OXFORD

Human Molecular Genetics, 2018, Vol. 27, No. 1 66–79

doi: 10.1093/hmg/ddx385
Advance Access Publication Date: 26 October 2017
Original Article

ORIGINAL ARTICLE

Combined microRNA and mRNA expression analysis in pediatric multiple sclerosis: an integrated approach to uncover novel pathogenic mechanisms of the disease

Maria Liguori[1,*], Nicoletta Nuzziello[1,2], Flavio Licciulli[1], Arianna Consiglio[1], Marta Simone[2], Rosa Gemma Viterbo[2], Teresa Maria Creanza[3], Nicola Ancona[3], Carla Tortorella[2,4], Lucia Margari[2], Giorgio Grillo[1], Paola Giordano[5], Sabino Liuni[1] and Maria Trojano[2]

[1]National Research Council of Italy, Department of Biomedicine, Institute of Biomedical Technologies, Bari Section, 70125 Bari, Italy, [2]Department of Basic Sciences, Neurosciences and Sense Organs, University of Bari, 70125 Bari, Italy, [3]National Research Council of Italy, Department of Engineering, ICT and Technology for Energy and Transportation, Institute of Intelligent Systems for Automation, 70125 Bari, Italy, [4]Department of Neurosciences, San Camillo Forlanini Hospital, 00185 Rome, Italy and [5]General Paediatric Unit "B. Trambusti", Azienda Policlinico-Giovanni XXIII and University of Bari, 70125 Bari, Italy

*To whom correspondence should be addressed at: National Research Council, Department of Biomedicine, Institute of Biomedical Technologies, Bari Section, via Amendola no. 122/D, 70126 Bari, Italy. Tel: +39 080 5929663; Fax: +39 080 5929690; Email: maria.liguori@ba.itb.cnr.it or maria.liguori@cnr.it

Abstract

Multiple sclerosis (MS) is a complex disease of the CNS that usually affects young adults, although 3–5% of cases are diagnosed in childhood and adolescence (hence called pediatric MS, PedMS). Genetic predisposition, among other factors, seems to contribute to the risk of the onset, in pediatric as in adult ages, but few studies have investigated the genetic 'environmentally naïve' load of PedMS. The main goal of this study was to identify circulating markers (miRNAs), target genes (mRNAs) and functional pathways associated with PedMS; we also verified the impact of miRNAs on clinical features, i.e. disability and cognitive performances. The investigation was performed in 19 PedMS and 20 pediatric controls (PCs) using a High-Throughput Next-generation Sequencing (HT-NGS) approach followed by an integrated bioinformatics/biostatistics analysis. Twelve miRNAs were significantly upregulated (let-7a-5p, let-7b-5p, miR-25–3p, miR-125a-5p, miR-942–5p, miR-221–3p, miR-652–3p, miR-182–5p, miR-185–5p, miR-181a-5p, miR-320a, miR-99b-5p) and 1 miRNA was downregulated (miR-148b-3p) in PedMS compared with PCs. The interactions between the significant miRNAs and their targets uncovered predicted genes (i.e. *TNFSF13B, TLR2, BACH2, KLF4*) related to immunological functions, as well as genes involved in autophagy-related processes (i.e. *ATG16L1, SORT1, LAMP2*) and ATPase activity (i.e. *ABCA1, GPX3*). No significant molecular profiles were associated with any PedMS demographic/clinical features. Both miRNAs and mRNA expressions predicted the phenotypes (PedMS-PC) with an accuracy of 92% and 91%, respectively. In our view, this original strategy of contemporary miRNA/mRNA analysis may help to shed light in the genetic background of the disease, suggesting further molecular investigations in novel pathogenic mechanisms.

Received: September 4, 2017. **Revised:** October 23, 2017. **Accepted:** October 23, 2017

Introduction

Multiple sclerosis (MS) is a complex multifactorial disease of the CNS characterized by inflammatory autoimmune demyelination and neurodegeneration (1). Usually identified as a disease of young adults, improvements in diagnostic tools and/or an increased sensitivity to its first signs and symptoms may have also contributed to a better recognition of MS during the very early ages of life, whereas in the past it had been mostly retrospectively tracked. At present, pediatric MS (PedMS) represents 3–5% of all MS cases; Relapsing Remitting MS (RRMS) is considered the most frequent form of the disease, with a relapse rate higher than the adult RRMS, in line with the evidence of more pronounced inflammatory MRI features (2–5).

Genetic predisposition, lifestyle, and environmental factors seem to contribute significantly to the overall risk of the onset of MS (6). Since the rarity of PedMS occurrence, and considering the general assumption that pediatric and adult MS patients should share the same risk factors (7), very few studies have investigated the genetic involvement of PedMS. However, children with MS can be considered 'environmentally naïve' so their genetic load may provide more significant information than in adult patients, especially about those genetic factors that trigger the onset of the disease at younger ages (8) or modulate its phenotypes. Identifying these peculiarities would be of valuable help, for example, in the search of more targeted therapeutic approaches.

Indeed, recent studies demonstrated that a genetic architecture of the normal human developing brain does exist, mostly driven by a given set of microRNAs (miRNAs) acting as age-related transcription factors on genes enclosed in specialized functional networks (9). miRNAs are a class of non-coding RNA (ncRNA) genes whose products are small single-strand RNAs (with lengths of about 22 nucleotides) involved in the regulation of translation of over 60% mammalian genes (10,11). Deregulation of miRNAs expression may be involved in a broad spectrum of cellular and biological processes including apoptosis and cell proliferation, and it has been also observed in molecular dysfunctions underlying several neurodegenerative diseases like Amyotrophic Lateral Sclerosis (ALS) (12), Parkinson's Disease (PD) (13), and Dementias (14).

In MS, it has been reported that miRNAs are not only involved in the pathogenesis of the disease (15) but also associated with different MS subtypes (16,17), as well as with the response to specific treatments like Interferon Beta-1b (18), Natalizumab (19), or Glutiramer Acetate (20), thus confirming their potential as biomarkers of disease monitoring. So far, no miRNAs investigations have been performed in PedMS.

The main goal of this study was to identify patterns of circulating markers (miRNAs) associated with a selected population of PedMS submitted to periodical clinical examinations and compared with a pediatric control population of healthy subjects (PCs). By using a High-Throughput Next-generation Sequencing (HT-NGS) approach, we were able to characterize different miRNAs and their target genes, and analysed the output data *in house* by an integrated team of bioinformatics and biostatistics personnel with long-standing expertise in NGS data analysis (Fig. 1), in order to uncover molecular pathway/s associated with PedMS. Additional aim was to search for transcriptomic profiles (miRNAs/mRNAs) distinctive of PedMS features, such as clinical disabilities or cognitive dysfunctions.

Results

The study population consisted of 19 patients with PedMS and 20 PCs (Table 1); they differed significantly for age at the study entry ($P = 0.001$), whereas no significant gender difference was found between them ($P > 0.05$) (Please also visit: http://www.ba.itb.cnr.it/pedms/ for graphical and data support).

PedMS patients were analysed during the inactive phase of a RR course of the disease (no clinical relapses or steroid assumption within 30 days of the study entry, no Gd-enhancing lesions at the concomitant MRI) and were steroid-free (at least one month prior to the blood sample). Ten out of the 19 patients had never assumed disease-modifying treatment (DMT) at the time

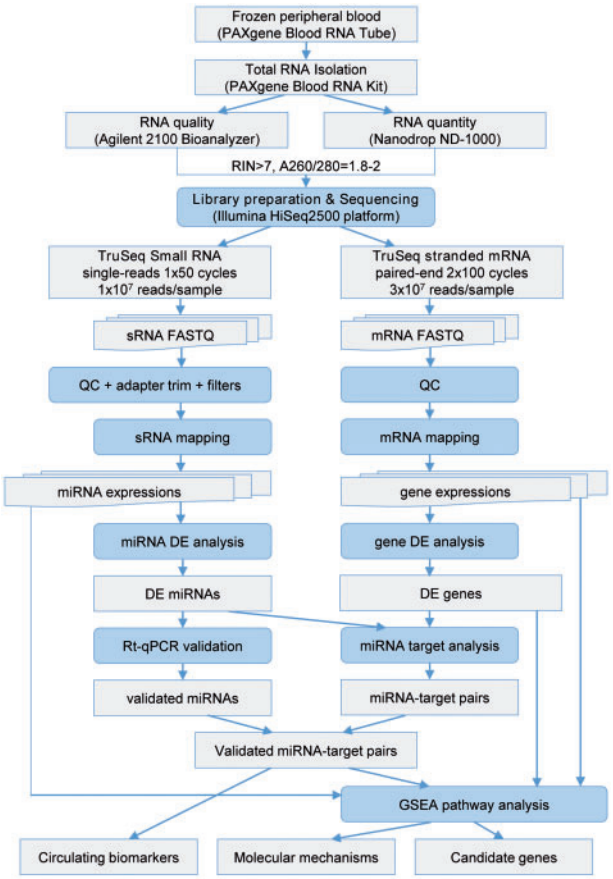


Figure 1. Workflow of the study.

Table 1. Characteristics of the study sample

	PedMS ($n = 19$)	PC ($n = 20$)
Female/Male	10/9	6/14
Mean age (SD), years	15.48 (2.74)[b]	8.83 (3.26)
Mean age at MS onset (SD), y	12.64 (3.20)	–
MS course	RR	–
Mean EDSS score (SD)	2.69 (1.07)	–
Mean disease duration (SD), years	2.84 (3.31)	–
Mean education (SD), years	9.56 (2.53)	3.39 (2.64)
DMT treatment° (yes/no)	9/10	–
Cognitive Abilities (CI/CP/NA)[a]	7/10/2	NA

EDSS= expanded Disability Status Scale.

At the time of the blood sample (all treated patients were under Interferon beta-1a therapy).

[a]CI= cognitively impaired (failure ≥ 3 tests); CP= cognitively preserved; NA= not available

[b]Mann–Withney U-test; $P < 0.001$.

of the blood test, whereas the remaining 9 had taken Interferon Beta-1a for a range of 21–673 days from the blood samples.

Seventeen PedMS patients completed the neuropsychiatric evaluation; 7 of them (41%) tested cognitively impaired (CI). No significant differences were found between CI_PedMS and the cognitively preserved ones (CP) for gender, age at MS onset, diseases duration, EDSS and years of education ($P > 0.05$).

Analysis of miRNAs and mRNAs differentially expressed (DE)

The bioinformatics analysis of HT-NGS data produced the following results. The comparisons of miRNAs expression levels within the study groups revealed 49 mature miRNAs significantly DE between PedMS and PCs (40 up-regulated and 9 downregulated) (Supplementary Material). No miRNAs discriminated CI-PedMS from CP-PedMS subjects.

After further selection (miRNAs with mean number of reads > 25, Fold Change > 1.5 and Dispersion Index range 0–1.6), we restricted the total list to 27 miRNAs that were subjected to qRT-PCR technical validation (Table 2).

