathogenic T cells during EAE are significantly enriched in dysregulated GWAS genes.[18] It is important that this observation is conserved in human PBMC, as we demonstrate a significantly high frequency of dysregulated GWAS genes in CIS and progressive forms of MS but not in RR-MS. As the potential to find transcriptional dysregulation in RR-MS data, however, was comparable with that in CIS (25.68% and 23.67% global dysregulation in CIS and RR-MS, respectively), the lack of enrichment in GWAS DEG in RR-MS has most probably a biological reason, whose explanation remains obscure.

Thus, to our knowledge, we report the first systematic expression analysis for MS GWAS genes using large cohorts of untreated subjects with distinct MS stages and HCs, and describe several novel GWAS dysregulations, whose single functional effects on MS pathogenesis deserve further investigations in appropriate animal models. In addition, genome-wide expression profiling has provided relevant information about the overall quality and frequency of dysregulation in PBMC transcriptomes at distinct MS stages. Previous studies have searched for pathways in MS using the full list of GWAS genes under the assumption that all GWAS genes are dysregulated in MS in all tissues and at all disease stages.[23,24] Our study demonstrates that this hypothesis may be wrong, as MS GWAS genes are not all dysregulated at the transcriptional level in PBMC and that there may be differences among disease courses. To investigate the combined effects of multiple dysregulations, we have searched for biological processes significantly enriched in DEG, including MS risk genes, as we have considered that MS is not a pure genetic disorder and pathogenic pathways may be contributed by transcriptional dysregulation determined by environmental factors as well. We report here the complete list of pathways obtained for each disease stage. Notably, several of them were detected in single forms of disease, whereas others were common to more than 1 MS stage with 65 pathways appearing in all the distinct MS subtypes. The shared pathways involve well-known functions in MS pathogenesis as cell signaling and migration, immunity, and apoptosis. Scientific evidence about a pathogenic role for most of these shared pathways is already available in the literature; however the contribution of the single pathways to the distinct MS stages still needs to be clarified. Furthermore, we

describe a series of novel candidate pathogenic pathways in MS, which deserve investigation and validation.

Overall, we offer the scientific community a large transcriptomic data set for expression studies in MS, the evidence about dysregulation of several known MS risk genes in blood at distinct MS stages, and the complete list of significant pathways contributed by GWAS genes for each disease form. This information may serve as the basis for novel translational medicine studies investigating the molecular framework underlying MS pathogenesis.

## AUTHOR CONTRIBUTIONS

S.S. performed transcriptome and pathway analysis, carried out statistical analyses, and wrote the paper. M.D.D. and A.R. prepared PBMC, extracted RNA, synthetized cDNA, and performed quantitative PCR experiments. R.M. performed quality control steps for transcriptomic data sets. M. Romeo, F.S., G.D.C., M. Rodegher, M. Radaelli., L.M., and V.M. enrolled subjects for the study and provided clinical information. D.C. and E.M. synthetized cRNA and hybridized microarray chips. E.B., G.M., and R.F. provided the Affymetrix data set. G.C. critically discussed the project and the results. C.F. conceived and designed the experiments, coordinated the study, analyzed the data, and wrote the paper.


## ACKNOWLEDGMENT

The authors thank Amici Centro Sclerosi Multipla for support and all the participants who donated blood for this study.

## STUDY FUNDING

This study was supported by the Italian Ministry of Health (RF-2011-02349698) and Merck-Serono.


## DISCLOSURE

S. Srinivasan, M. Di Dario, A. Russo, R. Menon, E. Brini, and M. Romeo report no disclosures. F. Sangalli received travel funding and/or speaker honoraria from Merck Serono, Biogen, Novartis, and Genzyme. G.D. Costa, D. Cantarella, and G. Martino report no disclosures. M. Rodegher served on the scientific advisory board for Genzyme, received speaker honoraria from Novartis, Merck Serono, and Biogen. M. Radaelli and L. Moiola received travel funding and speaker honoraria from Novartis, Teva Pharmaceutical Ltd., Sanofi, Genzyme, Merck Serono, and Biogen. E. Medico is on the editorial board for *Scientific Reports*. R. Furlan received speaker honoraria from Merck Serono, Novartis, and Roche, is an academic editor for *PLoS One*, holds a patent for CSF microvesicles, which are a marker of microglial activation, received research support from EMD Serono, Italian Ministry of Health, Italian Ministry of University and Research, and Italian Multiple Sclerosis Foundation. V. Martinelli participated in scientific advisory board for Merck Serono and Genzyme, received travel support and/or speaker honoraria from Biogen Idec, Merck Serono, Bayer, Teva Pharmaceutical, Novartis, and Genzyme. G. Comi received personal fees from Novartis, Teva, Sanofi, Genzyme, Merck Serono, Biogen, Bayer, Serono Symposia Int. Foundation, EXCEMED, Almirall, and Receptos. C. Farina is on the editoral board for *Journal of Advances in Medicine*, received research support from Merck Serono, Teva, Novartis, Italian Ministry of Health, and Fondazione Italiana Sclerosi Multipla. Go to Neurology.org/nn for full disclosure forms.