The contemporary comparative analysis of mRNA-seq reads identified 4,306 genes significantly DE between PedMS and PCs (2215 upregulated and 2091 downregulated). No significantly DE miRNAs/genes resulted associated with any clinical features or DMT treatment in PedMS (*data available not showed*).

Prediction accuracy of DE analysis

The differences at molecular levels, highlighted by the DE analyses in which the transcripts were analysed singularly, were confirmed by a global analysis. In particular, the differences between two phenotypic conditions were analysed by comparing the whole profile of the subjects using the prediction accuracy (ACC) method. The higher the ACC value, the stronger the difference at the molecular level between the two phenotypic conditions analysed. Patients with PedMS showed deeply altered miRNA and mRNA expressions when compared with PCs subjects, since we were able to predict the phenotypic status of a subject with ACC values of 92% and 91%, respectively, using only their expression profiles.

qRT-PCR technical validation (miRNAs)

Five candidate endogenous reference genes (miR-93–5p, miR-103a-3p, miR-106b-5p, miR-191–5p, and miR-423–5p) were selected from qRT-PCR blood studies (21,22) and from suggested miRNA endogenous controls in the TaqMan Advanced miRNA Assays white paper (Applied Biosystem, Thermo Fisher Scientific). The candidate reference genes were tested for stable expression across sRNA-Seq data results with the following criteria: (a) high read count in all samples; (b) no intra- and intergroup DE (P-value < 0.05). miR-103a-3p and miR-191–5p were found to fit these criteria, so they were selected as internal references.

In the comparison between subjects with PedMS and PC, 13 miRNAs (let-7a-5p, let-7b-5p, miR-25–3p, miR-125a-5p, miR-942–5p, miR-221–3p, miR-652–3p, miR-182–5p, miR-185–5p, miR-181a-5p, miR-320a, miR-99b-5p, and miR-148b-3p) were confirmed significantly DE ($P < 0.05$) (with only the last miRNA downregulated and the others upregulated, as in the sRNA-seq analysis) (Fig. 2 and Supplementary Material).

Gene target prediction

Among the DE genes, we searched for targets of the overall 27 mature miRNAs resulted from the PedMS *vs*. PC comparison (the total numbers of genes are indicated in Table 2). Following both the approaches detailed in the methodological section, for the 13 validated miRNAs that resulted from the comparison PedMS *versus* PC, we found that 68 significantly down-regulated target genes resulted experimentally validated or computationally predicted (Table 2). Interestingly, *E2F2, DIP2A, GID4, GOLGA8A, KLHL14, CDC34, RPL35A, MICAL3*, and *DNAJA4* were shared by at least two of the following miRNAs: let-7a-5p, let-7b-5p, miR-181a-5p, miR-320a, miR-25–3p, miR-185–5p, and miR-125a-5p (Fig. 3A).

Table 2 also shows the total numbers of target genes that were significantly DE in our integrated analysis, including those upregulated; as already reported, no complete overlap between the overall target genes and their validated miRNA expressions was found in our combined data (23) (Supplementary Material for details).

Pathway analysis

Cluster analysis of the 13 validated miRNAs identified several common KEGG pathways, the most significant being those implicated in Prion disease (hsa05020), fatty acid biosynthesis (hsa00061), cell cycles (hsa04110), Hippo signaling pathway (hsa04390) and adherence junction processes (hsa04520) ($P < 0.00001$; Fig. 4A). Together with their target mRNAs that were significantly DE in our analysis (926, up-/downregulated), GSEA showed that significant immunological signatures were implicated (Fig. 4B). The most representative (10/13 miRNAs) was the network of genes downregulated in peripheral blood monocytes (PBMC), which helps to distinguish from infectious from non-infectious causes of inflammatory response (GSE9960), followed by the one identified by genes upregulated in hematopoietic stem cells (HSCs) *versus* granulo-monocyte progenitors (GSE15330) (further details in Supplementary Material).

In the above-mentioned sub-network composed of 7-miRNAs (let-7a-5p, let-7b-5p, miR-181a-5p, miR-320a, miR-25–3p, miR-185–5p, and miR-125a-5p), molecular mechanisms related to lysosomial activities, such as endocytosis, phagocytosis, and autophagy, were the main features of the most representative canonical pathway (KEGG_lysosome), whereas among the functional categories of biological processes (GO-terms), the most representative network enclosed any response to oxygen molecular compounds. Finally, GSEA significant pathways evoked the involvement of genes implicated in innate immune responses of PBMCs, dendritic cells (DCs) HSCs (Fig. 3B).

Discussion

As far as we know, this is the first investigation combining original individual miRNAs/mRNAs data of patients with PedMS. The analysis showed that 12 mature miRNAs were significantly upregulated (let-7a-5p, let-7b-5p, miR-25–3p, miR-125a-5p, miR-942–5p, miR-221–3p, miR-652–3p, miR-182–5p, miR-185–5p, miR-181a-5p, miR-320a, and miR-99b-5p) and 1 miRNA was downregulated (miR-148b-3p) in PedMS patients compared with PC subjects. Furthermore, the analysis of the interactions between the DE miRNAs and their DE targets (mRNAs) uncovered several predicted genes, which a subsequent enrichment analysis categorized in functional pathways. Taken together, in

Table 2. List of the significant miRNAs (discovery/validation steps), and the most significant overlapping target genes. The *first section (from the left)* listed the 27 miRNAs that resulted significantly different (up/down-regulated: *second column*) in the comparison between PedMS and PCs performed in the discovery phase (NGS); for each miRNA, the main attributes from the sRNA-seq analysis (mean reads, log-Fold change, dispersion index and adjusted *P*-value) have been detailed. The *central section* shows the results of the technical validation (log-Fold change, *P*-value), confirming 13 miRNAs (top 13 in the list). In the *last section* (*on the right*), for each miRNA we summarized the total numbers of target genes that resulted significantly DE in our data, and among them only the overlapping genes reported computationally predicted or experimentally validated as target by published studies

	Discovery phase					Technical validation			
Transcript_id	Regulation	Mean reads	logFC	Dispersion index	Adjusted P-val.	logFC	P-value	Targets	validated/predicted overlapping genes
let-7a-5p[b]	up	3827.47	0.900599	0.422591	0.022873	1.489242	**0.005564**	**128**	**E2F2**[a], *RRM2*[a], **CDC34, RPL35A, DIP2A, MICAL3**
let-7b-5p[b]	up	2820.64	1.245435	0.575656	0.006478	1.745702	**0.000252**	**304**	*IGF2BP2*[a], **CDC34**[a], **MICAL3, E2F2, RPL35A,** *ANKRD9, CHPT1*
miR-25-3p	up	13167.76	0.754643	0.218633	0.006764	0.983619	**0.005749**	**157**	**KLHL14,** *BCAM*, **GID4**, *FHL2*, **GOLGA8A**
miR-99b-5p	up	66.89	1.035963	0.342344	0.004245	0.874308	**0.008515**	**51**	
miR-125a-5p[b]	up	273.04	1.384516	0.245282	0.000004	1.674539	**0.000225**	**193**	**E2F2, DIP2A,** *HIP1R, ADD2*, **DNAJA4**
miR-148b-3p	down	130.31	−0.894901	0.468691	0.020605	−0.749805	**0.043730**	**107**	*RTN3, CATSPER1, TMED7*, **KLF4,** *B4GALT5, MAFB*
miR-181a-5p[b]	up	20258.98	0.732759	0.314549	0.034282	0.949650	**0.004887**	**148**	*CARM1, DIP2C*, **GID4,** *POMC*, **DNAJA4, GOLGA8A**, *PAX5, ZNF594*
miR-182-5p	up	4104.76	0.916624	0.361750	0.011115	1.148772	**0.001568**	**183**	*MZB1*[a], *CEP128, RAB6B, FAM210B*
miR-185-5p	up	203.32	1.326069	0.544402	0.002817	0.799854	**0.002469**	**235**	*FCRL3, BACH2, PLXDC1, TUBB6, CPNE5*, **MICAL3,** *TFDP1, IGLL5, TCL1A, OSBP2*
miR-221-3p[b]	up	79.08	1.975037	0.538890	0.000011	1.056404	**0.001073**	**196**	*EML6, GNAS, SLC25A37, RGS6, TFR2*
miR-320a[b]	up	2412.66	0.618377	0.238738	0.040466	1.327173	**0.007294**	**358**	*AQP1*[a], *QSOX2, CELSR3, ANKH*, **KLHL14, DIP2A,** *FKBP8*
miR-652-3p	up	1522.32	0.892013	0.222386	0.001715	1.228058	**0.001181**	**184**	*PLEK2*
miR-942-5p	up	104.67	1.412644	0.352011	0.000158	1.565965	**0.000974**	**189**	*RANBP10, RPL36A, SMIM5, FAM117A*
let-7i-3p	up	110.87	0.841634	0.253236	0.006669	0.334213	0.2881	157	
let-7i-5p	up	2308.89	1.172369	0.558958	0.009023	−1.476012	0.3661	183	*ANKDRD9, DUSP2, E2F2, DIP2A, MICAL3, CDC34*
miR-10a-5p	up	38.36	1.161355	0.722607	0.027450	Undetermined		77	*LIMS2, ZNF154*
miR-21-5p[b]	up	41.50	1.385189	0.656703	0.006478	0.347856	0.2002	78	*TPM1*[a]
miR-29a-3p	down	30.43	−0.982972	0.050754	0.021160	Undetermined		117	*RAB31, GNB4, PTP4A1, KLF4*[a], *EPSTI1*
miR-30e-5p	down	119.45	−0.839242	0.410514	0.039331	Undetermined		125	*TMEM170B, ZCCHC24, FGD6, PTP4A1*
miR-130b-3p	up	81.33	1.154808	0.332325	0.001154	0.390139	0.6006	156	*XK, TCL1A, ZNF594*
miR-140-3p[b]	down	2701.32	−0.669809	0.298620	0.001350	−0.275507	0.37790	125	*CTSD, CITED4, DIAPH2, MARCKS, ARSD, XAF1, GNB4, PDLIM7, KLF4*
miR-144-5p[b]	up	26.91	3.061378	1.581342	0.000187	0.947559	0.7879	45	*CPNE5, BBOF1*
miR-151b	up	33.09	1.403847	0.772518	0.012100	0.723572	0.09075	64	
miR-451b	down	2658.08	−1.950685	1.035467	0.002311	Undetermined		78	*DIAPH2, EPB41L3, STS, FAM46A, SHTN1, PTP4A1, PARP9, USP32*
miR-484	up	14846.77	0.587641	0.218315	0.041494	0.925373	0.1059	346	*DENND58, DTX1, PASK, ANK1, CCDC191, SMIM5, MRC2, CDC34*
miR-501-3p	up	72.07	0.672652	0.295001	0.048552	0.673967	0.2373	136	*MYBL2, RGS6*
miR-3605-3p	up	143.22	0.928963	0.319553	0.007265	0.905319	0.2328	173	*ZWINT, E2F2, MCF2L, FIS1, ELOB, CDC34*

In the last column, those genes that targeted more than one miRNA are identified in **bold** (results outlined only within the 13 confirmed miRNAs, although some of these common genes were also targeted by the non-validated miRNAs; details in Fig. 2 and Supplementary Material).

[a]Experimentally validated target genes (Luciferase assay).

[b]miRNAs that have been already reported associated with MS (*see* Discussion for comments and references).

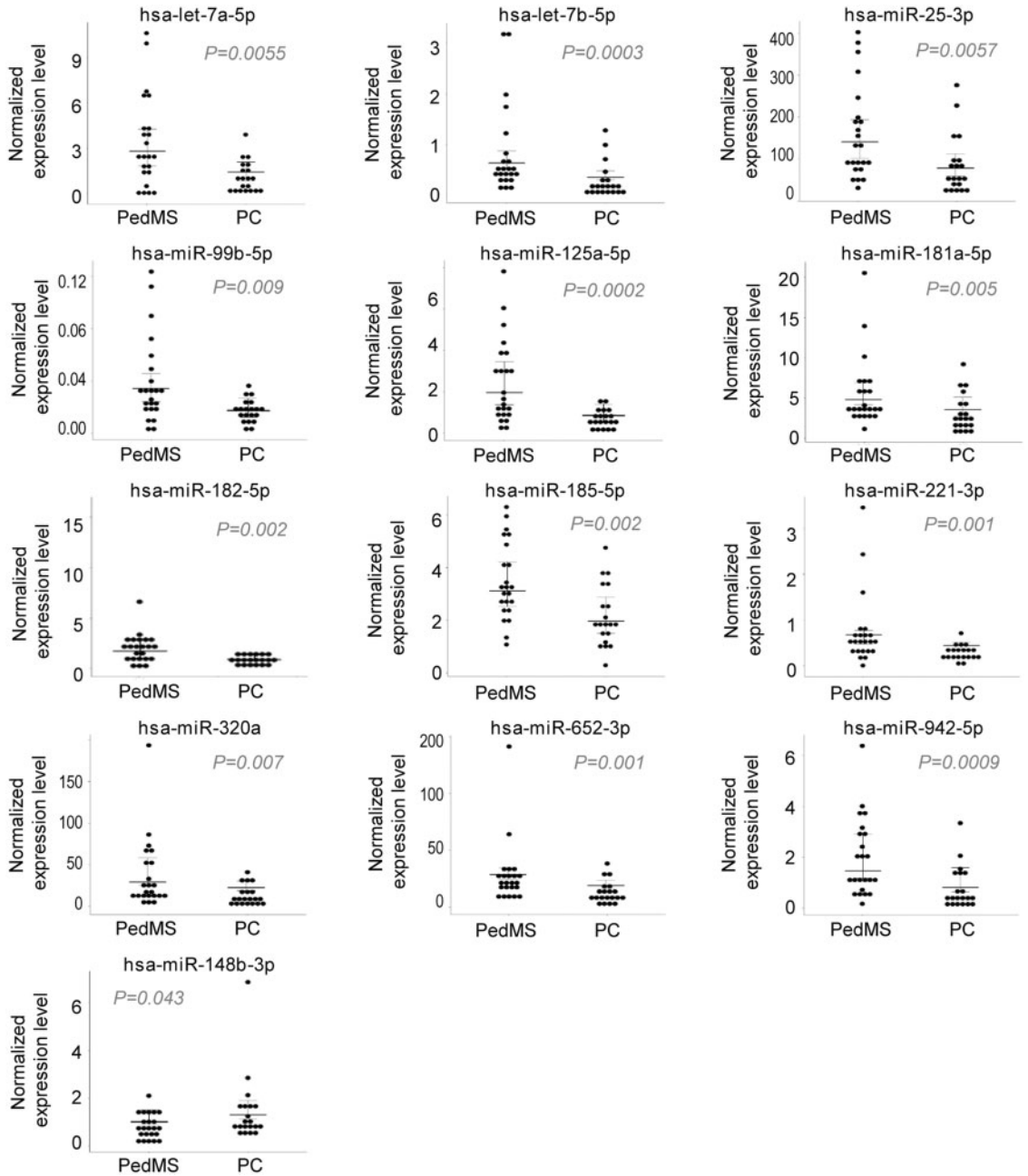
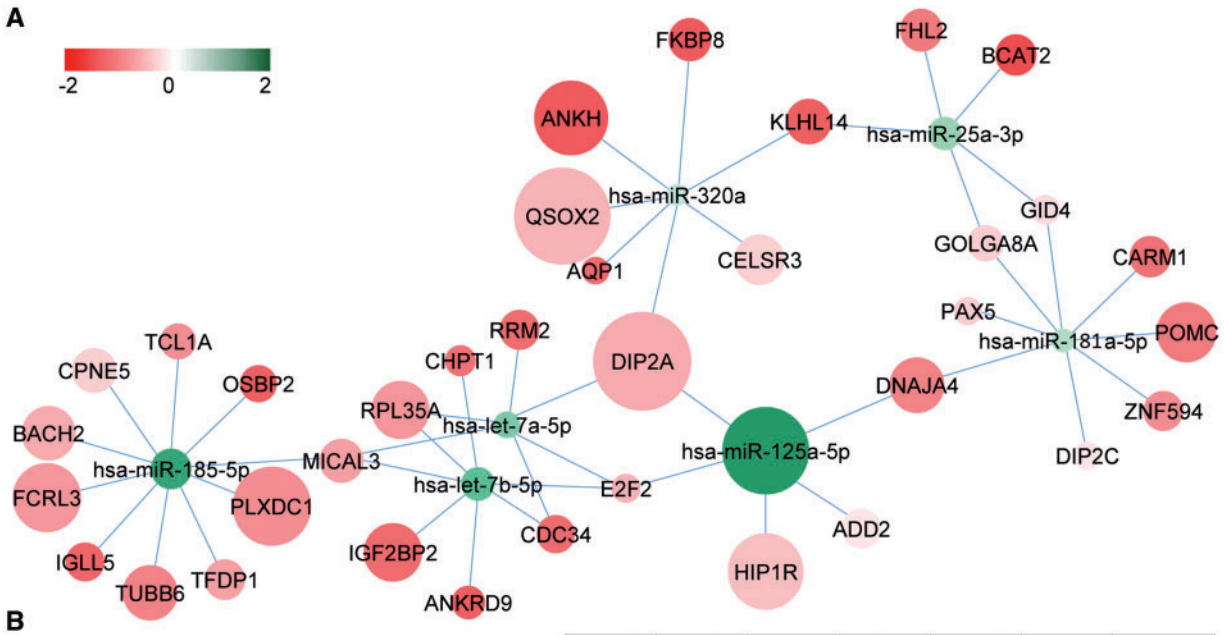


Figure 2. Differentially expressed miRNAs validated by qRT-PCR. Scatter dot plots of the significant miRNAs resulted from the analysis of PedMS patients compared with PC subjects ($P < 0.05$). The comparisons of normalized values between the subgroups was obtained according to the $2-\Delta\Delta Ct$ method. Y-axis represents the expression levels normalized to the geometric mean of endogenous references (miR-103a-3p and miR-191-5p). The horizontal line indicates the median. Two replicates were performed for each sample. Statistical analysis was carried out using Student's unpaired two-tailed t-test or two-tailed Mann–Whitney U-test, when appropriate (Supplementary Material, Table S1).

our view, these results draw a plausible picture of the PedMS molecular signature and suggest in-depth molecular investigations on novel circulating biomarkers and candidate genes.

In the last few years, many published reports addressed the attention to the dysregulation of miRNAs as deeply implicated in the pathogenesis of MS (15,24) or correlated their circulating expressions, for example, with MRI measures of MS severity (25), or with responsiveness to specific MS therapies (26,27).

In order to improve the accuracy of identifying miRNAs, genes and functional pathways possibly implicated in MS pathogenesis, and to overcome the limitation of small cohort studies performed, for example, in different tissues, recent consensus-based studies analysed and integrated publically available microarray data (GEO dataset) (23,28). In particular, this effort produced two regulatory networks: a miRNA network and a gene-based network *independent* from each other; as an overall take-home-message, they suggested that DE miRNAs may be more informative than DE genes in uncovering molecular pathways of interest in complex diseases like MS (23). However, as discussed in the same report, the microarray