*Received October 17, 2016. Accepted in final form January 23, 2017.*


## REFERENCES


1. Achiron A, Barak Y. Multiple sclerosis-from probable to definite diagnosis: a 7-year prospective study. Arch Neurol 2000;57:974–979.
2. Miller DH, Chard DT, Ciccarelli O. Clinically isolated syndromes. Lancet Neurol 2012;11:157–169.
3. Confavreux C, Vukusic S. Natural history of multiple sclerosis: a unifying concept. Brain 2006;129:606–616.
4. International Multiple Sclerosis Genetics Consortium (IMSGC), Beecham AH, Patsopoulos NA, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat Genet 2013;45:1353–1360.
5. Menon R, Di Dario M, Cordiglieri C, et al. Gender-based blood transcriptomes and interactomes in multiple sclerosis: involvement of SP1 dependent gene transcription. J Autoimmun 2012;38:J144–J155.
6. McDonald WI, Compston A, Edan G, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the international panel on the diagnosis of multiple sclerosis. Ann Neurol 2001;50:121–127.
7. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical bayes methods. Biostatistics 2007;8:118–127.
8. International Multiple Sclerosis Genetics Consortium, Wellcome Trust Case Control Consortium 2, Sawcer S, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 2011;476:214–219.
9. Patsopoulos NA, Bayer Pharma MS Genetics Working Group, Steering Committees of Studies Evaluating IFNbeta-1b and a CCR1-Antagonist, et al. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. Ann Neurol 2011;70:897–912.
10. Lill CM, Schjeide BM, Graetz C, et al. Genome-wide significant association of ANKRD55 rs6859219 and multiple sclerosis risk. J Med Genet 2013;50:140–143.
11. Dennis G Jr, Sherman BT, Hosack DA, et al. DAVID: database for annotation, visualization, and integrated discovery. Genome Biol 2003;4:P3.
12. Kasprzyk A. BioMart: driving a paradigm change in biological data management. Database (Oxford) 2011;2011:bar049.
13. Reimers M, Carey VJ. Bioconductor: an open source framework for bioinformatics and computational biology. Methods Enzymol 2006;411:119–134.
14. O'Neill SK, Cao Y, Hamel KM, Doodes PD, Hutas G, Finnegan A. Expression of CD80/86 on B cells is essential for autoreactive T cell activation and the development of arthritis. J Immunol 2007;179:5109–5116.
15. Sharpe AH, Freeman GJ. The B7-CD28 superfamily. Nat Rev Immunol 2002;2:116–126.
16. Windhagen A, Newcombe J, Dangond F, et al. Expression of costimulatory molecules B7-1 (CD80), B7-2 (CD86), and interleukin 12 cytokine in multiple sclerosis lesions. J Exp Med 1995;182:1985–1996.
17. Filion LG, Matusevicius D, Graziani-Bowering GM, Kumar A, Freedman MS. Monocyte-derived IL12, CD86 (B7-2) and CD40L expression in relapsing and progressive multiple sclerosis. Clin Immunol 2003;106:127–138.
18. Hoppmann N, Graetz C, Paterka M, et al. New candidates for CD4 T cell pathogenicity in experimental neuroinflammation and multiple sclerosis. Brain 2015;138:902–917.
19. Roychoudhuri R, Restifo NP. BACH2 represses effector programs to stabilize T(reg)-mediated immune homeostasis. Nature 2013;498:506–510.
20. Perga S, Montarolo F, Martire S, Berchialla P, Malucchi S, Bertolotto A. Anti-inflammatory genes associated with

multiple sclerosis: a gene expression study. J Neuroimmunol 2015;279:75–78.
21. Parnell GP, Gatt PN, Krupa M, et al. The autoimmune disease-associated transcription factors EOMES and TBX21 are dysregulated in multiple sclerosis and define a molecular subtype of disease. Clin Immunol 2014;151: 16–24.
22. Booth DR, Arthur AT, Teutsch SM, et al. Gene expression and genotyping studies implicate the interleukin 7 receptor in the pathogenesis of primary progressive multiple sclerosis. J Mol Med 2005;83:822–830.
23. Baranzini SE, Galwey NW, Wang J, et al. Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. Hum Mol Genet 2009;18:2078–2090.
24. International Multiple Sclerosis Genetics Consortium. Network-based multiple sclerosis pathway analysis with GWAS data from 15,000 cases and 30,000 controls. Am J Hum Genet 2013;92:854–865.