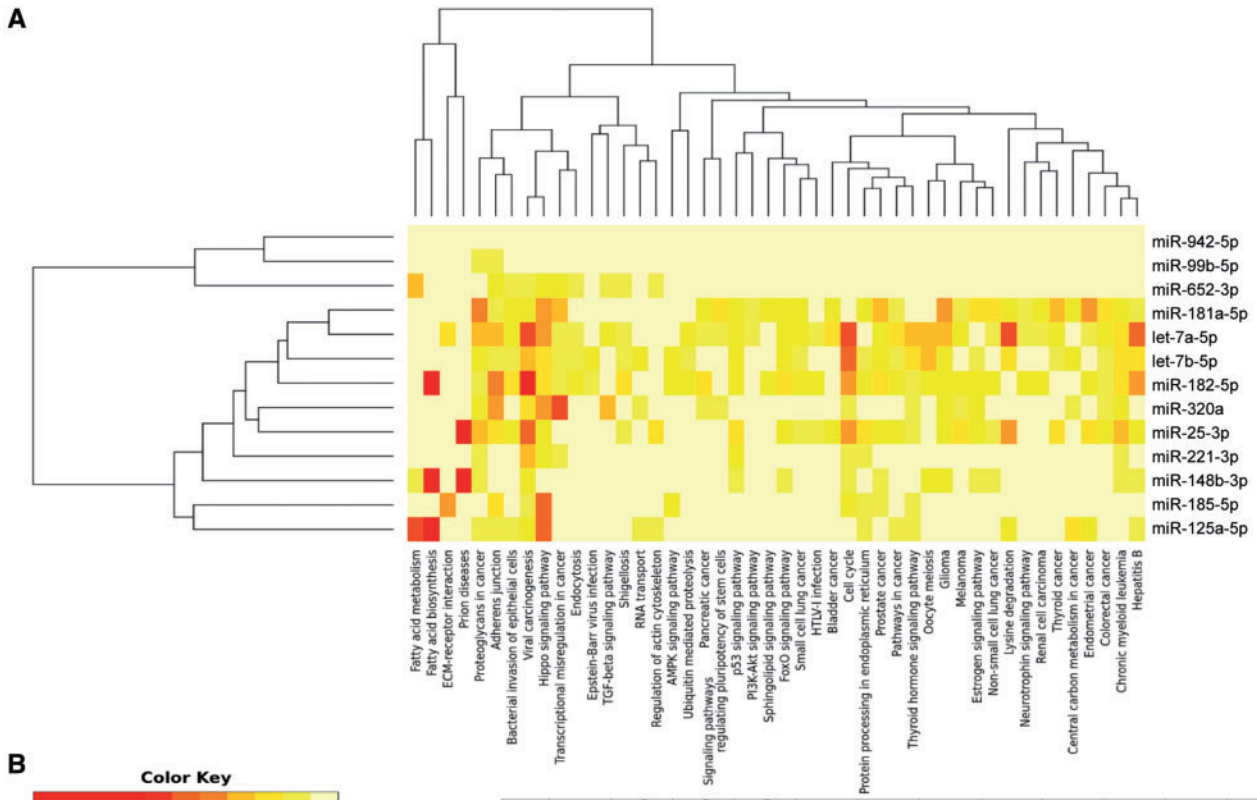
B

DC = dendritic cells; PBMC = peripheral blood mononuclear cells; HSCs = hematopoietic stem cells; ITGAM = integrin subunit alpha-M; IFNAR1 = interferon alpha and beta-receptor subunit-100	let-7a-5p	let-7b-5p	miR-125a-5p	miR-181a-5p	miR-185-5p	miR-25-3p	miR-320a
Canonical pathways							
Lysosome cycle (KEGG: hsa04142)			P=0.008	P=0.005	P=0.004	P<0.001	P=0.002
Cellular components							
Vacuole (GO:005773)	P=0.006	P<0.001	P<0.001	P=0.006			
Molecular functions							
ATPase Activator Activity (GO:8081671)	P=0.008	P=0.009	P=0.008		P=0.007	P=0.008	P=0.006
ATPase Binding (GO:8051117)	P=0.008	P=0.009			P=0.007		P=0.006
Immunological signatures							
Genes up-regulated in HSC vs. granulo-monocyte progenitors (GSE1533)	P=0.003	P=0.002	P=0.006		P<0.001	P<0.001	
Genes down-regulated in comparison of naïve-CD8 T-cells vs. unstimulated DCs (GSE22886)	P=0.007	P=0.002	P=0.007			P=0.01	P=0.01
Genes involved in molecular mechanisms underlying the innate response to influenza vaccination (GSE29614)	P=0.008	P=0.009	P=0.008		P=0.007	P=0.008	P=0.006
Genes down-regulated in comparison of B cells vs. myeloid DCs (GSE29618)		P=0.001			P=0.006	P=0.004	P=0.005
Genes down-regulated in comparison of plasmacytoid DCs vs. myeloid DCs (GSE29618)	P=0.003	P=0.003				P=0.01	P=0.001
Genes down-regulated in ITGAM wildtype vs. IFNAR1 (GSE45365)	P<0.001	P=0.003	P=0.008		P=0.006		P=0.01
Genes down-regulated in PBMC during Gram(-) vs. Gram (+) sepsis (GSE9960)		P<0.001	P<0.001	P=0.002		P=0.01	P<0.001

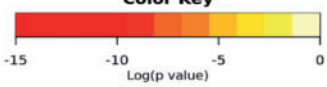
Figure 3. (**A**) Target genes that resulted significantly downregulated in PedMS shared by seven validated miRNAs ('7-miRNAs-loop' in the manuscript). miRNA-mRNA interactions were visualized as a network using Cytoscape 3.5.1. The node intensity color is proportional to fold change between PedMS and PC (red: down-regulated, green: up-regulated). The node size is proportional to the *P*-value in the differential expression analysis. (**B**) The table indicates the significant results ($P < 0.01$) of the pathways analysis (KEGG, GEO, GSEA datasets) within the 7 miRNAs-loop; all pathways resulted up-regulated.

technology allows one to investigate a limited number of miRNAs (900 out of more than 2000), thus limiting the full coverage of these circulating molecular biomarkers.

Microarrays technologies have been gradually replaced by the newly developed high-throughput technologies such as NGS, an innovative approach that provides tremendous potential for profiling both mRNAs and miRNAs. Analysis of a larger set of miRNAs of interest with this technology confirmed the method's feasibility and its potential use in the evaluating miRNAs dysfunctions. However, most of these studies did not investigate other transcription products, such as the correspondent target genes: in fact, each miRNA can regulate hundreds of target genes, so the possibility to neglect the overall impact on the genome functional activities is consistent. Furthermore, post-processing analysis of such large amount of data (several million reads per sample) requires specialized bioinformatics abilities thorough biostatistics knowledge in order to obtain accurate, congruent, and reproducible results.



B



	let-7a-5p	let-7b-5p	miR-125a-5p	miR-148b-3p	miR-181a-5p	miR-182-5p	miR-185-5p	miR-221-3p	miR-25-3p	miR-320a	miR-652-3p	miR-942-5p	miR-99b-5p
Frequencies of canonical pathways (KEGG dataset)													
KEGG_GLYCOSAMINOGLYCAN_DEGRADATION	0	0.007	0	0	0.005	0	0	0.007	0	0.004	0	0	0.01
KEGG_LYSOSOME	0	0	0.008	0	0.005	0	0.004	0.01	0	0.002	0	0	0
REACTOME_SPHINGOLIPID_METABOLISM	0.001	0.009	0.001	0	0	0	0	0.003	0	0	0	0.005	0
Frequencies of biological processes (GO dataset)													
GO_RESPONSE_TO_OXYGEN_CONTAINING_COMPOUND	0.009	0.01	0	0	0	0.004	0	0.006	0.005	0	0	0.001	0
Frequencies of cellular components (GO dataset)													
GO_VACUOLAR_LUMEN	0	0.004	0	0.008	0	0	0	0.006	0.008	0	0	0.003	0
Frequencies of molecular functions (GO dataset)													
GO_ATPASE_ACTIVATOR_ACTIVITY	0.008	0.009	0.008	0.009	0	0.01	0.007	0	0.008	0.006	0.009	0	0
GO_ATPASE_BINDING	0.008	0.009	0	0	0	0.01	0.007	0	0	0.006	0	0	0
Immunological signatures (GSEA)													
GSE10325_LUPUS_BCELL_VS_LUPUS_MYELOID_DN	0	0.002	0	0	0.006	0.001	0.003	0	0	0	0	0.006	0.006
GSE15330_HSC_VS_GRANULOCYTE_MONOCYTE_PROGENITOR_IKAROS_KO_UP	0.003	0.002	0.006	0	0	0.007	0	0	0	0	0.004	0	0
GSE22886_NAIVE_CD8_TCELL_VS_DC_DN	0.007	0.002	0.007	0	0	0.007	0	0	0.01	0.01	0	0.002	0
GSE29614_CTRL_VS_DAY3_TIV_FLU_VACCINE_PBMC_DN	0.008	0.009	0.008	0.009	0	0.01	0	0	0.008	0	0	0	0
GSE29614_DAY3_VS_DAY7_TIV_FLU_VACCINE_PBMC_UP	0.008	0	0.008	0.009	0	0.01	0.007	0	0.008	0.006	0	0	0
GSE29618_BCELL_VS_MDC_DN	0	0.001	0	0.002	0	0	0.006	0	0.004	0.005	0	0	0
GSE29618_PDC_VS_MDC_DN	0.003	0.003	0	0	0	0	0	0	0.01	0.001	0.002	0	0
GSE45365_WT_VS_IFNAR_KO_CD11B_DC_DN	0	0.003	0.008	0.009	0	0	0.006	0	0	0.01	0.005	0	0
GSE9960_GRAM_NEG_VS_GRAM_POS_SEPSIS_PBMC_DN	0	0	0	0.004	0.002	0.001	0	0.009	0.01	0	0.009	0.003	0

Figure 4. Functional pathways evoked by the final 13 significant miRNAs (PedMS *versus* PC). (**A**) Heat map visualization of the most significant pathways involved (KEGG database); (**B**) significant miRNAs and their target genes resulted from our analysis were found to impact several immunological processes (GSEA). Details of significantly implicated pathways (*P*-values, up/down regulations) are reported in Supplementary Material, Table S3.

Since our strategy relied on the more extensive NGS approach for both miRNAs and mRNAs, this extended the analysis to additional molecular players, possibly exploring novel pathogenic hypothesis. In fact, the originality and strength of this investigation are the availability of combined miRNAs/mRNAs data from the same subjects obtained at the same time point, uncovering functional combinations that may be linked to PedMS. The search for target genes and the enrichment pathway analysis identified more complex molecular networks that pointed mainly to the immunological signatures, as expected, and to several other biological processes. It is also worthy to note that, despite the relatively small number of

recruited subjects, both miRNAs and mRNA expressions were able to accurately predict the phenotypes (PedMS and PC), as indicated by ACC values of 92% and 91%, respectively.

According to published reports, we confirmed the upregulation of miR-181a-5p, miR-320a, miR-25–3p, miR-221–3p, let-7a-5p, let-7b-5p, and miR-125a-5p in experimental allergic encephalomyelitis (EAE) or MS, their longitudinal changes often related to long-term immunomodulant/immunosuppressive treatments (18,19,23,24,29,30). Among their potential targets, significant changes in *HLA-DRB5* expression, one of the most confirmed MS genes (31), were combined with the dysregulation of let-7b-5p, miR-320a, and miR-221–3p, reinforcing their possible involvement in the MS pathological network. It is worthy to note that, as for several other target genes, the 3 mentioned miRNAs and *HLA-DRB5* have been found significantly upregulated in our analysis. As others have stated (23), we consider that this incomplete overlap between miRNAs and their target mRNAs may be the expression of a more complex network in which the posttranscriptional regulation of the first players impacts the modulations of the others at different degrees (e.g. without fully degrading their target mRNAs, or acting by combining the effect of several genes with opposite functions), thus explaining once more the difficulty to shed light on the genetic background of MS.

In our study, the technical validation failed to confirm the reported association between miR-21–5p, miR-140–3p, miR-144–5p and MS (33–35). On the contrary, the analysis showed that miR-185–5p, miR-182–5p, miR-99b-5p, miR-652a-3p, miR-942–5p, and miR-148b-3p were significantly dysregulated in subjects with PedMS and might be potential novel biomarkers of the disease at least in the early ages, if confirmed in larger populations. Although none of these miRNAs have been previously reported in MS association studies, miR-185–5p significantly downregulated the expression of genes known in MS (*BACH2*) or NMO (*FCRL3*), whereas the only downregulated miRNA (miR-148b-3p) led to a significantly increased expression of *TMED7*, a gene involved in the type-1 IFN signaling pathway TLR4-related (32), and *KLF4*, a retroviral transduction factor inducing experimental (iPS) neuronal differentiation in MS (36). Furthermore, miR-652–3p and miR-942–5p, together with miR-320a and miR-221–3p, were significantly involved in the down-regulation of immunoglobulins diversification and recombination (GO accession), as in the p38-MK2 pathway (KEGG nomenclature) already reported to contribute to the oxidative stress that has been observed during the inflammatory demyelination of EAE (37) (details in Supplementary Material).

Looking at our data, it is possible to highlight several other miRNA–mRNA combinations that may be worthy of discussion, like the '7-miRNAs loop' (Fig. 3A). In our analysis, these miRNAs shared 9 significant downregulated genes that were experimentally validated (by the Luciferase assay) or computationally confirmed as targets by at least 2 datasets, and were associated with significant expression changes of 592 other genes (up-/downregulated). Besides confirming the association between let-7a-5p, let-7b-5p, and MS (35), our combined analysis uncovered a possible functional connection between these 2 miRNAs and miR-125a-5p, *via* the *DIP2A* gene, whose expression was found to be related to the burden of another neurodegenerative disease like AD (38), and to miR-185a-5p *via MICAL3*, a flavoprotein involved in the cytokine regulation process (39). In the same network, we pointed to the novel association between miR-181a-5p and MS (miR-181 has been mentioned in EAE) via *DNAJA4*, a gene that belongs to the cholesterol biosynthesis pathways (40).

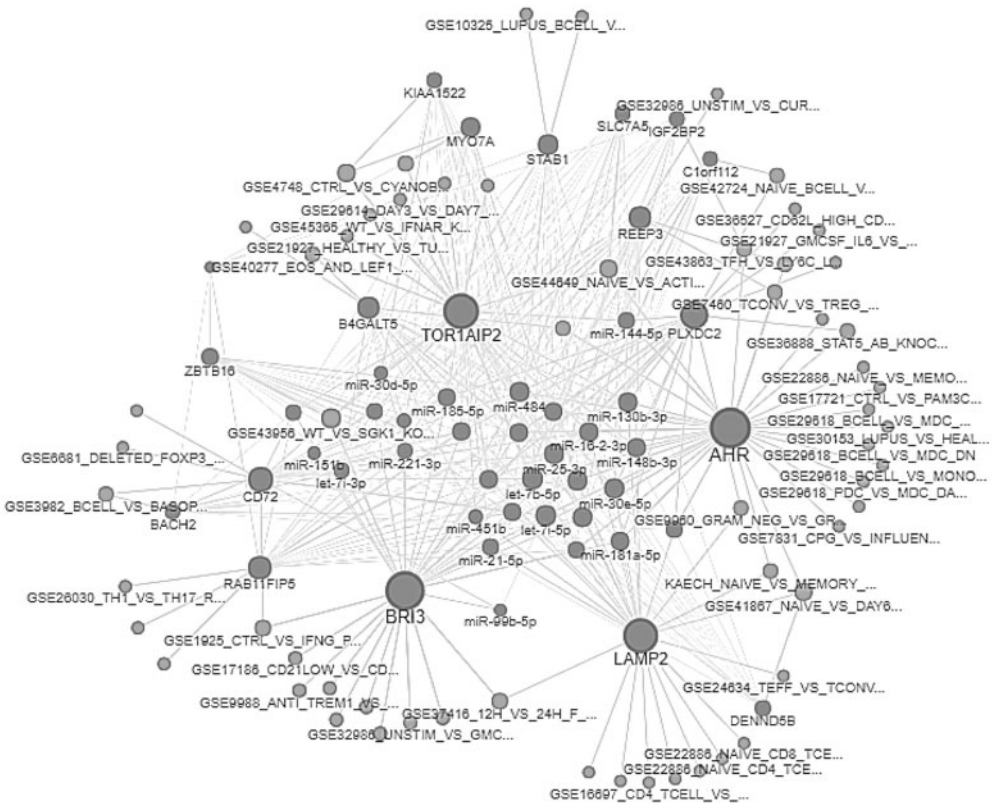
However, since we investigated the recruiting pathways using all gene-level data, we were able to reveal even weak but biologically relevant expression changes of genes, therefore enlightening small but coordinate changes that contribute to the functional networks at different levels. On this basis, we first confirmed that the final 13 validated miRNAs, together with their significant suggested targets (674 mRNAs, up-/downregulated), were implicated in several immunological pathways related to the innate responses of PBMCs, HSCs and DCs, as well as to the adaptive immunity of CD8+ and B-cells (Figs 3B and 4B). Interestingly, miR-320a, let-7a-5p, let-7b-5p, miR-185–5p, and miR-125a-5p and their target genes contributed to upregulate the network involved in the early activation of immune cells (lymphocytes T-B included) to viral infections, partly driven by innate cytokines and IFN type-1. Members of the *TNF* superfamily (e.g. *TNFSF13B*) and the *TLR2* gene were among those enclosed in these functional pathways (41,42), together with the *GRN* gene (targeted by let-7b-5p, miR-125–5p, and miR-320a), confirming once more their involvement in MS pathogenesis (43); all these genes were significantly DE in our analysis (adjusted P-value < 0.05, see Supplementary Material).

On the other hand, we also noted that 'less predicted' pathways or functional categories emerged from the analysis, hence drawing the attention to several of their gene members as promising candidate genes involved in MS susceptibility. Among the others, the KEGG lysosomal cycle was one of the most significantly involved in 5 out of the 7 outlined miRNAs (Fig. 3B and Supplementary Material). In line with recent studies establishing that functional SNPs in key genes of autophagy, like *ATG7* and *ATG16L1*, may be potentially responsible for the subsequent impact on adaptive immune cells' long-term survival (44,45), we found these two genes significantly dysregulated in PedMS compared with PC. Notably, *ATG16L1* was also reported to be implicated (although decreased) in the EAE pathogenesis together with *LAMP2* (46), the latter being one of the most frequently targeted by our validated miRNAs (7/13) and a member of the previously mentioned immunological GSE9960 network that was found to be significantly represented in our study. Furthermore, in the same pathway we discovered that *SORT1*, a recently evaluated – although controversial - player in the immune processes of autoimmune inflammation (47) was significantly upregulated in PedMS *versus* PC, targeted by miR-125–5p, miR-181a-5p and miR-185–5p.

Similarly, target genes related to the ATPase activator activity, ATPase binding and ATPase regulator activity, were found to be significantly implicated: *ABCA1* (ATP binding cassette subfamily A member 1), targeted by miR-25–3p and reported to be associated with MS (48,49); and *GPX3* (target of miR-185–5p), functionally implicated in ALS neurodegeneration (50,51).

Figure 5A (interactive at: http://www.ba.itb.cnr.it/pedms/) summarizes the most frequently implicated genes in our analysis, some of them already reported associated with MS, like ZBTB16, AHR, BACH2, CD72 and LAMP2 (details in Fig. 5B). Genes already mentioned in the previous paragraphs seem crucial, both as targets of several miRNAs of interest and intersectional nodes of functional metabolic network. In vivo functional validations of the most promising targets are in progress, in order to narrow the investigations to fewer candidate genes; subsequent step will include the search for SNPs that may justify the gene expression changes, possibly confirmed in larger MS populations.

This study suffers for some recruitment biases that unfortunately we could not avoid due to limitations imposed by our regional Ethical Committee. As indicated in the opportune



Gene	miRNA count	Functional associations (and references)
BRI3	23	AD (Matsuda S et al., J Biol Chem 2009)
ZBTB16	22	MS (Singh SP et al., J Immunol 2015); Chron Disease (Lin LJ et al, Mol Biol Rep 2014)
TOR1AIP2	19	
DENND5B	18	
RAB11FIP5	18	Autism (Roohi J et al., Am J Med B Neuropsichiatr Genet 2008)
STAB1	18	Bipolar disorder (Witt, Traansl Psy 2014); Schizophrenia (Lee, Gene 2013)
KIAA1522	17	
AHR	16	MS (Tettey P et al., J Neurol Neurosurg Psychiatry 2017)
BACH2	16	MS (Noori-Zadeh A et al., J Neurol Sci 2017)
CD72	16	MS (Kuklina EM et al., Bull Exp Bio Med 2014)
IGF2BP2	16	PD and AD (Chung SJ et al., Park Relat Disord 2015)
MYO7A	16	Leber's optic neuropathy (Wang X et al., Hum Mutat 2011)
PLXDC2	16	Axonal growth (McIntyre JC et al., J Neurosci Res 2010)
REEP3	16	Autism (Castermans D et al., Eur J Hum Genet 2007)
SLC7A5	16	Autism (Tărlungeanu DC et al., Cell 2016); Schizophrenia (Comasco E et al., Neuropsycol 2016)
ZBP1	16	Inflammasome and programmed cell death (Kuriakose T et al., Sci Immunol 2016)
B4GALT5	15	
C1orf112	15	Aging (van Dam S et al., BMC genomics 2012)
LAMP2	15	MS (Liang P et al., Neurosci Bull 2015); PD (Murphy KE et al., Mov Disord 2015);
PRRG4	15	Autism (Yamamoto T et al., Am J Med Genet 2014); PD (Latourelle JC et., BMC Med Genet 2009)

Figure 5. The top 20 target genes resulted from our analysis (interactive at: http://www.ba.itb.cnr.it/pedms/) (**A**) In this Google-graph we represented the 20 most targeted genes, together with their miRNAs (significantly DE between PedMS versus PC, $P < 0.05$) and the most significant pathways in which they were involved (GSEA database). For each gene, the numbers of miRNAs occurrences and pathways interceptions are visualized in the graph by the ring size (the higher number, the bigger ring) and detailed in (**B**). For each gene, we summarized the most interesting functional meaning and association with neurological diseases, if any.

section, PC subjects were significantly younger than those with PedMS: considering the age ranges of our study population, we did not obtain permission to recruit healthy subjects, for example, among PedMS patients' schoolmates, but we had the permission (and parents' consent) to use the blood samples of children who were routinely screened (and tested negative) for inflammatory and hematological diseases at the Hematology Day-Care of the Pediatric Hospital in Bari. Only few of them matched the same age, but we decided to include them anyway, also according to published evidences on age-related molecular signatures in healthy as in pathological conditions, establishing a temporal interval not far from our study range (52,53).

Another possible objection might be represented by the relatively small sample size and the lack of a replication study in an

independent sample, mostly due to the rarity of PedMS occurrence; however, in our view this flaw should not limit the validity of the results considering their prediction accuracy (over 90%).

In conclusion, in our view this extensive analysis enabled us to identify possible molecular signatures of PedMS, confirming the need of such an integrated approach for shedding light on complex diseases like MS. Pathways related to immunological functions were confirmed to be implicated, as well as genes involved in autophagy and ATPase activity, thus suggesting further molecular investigations that may expand the pathogenic hypothesis of MS (Supplementary Material). No significant molecular features were found to be associated with measures of clinical disability or cognitive deterioration. However, a longitudinal 1-year observation of the same subjects is in progress, looking for peculiar molecular profiles of clinical and MRI changes, and responses to DMT. During the developmental ages like in PedMS, this may carry the promising chance to intervene with targeted therapeutic approaches in order to prevent irreversible disabilities.

On the other hand, a validation study on an adult MS group is *in progress*, in order to verify whether the reported molecular profile is a peculiar signature of PedMS rather than a common feature of the disease independently from age.

Subjects and Methods

Study populations

The investigation was performed on samples belonging to Caucasian patients with childhood and juvenile MS (4,54,55) recruited within 5 years from the onset and followed up at the Department of Basic Sciences, Neurosciences and Sense Organs, University of Bari, by neurologists and neuropsychologists with a primary interest in MS. The start date of recruitment was September 2015. At study entry, patients' demographic and clinical features were collected: age, disease course and duration, and Expanded Disability Status Score (EDSS) (56). A validated neuropsychiatric (NPS) test battery, exploring several cognitive domains, was administered to patients with PedMS by a trained neuropsychologist (test duration: 45-min to 2 h). According to published guidelines, cognitive impairment was conferred after a failure of at least three tests (57).

PCs were selected among the young subjects routinely admitted at the Hematology Day-Care of the Giovanni XXIII Pediatric Hospital of Bari who tested negative for any inflammatory and hematologic diseases.

The study was approved by the Ethics Committee of Azienda Ospedaliera Policlinico, University of Bari. Since the study subjects were all under the age of 18 years, their legal tutors signed written informed consent forms (according to the Declaration of Helsinki) at the time of the enrollment.

Molecular analysis of peripheral blood samples

Sample preparation

Peripheral blood samples were taken from patients and controls and stored at −20 °C in 3ml PAXgene Blood RNA Tubes (PreAnalytiX Qiagen/BD, Hombrechtikon, Switzerland). Total RNA was isolated using the PAXgene Blood RNA Kit (PreAnalytiX Qiagen/BD, Hilden, Germany) at ITB CNR, Bari. RNA concentration and purity were measured by Nanodrop ND-1000 (Thermo Scientific, Wilmington, DE, USA) and RNA 6000 Pico chip on Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA), respectively. Samples with RNA Integrity Number (RIN) scores higher than 7 and with A260/A280 values in the 1.8-2.2 range were processed in downstream deep sequencing.

High-throughput next-generation sequencing (HT-NGS)

RNA samples were sequenced using an Illumina HiSeq2500 platform service (http://www.genomix4life.com/en/). Small RNA (sRNA) libraries were prepared using the TruSeq Small RNA Sample Preparation kit (Illumina, San Diego, CA, USA) following the manufacturer's instructions. Briefly, 1 μg of total RNA was ligated with 3′ and 5′ adapters, reverse transcribed, and amplified by 15 cycles of PCR using a unique index for each reaction. PCR products of 140–160bp were isolated following electrophoresis through a 6% Novex Tris-borate polyacrylamide gel, precipitated with ethanol, and fluorimetrically quantified. The quality of libraries was confirmed using a High Sensitivity DNA Analysis Kit on a Bioanalyzer 2100 instrument (Agilent Technologies, Santa Clara, CA, USA). A multiplexed pool consisting of equimolar amounts of individual sRNA-derived libraries was sequenced to generate 50-bp single-end reads, resulting in a final output of around 10 million reads per sample.

The mRNA libraries were prepared using the TruSeq Stranded mRNA Sample Preparation kit (Illumina, San Diego, CA, USA). Briefly, 1 μg of total RNA was used for poly-A mRNA selection with oligo-dT beads, followed by thermal mRNA fragmentation and reverse transcription (RT). The complementary DNAs (cDNAs) were 3′-end adenylated and ligated to Illumina Paired-end sequencing adapters and subsequently amplified by 12 cycles of PCR. The libraries were fluorimetrically quantified and analysed on a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), pooled equimolar into a multiplex sequencing pool and sequenced to generate 2x100bp paired-end reads, leading to a final output of around 30 million reads per sample.

RT and microfluidic qPCR

Approximately 8ng of total RNA/sample was reverse transcribed into cDNA using the TaqMan Advanced miRNA cDNA synthesis kit (Applied Biosystems, Thermo Fisher Scientific). The ends of each mature miRNAs were extended with 5'-end ligation of an adaptor sequence and 3' poly-A tailing, and recognized by universal RT primers (Applied Biosystems, Thermo Fisher Scientific). The obtained cDNA was amplified using the Universal miR-Amp Primers (Applied Biosystems, Thermo Fisher Scientific), diluted, and served as a template for microfluidic qPCR analysis with TaqMan Advanced miRNA Cards (Applied Biosystems, Thermo Fisher Scientific). The TaqMan microfluidic card is a custom-designed miRNA assay, preconfigured in a 384-well format and spotted onto a microfluidic card. The card contains eight sample-loading ports, each connected by a micro channel to 48 miniature reaction chambers, enabling the simultaneous quantitation of 42 human miRNAs plus 5 candidate endogenous reference genes for a total of 4 samples per card (2 replicates per assay). Briefly, 25 μl of diluted pre-amplified product was mixed with 50 μl of TaqMan® Fast Advanced Master Mix (Applied Biosystems, Thermo Fisher Scientific) and 25 μl of nuclease-free water; 100 μl of the PCR reaction mix was dispensed into each port of the TaqMan Advanced miRNA Card. Each card was centrifuged, sealed with a TaqMan low-density array sealer (Applied Biosystems, Life Technologies) and placed in the microfluidic card sample block of an ABI Prism® 7900HT sequence detection system (Applied Biosystems, Life Technologies). Finally, PCR amplification was performed. Thermal cycling parameters were

10min at 92°C to enzyme activation, 40 cycles of denaturation at 95°C for 1sec, and annealing and extension at 60°C for 20 sec. Raw Ct values were calculated using RQ manager software v.2.3 (ABI). (Fig. 1).

Statistical analysis

Two normalization tools, NormFinder (58) and geNorm (59), were used to identify the most suitable endogenous reference genes. NormFinder computes a stability measure based on intra- and inter-group variations of reference genes through an algorithm calculating a stability value inversely correlated with the stability of gene expression. geNorm computes a gene stability value (M) for each candidate gene based on pairwise comparisons of variability. Similar to the NormFinder stability value, lower M-values characterize greater gene stability. The comparison of normalized values between the subgroups was obtained according to the $2-\Delta\Delta Ct$ method (P-value <0.05 was considered statistically significant). Normality of data was assessed by the Shapiro–Wilk test. Statistical differences were verified by Student's unpaired two-tailed t-test or the two-tailed Mann–Whitney U-test, when appropriate. A comparison was performed for every miRNA, and ratios between miRNAs and multiple patterns, representing specific transcriptome profiles, were examined.

Bioinformatics and statistical analysis of HT-NGS data

a. **HT-NGS data acquisition:** Illumina NGS sequencing produced hundreds of millions of sequence (reads) per instrument run. The short reads obtained from sRNA sequencing contained parts of the 3'-end adapter that were removed with a combination of cutadapt software (Supplementary Material, Ref-S1) and *ad-hoc* developed Perl script. Long RNA sequences were split by the sequencing protocol into two paired reads representing the start and the end of the sequence.
b. **Quality check:** The quality control of the obtained reads was checked using the FastQC package (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). In particular, we checked and removed low-quality reads using the following criteria: base sequence quality, sequence quality scores, base sequence content, base GC content, sequence GC content, base N content, sequence length distribution, sequence duplication levels, overrepresented sequences, and kmer content.
c. **Read identification (sRNA):** The sRNA reads were mapped against an ncRNAdb (60), an *in-house-developed* reference database representing a comprehensive and non-redundant dataset of public ncRNA sequences and annotations. The ncRNAdb contains data collected from miRBase (Supplementary Material, Ref-S2), Vega (Supplementary Material, Ref-S3), Ensembl (Supplementary Material, Ref-S4), RefSeq (Supplementary Material, Ref-S5), piRNAbank (http://pirnabank.ibab.ac.in), GtRNAdb (http://gtrnadb.ucsc.edu) and HGNC (http://www.genenames.org). The mapping was performed with Bowtie, an ultrafast, memory-efficient short-read alignment software (Supplementary Material, Ref-S6). The reads that were not mapped to known ncRNAs were aligned against the human genome and passed to mirDeep2 software (Supplementary Material, Ref-S7), which computationally identifies novel miRNA and their mature miRNA products.
d. **Read identification (mRNA):** The reads obtained from total RNA were mapped against the human genome and known human transcripts (GRCh38), downloaded from Vega/Ensembl. The mapping was performed with Bowtie2 (Supplementary Material, Ref-S8), which supports gapped alignment and is faster than the other software on long paired-end reads.
e. **Expression quantification:** In order to obtain reliable read counts and to fix the problem of multireads (reads mapping to more than one reference location), the RSEM tool (Supplementary Material, Ref-S9) was employed for expression estimations. The count values produced by the Bayesian model implemented in RSEM were used as expression values in this work. When normalization of the expressions was necessary for some analysis steps, the trimmed mean of M-values (TMM) normalization method was used (Supplementary Material, Ref-S10).
f. **Differential expression (DE) analysis:** Expression estimations computed for mRNAs (coding genes) and small ncRNAs were compared among the different sample groups (PedMS, PC) with the aim of determining statistically significant changes in expression. Since this is a very crucial step in the bioinformatics workflow and there is no general consensus regarding which method performs best in a given situation, we combined the results of three different software packages for DE analysis: edgeR (Supplementary Material, Ref-S11), DESeq2 (Supplementary Material, Ref-S12) and limma (Supplementary Material, Ref-S13). While edgeR and DESeq2 were designed for NGS data and include data normalization and *P*-value correction for multiple testing by FDR (61), limma was recently enriched for being used with expression obtained from read counts, taking into account the peculiarities of RNA-seq data (Supplementary Material, Ref-S14). In particular, genes were filtered out if they failed to achieve a count per million (cpm) value of 1 in at least 20% of samples. The expressions were simultaneously scale normalized using TMM and variance was stabilized using the voom technique; the corresponding log-cpm values and associated weights were input to the limma standard linear modeling and empirical Bayes for DE analysis. An adjusted *P*-value <0.05 was considered significant for evaluating changes of the expression.
g. **Global differences between phenotypes:** Differences at the global level between different phenotypes by using molecular data were assessed through two methods: Principal Component Analysis (PCA) and Support Vector Machines (SVM). An unbiased estimation of the Prediction Accuracy was assessed using the Leave-One-Out Error (LOOE) procedure (62,63). The percentage of correctly classified samples is an unbiased estimate of the prediction accuracy. In our context, the higher the prediction accuracy, the greater the difference at a molecular level between the two different phenotypes analysed. When we estimated the prediction accuracy using an unbalanced data set, we randomly drew from the largest data set a number of samples equal to the smallest data set and measured the LOOE error on this reduced sample. We repeated this procedure 100 times and evaluated the mean LOOE (Fig. 1, *see* Supplementary Material for additional references).

MicroRNA target analysis

Starting from the results of the DE analysis performed on the two types of datasets (sRNAs and mRNAs), the relationships among differentially expressed miRNAs and

differentially expressed target genes were investigated through a bioinformatics approach. The interactions among miRNAs and their target genes were selected using two databases of experimentally validated bindings (miRtarbase: http://mirtarbase.mbc.nctu.edu.tw; DIANA-Tarbase: http://diana.imis.athena-innovation.gr/DianaTools/index.php? r=site/index), and the output of 5 algorithms for target predictions: miRanda (http://www.microrna.org/microrna/home.do), RNAhybrid (https://bibiserv.cebitec.uni-bielefeld.de/rnahybrid), rna22 (https://cm.jefferson.edu/rna22), mirDB (http://ophid.utoronto.ca/mirDIP), and TargetScan (http://www.targetscan.org/vert_71). The algorithms were run on 3'-UTR sequences from RefSeq human transcripts, with default parameters, except for miRanda (score: 150, energy: 7) and RNAhybrid (*P*-value: 0.1, energy: −22), as suggested in the literature (Supplementary Material, Ref-S15).

To address the low precision and poor sensitivity of the computed target predictions (Supplementary Material, Ref-S16 and S17) and to strengthen the results of interaction identification, we finally selected those bindings that were confirmed by the Luciferase assay in the previously mentioned databases or computationally predicted at least by 2 algorithms (seeds region at the same position reported in at least 2 datasets) (also for this section, see Supplementary Material for additional references).

Pathway analysis

To first unveil common functional pathways, if any, among the validated miRNAs, the web-based computational DIANA-mirPath software was used (64).

In order to determine the role of deregulated miRNAs and target genes in PedMS, we then performed a pathway analysis with a computational method that integrates HT-NGS expression data with miRNA target predictions (65). We considered the association between miRNAs and their target genes together with a collection of sets of genes, associated with particular biological processes (such as Gene Ontology categories), canonical pathways or experimental results. The hypothesis was that if the genes annotated with some biological processes that are targets of a specific miRNA differentially expressed, it was reasonable to infer that this miRNA was involved in the dysregulation of this process.

In detail, for each specific miRNA, the method considered the intersection among gene sets and miRNA-targets associations. An enrichment pathway analysis was performed on these gene sets to identify the biological processes in which these genes – which are also miRNA targets – resulted in up- or down-regulation.

To this aim, an enrichment analysis based on the Gene Set Enrichment Analysis (GSEA) algorithm with unweighted enrichment scores (exponent $P=0$), which distinguishes up- and down- regulated pathways, was adopted. GSEA is a computational method that allows one to retrieve gene sets that are enriched at either the top or the bottom of differentially expressed genes ranked by some measure of differences in gene expression across sample classes (http://software.broadinstitute.org/gsea/index.jsp). An unpaired t-test measures the gene DE in GSEA. GSEA adopts a threshold-free approach and uses the expression data of all the detected genes, thereby reducing the chance of missing weak but biologically relevant expression changes. The method can detect small but coordinated changes of genes that collectively contribute to some biological process.

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

The Authors are grateful to the subjects and their families who kindly accepted to be part of the project. We also thank: Dr. Nicola Losito for the informatics assistance provided during the study; the professional nurses who collected the peripheral blood samples (at the Multiple Sclerosis Center, Department of Basic Sciences, Neurosciences and Sense Organs, University of Bari, and at the Hematology Day-Care of the Giovanni XXIII Pediatric, Hospital of Bari); Dr. Francesco Mariano Caratozzolo for some technical support in qRT-PCR experiments.
Finally, Dr. Liguori wishes to dedicate this manuscript to the memory of Professor Francesco Fera, brilliant scientist and unforgettable friend.

Conflict of Interest statement. Prof. Maria Trojano has served on scientific Advisory Boards for Biogen Idec, Novartis, Almirall, Roche and Genzyme; received speaker honoraria from Biogen-Idec, Bayer-Schering, Sanofi Aventis, Merck-Serono, Teva, Genzyme, Almirall and Novartis; received research grants for her Institution from Biogen-Idec, Merck-Serono and Novartis. The remaining Authors have nothing to disclose.

Funding

This study was fully supported by Fondazione Italiana Sclerosi Multipla (FISM) Grant no. 2014/R/10.

References

1. Compston, A. and Coles, A. (2008) Multiple Sclerosis. *Lancet*, **372**, 1502–1507.
2. Ghezzi, A., Pozzilli, C., Liguori, M., Marrosu, M.G., Milani, N., Milanese, C., Simone, I.L. and Zaffaroni, M. (2002) Prospective study of multiple sclerosis with early onset. *Mult. Scler.*, **8**, 115–118.
3. Simone, I.L., Carrara, D., Tortorella, C., Liguori, M., Lepore, V., Pellegrini, F., Bellacosa, A., Ceccarelli, A., Pavone, I. and Livrea, P. (2002) Course and prognosis in early-onset MS: comparison with adult-onset forms. *Neurology*, **59**, 1922–1928.
4. Banwell, B., Ghezzi, A., Bar-Or, A., Mikaeloff, Y. and Tardieu, M. (2007) Multiple sclerosis in children: clinical diagnosis, therapeutic strategies, and future directions. *Lancet Neurol.*, **6**, 887–902.
5. Benson, L.A., Healy, B.C., Gorman, M.P., Baruch, N.F., Gholipour, T., Musallam, A. and Chitnis, T. (2014) Elevated relapse rates in pediatric compared to adult MS persist for at least 6 years. *Mult. Scler. Relat. Disord.*, **3**, 186–193.
6. Olsson, T., Barcellos, L.F. and Alfredsson, L. (2017) Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. *Nat. Rev. Neurol.*, **13**, 25–36.
7. Disanto, G. and Ramagopalan, S.V. (2013) Similar genetics of adult and pediatric MS: age is just a number. *Neurology*, **81**, 1974–1975.
8. Vargas-Lowy, D. and Chitnis, T. (2012) Pathogenesis of pediatric multiple sclerosis. *J. Child. Neurol.*, **27**, 1394–1407.

9. Ziats, M.N. and Rennert, O.M. (2014) Identification of differentially expressed microRNAs across the developing human brain. *Mol. Psychiatry*, **19**, 848–852.
10. Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
11. Pasquinelli, A.E. (2012) MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat. Rev. Genet.*, **13**, 271–282.
12. Waller, R., Goodall, E.F., Milo, M., Cooper-Knock, J., Da Costa, M., Hobson, E., Kazoka, M., Wollff, H., Heath, P.R., Shaw, P.J. and Kirby, J. (2017) Serum miRNAs miR-206, 143-3p and 374b-5p as potential biomarkers for amyotrophic lateral sclerosis (ALS). *Neurobiol. Aging*, **55**, 123–131.
13. da Silva, F.C., Iop, R.D., Vietta, G.G., Kair, D.A., Gutierres Filho, P.J., de Alvarenga, J.G. and da Silva, R. (2016) microRNAs involved in Parkinson's disease: A systematic review. *Mol. Med. Rep.*, **14**, 4015–4022.
14. Vijayan, M., Kumar, S., Bhatti, J.S. and Reddy, P.H. (2017) Molecular Links and Biomarkers of Stroke, Vascular Dementia, and Alzheimer's Disease. *Prog. Mol. Biol. Transl. Sci.*, **146**, 95–126.
15. Huang, Q., Xiao, B., Ma, X., Qu, M., Li, Y., Nagarkatti, P., Nagarkatti, M. and Zhou, J. (2016) MicroRNAs associated with the pathogenesis of multiple sclerosis. *J. Neuroimmunol.*, **295-296**, 148–161.
16. Gandhi, R., Healy, B., Gholipour, T., Egorova, S., Musallam, A., Hussain, M.S., Nejad, P., Patel, B., Hei, H., Khoury, S. *et al.* (2013) Circulating microRNAs as biomarkers for disease staging in multiple sclerosis. *Ann. Neurol.*, **73**, 729–740.
17. Vistbakka, J., Elovaara, I., Lehtimäki, T. and Hagman, S. (2017) Circulating microRNAs as biomarkers in progressive multiple sclerosis. *Mult. Scler.*, **23**, 403–412.
18. Hecker, M., Thamilarasan, M., Koczan, D., Schröder, I., Flechtner, K., Freiesleben, S., Füllen, G., Thiesen, H.J. and Zettl, U.K. (2013) MicroRNA expression changes during interferon-beta treatment in the peripheral blood of multiple sclerosis patients. *Int. J. Mol. Sci.*, **14**, 16087–16110.
19. Ingwersen, J., Menge, T., Wingerath, B., Kaya, D., Graf, J., Prozorovski, T., Keller, A., Backes, C., Beier, M., Scheffler, M. *et al.* (2015) Natalizumab restores aberrant miRNA expression profile in multiple sclerosis and reveals a critical role for miR-20b. *Ann. Clin. Transl. Neurol.*, **2**, 43–55.
20. Waschbisch, A., Atiya, M., Linker, R.A., Potapov, S., Schwab, S., Derfuss, T. and Kleinschnitz, C. (2011) Glatiramer acetate treatment normalizes deregulated microRNA expression in relapsing remitting multiple sclerosis. *PLoS One*, **6**, e24604.
21. Niu, Y., Wu, Y., Huang, J., Li, Q., Kang, K., Qu, J., Li, F. and Gou, D. (2016) Identification of reference genes for circulating microRNA analysis in colorectal cancer. *Sci. Rep.*, **6**, 35611.
22. Zheng, G., Wang, H., Zhang, X., Yang, Y., Wang, L., Du, L., Li, W., Li, J., Qu, A., Liu, Y., Wang, C. and Navarro, A. (2013) Identification and validation of reference genes for qPCR detection of serum microRNAs in colorectal adenocarcinoma patients. *PLoS One*, **8**, e83025.
23. Freiesleben, S., Hecker, M., Zettl, U.K., Fuellen, G. and Taher, L. (2016) Analysis of microRNA and Gene Expression Profiles in Multiple Sclerosis: Integrating Interaction Data to Uncover Regulatory Mechanisms. *Sci. Rep.*, **6**, 34512.
24. Regev, K., Paul, A., Healy, B., von Glenn, F., Diaz-Cruz, C., Gholipour, T., Mazzola, M.A., Raheja, R., Nejad, P., Glanz, B.I. *et al.* (2016) Comprehensive evaluation of serum microRNAs as biomarkers in multiple sclerosis. *Neurol. Neuroimmunol. Neuroinflamm.*, **3**, e267.
25. Regev, K., Healy, B.C., Khalid, F., Paul, A., Chu, R., Tauhid, S., Tummala, S., Diaz-Cruz, C., Raheja, R., Mazzola, M.A. *et al.* (2017) Association Between Serum MicroRNAs and Magnetic Resonance Imaging Measures Multiple Sclerosis Severity. *JAMA Neurol.*, **74**, 275–285.
26. Wu, T. and Chen, G. (2016) miRNAs participate in MS pathological processes and its therapeutic response. *Mediators Inflamm.*, 4578230.
27. Gandhi, R. (2015) miRNA in multiple sclerosis: search for novel biomarkers. *Mult. Scler.*, **21**, 1095–1103.
28. Yang, Q., Pan, W. and Qian, L. (2017) Identification of the miRNA-mRNA regulatory network in multiple sclerosis. *Neurol. Res.*, **39**, 142–151.
29. Bergman, P., James, T., Kular, L., Ruhrmann, S., Kramarova, T., Kvist, A., Supic, G., Gillett, A., Pivarcsi, A. and Jagodic, M. (2013) Next-generation sequencing identifies microRNAs that associate with pathogenic autoimmune neuroinflammation in rats. *J. Immunol.*, **190**, 4066–4075.
30. Muñoz-Culla, M., Irizar, H., Castillo-Triviño, T., Sáenz-Cuesta, M., Sepúlveda, L., Lopetegi, I., López de Munain, A., Olascoaga, J., Baranzini, S.E. and Otaegui, D. (2014) Blood miRNA expression pattern is a possible risk marker for natalizumab-associated progressive multifocal leukoencephalopathy in multiple sclerosis patients. *Mult. Scler.*, **20**, 1851–1859.
31. Caillier, S.J., Briggs, F., Cree, B.A., Baranzini, S.E., Fernandez-Viña, M., Ramsay, P.P., Khan, O., Royal, W., 3rd, Hauser, S.L., Barcellos, L.F. and Oksenberg, J.R. (2008) Uncoupling the roles of HLA-DRB1 and HLA-DRB5 genes in multiple sclerosis. *J. Immunol.*, **181**, 5473–5480.
32. Zhou, Q., Hao, L., Huang, W. and Cai, Z. (2016) The Golgi-Associated Plant Pathogenesis-Related Protein GAPR-1 Enhances Type I Interferon Signaling Pathway in Response to Toll-Like Receptor 4. *Inflammation*, **39**, 706–717.
33. Sanders, K.A., Benton, M.C., Lea, R.A., Maltby, V.E., Agland, S., Griffin, N., Scott, R.J., Tajouri, L. and Lechner-Scott, J. (2016) Next-generation sequencing reveals broad down-regulation of microRNAs in secondary progressive multiple sclerosis CD4+ T cells. *Clin. Epigenetics*, **8**, e87.
34. Ma, X., Zhou, J., Zhong, Y., Jiang, L., Mu, P., Li, Y., Singh, N., Nagarkatti, M. and Nagarkatti, P. (2014) Expression, regulation and function of microRNAs in multiple sclerosis. *Int. J. Med. Sci.*, **11**, 810–818.
35. Magner, W.J., Weinstock-Guttman, B., Rho, M., Hojnacki, D., Ghazi, R., Ramanathan, M. and Tomasi, T.B. (2016) Dicer and microRNA expression in multiple sclerosis and response to interferon therapy. *J. Neuroimmunol.*, **292**, 68–78.
36. Song, B., Sun, G., Herszfeld, D., Sylvain, A., Campanale, N.V., Hirst, C.E., Caine, S., Parkington, H.C., Tonta, M.A. and Coleman, H.A. (2012) Neural differentiation of patient specific iPS cells as a novel approach to study the pathophysiology of multiple sclerosis. *Stem Cell Res.*, **8**, 259–273.
37. Wang, L., Li, B., Quan, M.Y., Li, L., Chen, Y., Tan, G.J., Zhang, J., Liu, X.P. and Guo, L. (2017) Mechanism of oxidative stress p38MAPK-SGK1 signaling axis in experimental autoimmune encephalomyelitis (EAE). *Oncotarget*, **8**, 42808–42816.
38. De Jager, P.L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L.C., Yu, L., Eaton, M.L., Keenan, B.T., Ernst, J. and McCabe, C. (2014) Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat. Neurosci.*, **17**, 1156–1163.
39. Liu, Q., Liu, F., Yu, K.L., Tas, R., Grigoriev, I., Remmelzwaal, S., Serra-Marques, A., Kapitein, L.C., Heck, A.J.R. and Akhmanova, A. (2016) MICAL3 flavoprotein monooxygenase

forms a complex with centralspindlin and regulates cytokinesis. *J. Biol. Chem.*, **291**, 20617–20629.
40. Robichon, C., Varret, M., Le Liepvre, X., Lasnier, F., Hajduch, E., Ferré, P. and Dugail, I. (2006) DnaJA4 is a SREBP-regulated chaperone involved in the cholesterol biosynthesis pathway. *Biochim. Biophys. Acta*, **1761**, 1107–1113.
41. Sonar, S. and Lal, G. (2015) Role of Tumor necrosis factor superfamily in neuroinflammation and autoimmunity. *Front. Immunol.*, **6**, 634.
42. Hossain, M.J., Tanasescu, R. and Gran, B. (2017) Innate immune regulation of autoimmunity in multiple sclerosis: Focus on the role of Toll-like receptor 2. *J. Neuroimmunol.*, **304**, 11–20.
43. Vercellino, M., Fenoglio, C., Galimberti, D., Mattioda, A., Chiavazza, C., Binello, E., Pinessi, L., Giobbe, D., Scarpini, E. and Cavalla, P. (2016) Progranulin genetic polymorphisms influence progression of disability and relapse recovery in multiple sclerosis. *Mult. Scler.*, **22**, 1007–1012.
44. Shaw, S.Y., Tran, K., Castoreno, A.B., Peloquin, J.M., Lassen, K.G., Khor, B., Aldrich, L.N., Tan, P.H., Graham, D.B., Kuballa, P. *et al.* (2013) Selective modulation of autophagy, innate immunity, and adaptive immunity by small molecules. *ACS Chem. Biol.*, **8**, 2724–2733.
45. Stranks, A.J., Hansen, A.L., Panse, I., Mortensen, M., Ferguson, D.J., Puleston, D.J., Shenderov, K., Watson, A.S., Veldhoen, M., Phadwal, K. *et al.* (2015) Autophagy controls acquisition of aging features in macrophages. *J. Innate Immun.*, **7**, 375–391.
46. Liang, P. and Le, W. (2015) Role of autophagy in the pathogenesis of multiple sclerosis. *Neurosci. Bull.*, **31**, 435–444.
47. Reuter, E., Weber, J., Paterka, M., Ploen, R., Breiderhoff, T., van Horssen, J., Willnow, T.E., Siffrin, V. and Zipp, F. (2015) Role of sortilin in models of autoimmune neuroinflammation. *J Immunol.*, **195**, 5762–5769.
48. Nelissen, K., Mulder, M., Smets, I., Timmermans, S., Smeets, K., Ameloot, M. and Hendriks, J.J.A. (2012) Liver X receptors regulate cholesterol homeostasis in oligodendrocytes. *J. Neurosci. Res*, **90**, 60–71.
49. Mailleux, J., Vanmierlo, T., Bogie, J.F., Wouters, E., Lütjohann, D., Hendriks, J.J. and van Horssen, J. (2017) Active liver X receptor signaling in phagocytes in multiple sclerosis lesions. *Mult. Scler.*, In Press, doi: 10.1177/1352458517696595.
50. Tanaka, H., Shimazawa, M., Takata, M., Kaneko, H., Tsuruma, K., Ikeda, T., Warita, H., Aoki, M., Yamada, M., Takahashi, H. *et al.* (2013) ITIH4 and Gpx3 are potential biomarkers for amyotrophic lateral sclerosis. *J. Neurol.*, **260**, 1782–1797.
51. Merino, J.J., Roncero, C., Oset-Gasque, M.J., Naddaf, A. and González, M.P. (2014) Antioxidant and protective mechanisms against hypoxia and hypoglycaemia in cortical neurons in vitro. *Int. J. Mol. Sci.*, **15**, 2475–2493.
52. Xu, B., Hsu, P.K., Karayiorgou, M. and Gogos, J.A. (2012) MicroRNA dysregulation in neuropsychiatric disorders and cognitive dysfunction. *Neurobiol. Dis.*, **46**, 291–301.
53. Cristino, A.S., Williams, S.M., Hawi, Z., An, J.-Y., Bellgrove, M.A., Schwartz, C.E., Costa, L. d F. and Claudianos, C. (2014) Neurodevelopmental and neuropsychiatric disorders represent an interconnected molecular system. *Mol. Psychiatry*, **19**, 294–301.
54. Chitnis, T., Tardieu, M., Amato, M.P., Banwell, B., Bar-Or, A., Ghezzi, A., Kornberg, A., Krupp, L.B., Pohl, D., Rostasy, K. *et al.* (2013) International Pediatric MS Study Group Clinical Trials Summit: meeting report. *Neurology*, **80**, 1161–1168.
55. Krupp, L.B., Tardieu, M., Amato, M.P., Banwell, B., Chitnis, T., Dale, R.C., Ghezzi, A., Hintzen, R., Kornberg, A., Pohl, D. *et al.* (2013) International Pediatric Multiple Sclerosis Study Group criteria for pediatric multiple sclerosis and immune-mediated central nervous system demyelinating disorders: revisions to the 2007 definitions. *Mult. Scler.*, **19**, 1261–1267.
56. Kurtzke, J.F. (1983) Rating neurological impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*, **33**, 1444–1452.
57. Amato, M.P., Goretti, B., Ghezzi, A., Lori, S., Zipoli, V., Moiola, L., Falautano, M., De Caro, M.F., Viterbo, R., Patti, F. *et al.* (2010) Cognitive and psychosocial features in childhood and juvenile MS: two-year follow-up. *Neurology*, **75**, 1134–1140.
58. Andersen, C.L., Jensen, J.L. and Ørntoft, T.F. (2004) Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer. Res*, **64**, 5245–5250.
59. Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*, **18**, RESEARCH0034.
60. De Caro, G., Consiglio, A., D'Elia, D., Gisel, A., Grillo, G., Liuni, S., Tulipano, A. and Licciulli, F. (2015) The NonCode aReNA DB: a non-redundant and integrated collection of non-coding RNAs. *EMBnet J.*, **21**, e834.
61. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.*, Series B, **57**, 289–300.
62. Vapnik, V. (1998) *Statistical Learning Theory*. John Wiley & Sons editors. NJ.
63. Luntz, A. and Brailovsky, V. (1969) On estimation of characters obtained in statistical procedure of recognition. *Tehnicheski Kibernetika*, vol. **3**,
64. Vlachos, I.S., Zagganas, K., Paraskevopoulou, M.D., Georgakilas, G., Karagkouni, D., Vergoulis, T., Dalamagas, T. and Hatzigeorgiou, A.G. (2015) DIANA-miRPath v3.0: deciphering microRNA function with experimental support. *Nucleic Acids Res.*, **43**, W460–W466.
65. Lee, T.I. and Young, R.A. (2013) Transcriptional regulation and its misregulation in disease. *Cell*, **152**, 1237–1251.